Title
Essays on Product Strategies through Consideration of Individual Distributions

Author
Perry, Hagit

Publication Date
2016

Essays on Product Strategies through Consideration of Individual Distributions

by

Hagit Perry

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor J. Miguel Villas Boas, Chair
Assistant Professor Minjung Park
Assistant Professor Przemyslaw Jeziorski
Professor Sofia Berto Villas-Boas

Spring 2016

Essays on Product Strategies through Consideration of Individual Distributions

Abstract

Essays on Product Strategies through Consideration of Individual Distributions

by

Hagit Perry

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor J. Miguel Villas Boas, Chair

Marketing literature and practitioners are in agreement that it is essential for brands in competitive markets to identify segments that should be targeted and to build rational product strategies that target these segments. It is essential because most markets include consumers with heterogeneous preferences and precise segmentation and targeting creates product differentiation, which prevents direct competition and allow the market to reach an optimal profit optimization equilibrium.

In this consumer markets' era, defined by practitioners as the big data era, consumers' individual transactions and actions, which reveal their preferences, became highly available to marketers. This allows marketers to greatly improve their targeting and to optimize their profits through that.

This dissertation contains three essays that examine optimal products strategies with consideration of individual distributions. Through the models that are built and estimated, individual preferences are identified. Following that, individuals are aggregated into clustered segments, and clear optimization strategy is designed.

All the essays build and discuss structural models and estimation strategies. Each estimation uses unique datasets that were selected and organized carefully for the purpose of robust identification of the varied effects that are examined and analyzed.

Each essay identifies and considers the individual distributions in the analyses. Altogether, the essays provide a deeper understanding of how to consider individual distributions in varied settings and marketing needs that marketers face frequently.

Chapter 1 examines the theory of trying, forgetting, and sales in empirical settings. This is an important model as there are many markets where consumers need to try products for realizing their fit, however after trying, some consumers may forget the fit over time through learning processes of competitive information and other processes.

The theory shows that the trying and forgetting model predicts that sales will occur periodically according to the magnitudes of the effects as the sales are used by the brands as product-fit reminders to the targeted consumers.

For the empirical examination of this theory, a model that includes trying and forgetting

effects within the standard demand side model is built and estimated. The model allows consumers to have heterogeneous tastes and includes treatment for possible endogeneity.

Using the demand estimation and including an individual level distribution estimation, the population is divided into segments. Consumers are divided by their utilities for products as it is optimal for firms to target with the regular price the segments that favor the and when they launch sales, they may target more segments as the trying experience may affect their utility and make them be included in the main segment that this firm targets.

This segmentation of the data makes it possible to find the equilibrium in which each firm optimizes profits and the market does not enter a situation of direct competition and a Bertrand game as the firms focus on the segments that favor them and launch temporal sales to introduce or remind consumers of the products fit. This allows the identification prices strategies that optimize profits.

Chapter 1 also builds a novel dynamic game supply side model together with simulation strategy and technique for that. This is a major contribution as it finds the equilibrium of a multi agent, segments, states, and periods dynamic game for these common settings where firms need to design a long-term, per period, pricing menu as they cannot change their product pricing often.

The results of the estimation and simulation show that the trying and forgetting effects are highly significant on the demand side, but are not used well by some brands through their introduction period and afterward, which greatly and negatively influence their market share and long-term profits.

Chapter 2 examines a method of finding individual level preference for attributes across products and the importance that it can have on policy makers, marketers, and consumers. It specifically discusses the case of reducing overweight in the population through finding the willingness to pay for the fat attribute of products among consumers that consistently buy fattier products at varied categories and introducing these consumers to products that are healthier for them through promotions on those products.

This is an important question as overweight is was recognized as a global epidemic and thus researchers and policy makers are consistently looking for solutions with no consistent finding yet as neither macro taxes of attributes such as sugar or fat nor or macro subsidies of healthier products were feasible, effective, or efficient.

It shows that the standard model does not allow targeted and effective promotions to these consumers as there is a gap in willingness to pay for fat through the population compared to the targeted group. However, using the estimation of the individual level distributions, this part shows that it is possible to convert this segment of consumers to choose healthier products through small magnitude promotional pricing.

Chapter 3 examines a case that is opposite to the previous chapters. While in the previous chapters the segments were revealed through the estimation and individual distribution estimation methods. The data in this chapter saliently reveals that 20% of consumers increased their per unit spend in a durable goods category at the first months of the US sub-prime recession of 2008. This hints that a large portion of the consumers became price loving at the beginning of one of the most difficult periods of the US economy.

This is clearly the opposite to the expectation, thus chapter examines the data carefully and suggest varied models. Finally, it shows that in this case, a well specified demand model can identify the reasons for the initial confusion coming from the data.
Altogether, the essays examine frequent market settings that were not examined before and provide models together with estimation strategies and methods, which allow better optimization of product strategies through the consideration of individual level distributions and through segmenting the population accordingly.

To Mom and Dad

Words cannot express my gratitude for all that you are and all that you do

and

To Shirley

Who gives me the inspiration and energies to everything that I do

Contents

List of Figures

List of Tables

Acknowledgments

I would like to deeply thank my advisors Professor J. Miguel Villas-Boas, Assistant Prof. Minjung Park, and Assistant Prof. Przemyslaw Jeziorski for advising me through the development of the essays in this thesis, undivided attention and many helpful discussions.
I would also like to greatly thank Prof. Sofia Berto Villas-Boas, for her support both on advising me and on finding and requesting the data for two of this dissertation's chapters.
Additionally, I would like to thank the marketing group at UC Berkeley, Haas School of Business, for their commitment and support throughout the process of my PhD studies and candidacy. Special thanks go to Professor Ganesh Iyer for his support in me as a PhD student and candidate and to Professor Wasim Azhar for his support in me as a lecturer of business courses.

Chapter 1

Optimal Pricing when Trying and Forgetting affect Consumers' Choices

1.1 Overview

Consumers' level of information about the fit of a product or a service was long recognized by the literature as a key factor in consumers decision-making process when facing a choice menu. Hence, marketing and economics literature has considered processes of consumers' information gain through varied channels and settings, which can be divided to pre-trial and post-trial as there are major differences between these processes.

The pre-trial literature is focused on consumers' information gain through marketing tools, such as advertising and signaling, and the post-trial aims at identifying the effect of each product or service's trials on the consumers who tried.

While both types of processes of information-gain are very important to marketers, practitioners note that most promotional pricing is intended to induce buyers to try a product (Nagle, 2006) and experiments show that trying has a significant effect on consumers' choices and market share, e.g., Heiman and Muller, 1996. Moreover, literature discusses how uncertainty (e.g., Tversky and Kahneman, 1974) and switching costs (e.g., Farrel and Klemperer, 2007) cause market inefficiencies where new products are available to consumers, but consumers do not try them and eventually these products cannot gain enough market share.

In fact, research groups note that most new products that hit the shelves go unnoticed and the typical failure rate of new product launches can be anywhere in the 85% to 95% range. [1] While these give a high motivation for marketers to identify the effect that trying has on their targeted consumers because the identification can allow quantifying the required investment in providing ways for consumers to try products, there are still many cases, such as the set of cases that this chapter considers, for which the trying effect was not identified by the literature yet through a structural model.

Two major consumers' information-gain processes that are relevant to trying were studied

[1] http://www.forbes.com/2010/12/03/most-memorable-products-leadership-cmo-network.html

in the literature. One process is when consumers gradually gain information about product fit through a few purchases of the product. This process is highly relevant to services like online shopping and to categories like medications (e.g., Crawford and Shum, 2005) because consumers gain information about the complete features and effects of the products and services over time and through varied experiences as the results of the trial depend on many changing variables, such as the initial condition of the patient or the service providers each purchase time (e.g., online shopping depends on the delivery person).

The other process is when consumers gain enough information about a product (product-A) after they purchase it, however gradually forget some of the information till they face the next choice situation and even afterwards if they did not choose product-A at that choice opportunity.

This process is relevant to durable goods where products and services can be consumed a few times between one purchase to the next. Products that are relevant to this are groceries like packaged coffee and ice cream, home supplies like laundry detergents, cosmetics like perfumes, software and services like music-streaming applications, and more. For example, average coffee packages should last about six weeks of daily household consumption.

These products do not change between one purchase to the other and consumers that had partial information about the fit of a product prior to purchasing can use them many times before the next purchase.

While consumers gain information about the product that they purchased over the period of time that they consume it till the next purchase, in many cases, consumers that have purchased another product at time t_0, will not consume product-A for a long while. Thus, they might forget the information about the fit that they might gained through trying product-A.

Related literature can be divided to a few streams. The first is the literature that studies state dependence from perspectives of inertia and variety seeking (e.g., Erdem 1996 and Chintagunta, 1998), the second, is the stream that studies models of gradual learning process through a few purchases (e.g., Erdem and Keane 1996), and the last is related to choices of consumers when samples of products are available to them (e.g., Bawa and Shoemaker, 2004).

On the demand side, the model in this chapter is an extension of the standard choice model (e.g., Berry 1994) and is highly related to the state dependence literature. However, it finds that inertia or variety seeking, which are commonly studied in this literature, do not explain the choices in the data well. On the other hand, prior choices influence the level of posterior information, thus they influence the rational decision of which product to choose, which may eventually seem like inertia, but are actually explained by rational expectations about the utility from each of the products.

The model builds upon the conclusions that are presented in Chintagunta, 1998, which are that repetitive choice of products is influenced by the utility derived from their attributes. This means that when consumers try a product, they enjoy it as a whole, but they also enjoy the different attributes of it separately.

For example, suppose consumers try a new product of French coffee from the brand Peets-coffee and like the product more than they expected. Other than learning about the specific product that they tried, the information that they gain is also that they liked French coffee as a sub-category and that they liked Peets-coffee as a brand. Chintagunta, 1998 shows that consumers will tend to buy products that include a set of attributes that is close to the set of attributes of the previous product that they purchased earlier. This chapter uses this understanding and trying effects are separated by attributes, such as brand and sub-category. This chapter analyzes these settings, presents a structural model of demand and supply, which consider possible endogeneity of pricing and heterogeneous trying-forgetting (TF) effects, uses a unique dataset to estimate the demand side, and develops a technique, with which the oblivious equilibrium of the pricing strategies is simulated.

The estimation of the trying and forgetting effect for each product in the consideration set allows marketers to quantify how much to invest in providing ways for targeted consumers to try a product and how often they should remind the consumers about the product's fit through inducing them to try, given that some consumers will gradually forget the information that they gained through trying.

Marketing methods that may be used by the marketers are varied. They include promotional pricing, extension of the product line with smaller packages in groceries, development of entry-level versions of software products, and more.

One category that is known for practicing the method of allowing consumers to sample is the cosmetics industry. It is very common for potential customers to enter stores that sell cosmetics, such as Sephora and Nordstrom, and use the tester products on the shelves for trying makeup products, perfumes, and more. Consumers can also ask for small samples for continuing the trial at home.

Another example is software, where consumers can try the products for a limited period before they commit to purchase. Products that use this method are music streaming service, such as the popular service, Spotify, and home-deliveries services, such as the popular services, Google express and Amazon Prime.

However, while these practices are common for cosmetics and software, identifying the trying effect using transactional data in these categories is not obvious because while varied products are tried, once a purchase decision was made, a very long time will pass till the next one will occur. In fact, a specific perfume can be purchased once a year. On the other hand, while there are many durable goods in groceries, for most grocery categories, samples are not used frequently. Thus, trying should be done through purchasing a package of a product.

Together with that, for a few retail-store categories, such as coffee, ice cream, and wine, trying of products that are sold in the supermarket can be done through the branded-stores, restaurants, and more.

In the ice cream category for example, Ben and Jerry's products, are sold both in the supermarket and in branded shops. For coffee, this is even more prevalent. There are many coffee-shops' brands like Starbucks, Dunkin' Donuts, and Peet's coffee, which sell packaged coffee through their own shops and in the supermarket, next to brands that are sold in supermarkets only.

Moreover, as discussed earlier, trying effect of a product is a combination of trying varied attributes or aspects of the product. For example, if a new product is launched by a brand that already sell other products, consumers who tried other products of the same brand, might have more information about the fit of the new product than consumers who do not have prior information about the brand. In the same manner, forgetting of the information that is gained through trying might vary by the brand assortment size in the category that is considered.

Given these, this chapter considers the premium grounded and whole bean coffee between the third quarter of 2007 and the second quarter of 2009. The data include all the individual level transactions that were done by the consumers at a US leading supermarkets chain's Northern California stores.

The market leaders were the strongest coffee-shop brands, Peet's and Starbucks. At the end of 2007, two strong brands entered the house-blend sub-category, the supermarket's premium private label and Dunkin' Donuts. The latter brand also entered the market at that time as it was not sold in retail stores till then.

The launch of the products is important for the estimation as it is clear that consumers could not buy and try the products that are on the shelves earlier than the first time that we observe the purchase in the data. This phenomenon in the data that is used for analyzing differences between prior and posterior were discussed in previous research, such as Erdem and Keane, 1996 and Osborne, 2011.

Moreover, in this case, following the introduction in the US Dunkin' Donuts gained a large market share in the US and within a year, at the end of 2008, became the second largest premium coffee brand after Starbucks in the supermarkets. However, it was not successful in the data taken from Northern California, which raises the question of what could have caused this.

One highly apparent difference in the market of the data that is used, is that Dunkin Donuts had a large coffee-shops' representation in the US, but no coffee-shops in California (till 2015). [2] However, the strongest brand in Northern California was Peet's coffee, which had a strong presence in Northern California between the years 2007 and 2009.

While Peet's has a very large market share in California, it has a tiny market share in the rest of the US where the brand does not have a coffee-shop representation.

These give another perspective to trying products through varied channels, which makes a trying-forgetting model seem natural for studying the consumers' choices in this market.

Lastly for the demand side, this chapter shows that the trying-effect gradually fades. The principle is that using the standard model, consumers must forget the exact fit over time because otherwise for consumers who usually stick to a limited number of product (do not seem variety seeking), who are most of the consumer in this data for example, who liked a new product significantly more than other products post trying it, do not stick to it in most of the other purchases when prices and market conditions do not significantly change.

According to the market research groups, consumers sample varied coffees along each deci-

[2]Dunkin' Donuts opened the first store in California in 2015, however this is not relevant to the data that are used here.

sion period, which is about two months. This provides a lot of varied information to the consumers. Varied literature shows that consumers get confused about which product made the best fit for them.

Moreover if they purchase a product, it seems highly probable that they will forget some of the fit of the other products that they did not choose. The chapter shows that the process of forgetting is highly probable for this market.

It is easy to imagine trying one perfume on one hand and another perfume on the other hand, and after smelling the left hand and then the right hand, the first fragrance is not very well remembered, unless the differentiation was very large and the consumer has a strong preference for one fragrance compared to the other.

This chapter shows that the fit is better if both trying and forgetting dynamics are included in the choice model than if only trying is included in the model. Moreover, as mentioned earlier, the fit is better for trying and forgetting than for the standard model.

Given these, the theoretical model of trying, forgetting, and sales that is discussed in Villas-Boas and Villas-Boas, 2008, which considers the case when consumers are trying products and are gradually forgetting the fit of the products, is the main theoretical concept for the structural model in this chapter.

In the theoretical chapter, the expected pricing equilibrium for the monopoly is of periodic sales according to parameters of forgetting rate and change in level of information post the trial.

This chapter builds a supply side model of a stochastic dynamic-game and simulates it using the estimated parameters to find the optimal pricing strategies of the brands. It uses the concept of oblivious equilibrium, which fits this market very well, and finds a pricing equilibrium of commitment strategies where the behavior of the brands should follow the theory. The simulation's methods are defined to solve the game where each segment holds a different state of information, there are multiple segments with heterogeneous tastes for each of the parameters, and the firms optimize present profits on the long term. The methods that this chapter develops are inspired by the principles that are considered in Weintraub et al., 2008 of what is defined as an *oblivious equilibrium* (OE). This is because partial competition makes a good fit for the settings where the retailers guide the pricing in the market and not each of the competitors as pricing decisions by the brands cannot be made on a frequent basis as a response to competitors' pricing by the competing brands. In fact, the brands are required to define for the retailers a long term pricing menu, and they have limited number of opportunities to change the long term menu or ask for a special promotions. While each state's pricing depends on the previous pricing decisions of each firm, the methods of finding the MPE per period are not efficient and not required for these cases, and this solve the problem of dimensionality. This chapter is among very few chapters that simulates dynamic-game of more than two players with many possible states per discrete time, over more than 50 time periods. The section that discusses the supply model, 1.3, elaborates on these.

The simulation presents the optimal pricing strategies, however the results of the simulation allow an estimation of the optimal investment in providing opportunities for consumers to try the products, which as shown in the chapter, can be through collaborating with channels that sell samples of the products. By that, this chapter also adds to the literature on sampling (e.g., theoretical chapters like Heiman et al., 2001, and empirical chapters like Bawa and Shoemaker, 2004.)

As will be presented in the chapter, the empirical findings are that the optimal pricing behavior given the estimated parameters is of periodic sales. As the information gains is greater the sales will be more frequent and larger in discount. This is as the theoretical model predicts and another important support to the validity of the empirical model.

The rest of the chapter is organized into five sections. Section 1.2 describes the data, section 1.3 presents the models of demand and supply and the equilibrium concept, 1.4 discusses the estimation and its results, and 1.5 discusses the simulation in detail and presents the results. Finally, 3.5 summarizes the contributions of this chapter and addresses future research issues.

1.2 The Coffee Market and the Data

The data that is used for the estimation is taken from one of the largest grocery supermarket chains in the USA. The dataset is based on a panel of three years of all individual level transactional data between 2007 and 2009 that were taken place in the Northern-California stores of the chain. The data for the estimation itself are all the transactions between the third quarter of 2007 and the second quarter of 2009.

The data include all information about the products that were purchased including the size and the price paid in each transaction. The dataset includes 580 thousand households who made 4.5 million transactions. Out of them, 400 thousand transactions were of packaged coffee.

The Northern California market was a mature market in 2007 and it was selected accordingly as it could have been assumed that the market was at an equilibrium. Moreover, it was a good representative of the US market with a few small differences that will be discussed here and will be particularly defined in the summary statistics section.

As briefly mentioned in 1.1, the data that was selected from the coffee market in those years, and especially the one in the Northern California, serves the purpose of this chapter well. One perspective of using data on coffee is that the panel data is very reliable because coffee was Americas most popular beverage after water between 2007 and 2009 [3], with an annual average of more than 55% of American adults that consumed coffee daily, and an overall of 80% of Americans surveyed for the NCAs 2008 National Coffee Drinking Trends survey who reported that they drank coffee during that year. Thus, through these years consumers shopped for coffee frequently and it can be assumed that the consumers' in the data are heterogeneous and that the results represent general phenomena that are relevant to many markets.

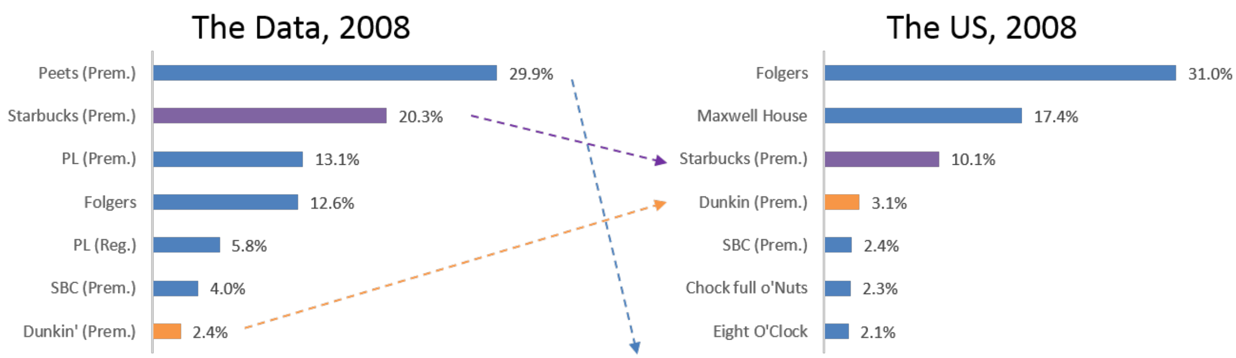
[3] according to data from the National Coffee Association (NCA) annual National Coffee Drinking Trends (NCDT) market study

This is also an important market by itself. Between 2007 and 2009 the total retail sales in the US between $5.7 and $6 billion dollars. Over these years there was a continuous growth and never a decrease, even along the recession years of 2008 and 2009.
According to the National Coffee Association (NCA), three of every four cups of coffee consumed in the United States are made a home. This makes the packaged coffee be a very important part of this market and over 55% of the sales are made in supermarkets.
It is very profitable to sell coffee through branded-channels, such as coffeehouses. In fact, while specialty coffee had long been available from gourmet and other specialty stores, there was no mass exposure to its delights until the 1990s, when coffeehouse chains began proliferating rapidly (correlated with Starbucks' rapid growth). Coffeehouses have exploded across the landscape, going from just 500 in 1991 to more than 25,300 in 2008 [4]. The rise of the coffeehouse culture has had a profound effect on the coffee market, as these venues introduced the general public to specialty coffee, thus creating loyal followings both for higher quality coffee and for specialty coffee drinks.
As discussed in this chapter, the vast availability of coffeehouses has varied implications, and one of them is related to the trying-effect of products that are sold through supermarkets.
There are many categories in coffee, however the ground and whole bean coffees were the largest packaged coffee sub-categories at retail with 62% share of dollar sales through mass-market channels and premium Coffees have started to gradually gain a large market share compared to mainstream coffees, especially in Northern California. Thus, this chapter focuses on ground and whole bean, premium category.
Hundreds of companies market coffee in the United States. These companies range in size from giant global consumer product companies such as Folgers and Starbucks, to domestic specialists such as Peets Coffee & Tea, Inc., to many tiny regional microroasters. Nonetheless, as of April 2008 the top one dozen companies controlled 84% of retail dollar sales through mass-market channels, according to IRI. This was also the case in Northern California. The share spread is presented in figure 1.1.

As can be seen in figure 1.1, the brands' shares in the data are not following the general trends in the US. Firstly, the premium coffee was much favorable than the regular coffee in this market, even though there was a significant difference in unit price between the regular and the premium products, where the mean price for regular coffees was 0.31 per unit and the mean price for premium was 0.87. For example, the strongest market leaders which capture almost 50% in the general US market mainly through their regular coffee products, Folgers and Maxwell House, gained about a third of their total market share in the data. In fact, products that are defined premium gained more than 70% of the market compared to 33% in the US market between the years 2007 and 2009. This comparison is important because it raises an interesting question about the introduction of Dunkin' Donuts in this leading chain's retail-stores in Northern California, which represent at least 15% of the households in Northern California and more of that in proportions of the targeted markets for the premium

[4] by Packaged Facts estimate

Figure 1.1: Brand-share in the data and Comparison to the US Market



Note: The figure for the data is based on the aggregate data from the supermarket chain. The figure for the US is based on IRI sales tracking data through U.S. supermarkets, drugstores and mass merchandisers except Wal-Mart for the 52 weeks ending April 20, 2008. Source: Compiled by Packaged Facts based on data from Information Resources, Inc. InfoScan Review.

coffee brands.

In fall-2007, Procter & Gamble had launched the Dunkin' Donuts packaged coffee brand under license from the regional quick-service chain in what was defined as a strong nationwide rollout. [5]. The brand has realized sales increases during 2007 of 69.7% and the 52 weeks ending April 20, 2008 were of another increase of 83.0%. However, as presented in figure 1.2, in the data the diffusion of Dunkin' Donuts was much slower and eventually gained about half of the market share than in the US, this is even though it potentially had a higher probability to gain a high market share because the market in the data prefers premium brands and according to the summary statistics its consumers like to try new products.

While the rollout in terms of advertising does not seem to be the problem because as discussed earlier the rollout was very strong nationwide, there can be varied reasons for the differences between the US market and the Northern California market through the channel that is observed. However one correlation is interesting in the terms of trying the coffees in the coffeehouses that were discussed in the introduction section. This is that while there was a wide spread of Dunkin' Donuts' coffee houses in east coast, there were no Dunkin' Donuts' coffeehouses in California [6]. At the same time, while Peet's coffee was wide spread in Northern California, there were no Peet's coffeehouses in the rest of the US. This is compared to the spread of Starbucks, which had about equal number of stores in the east-coast and west-coast.

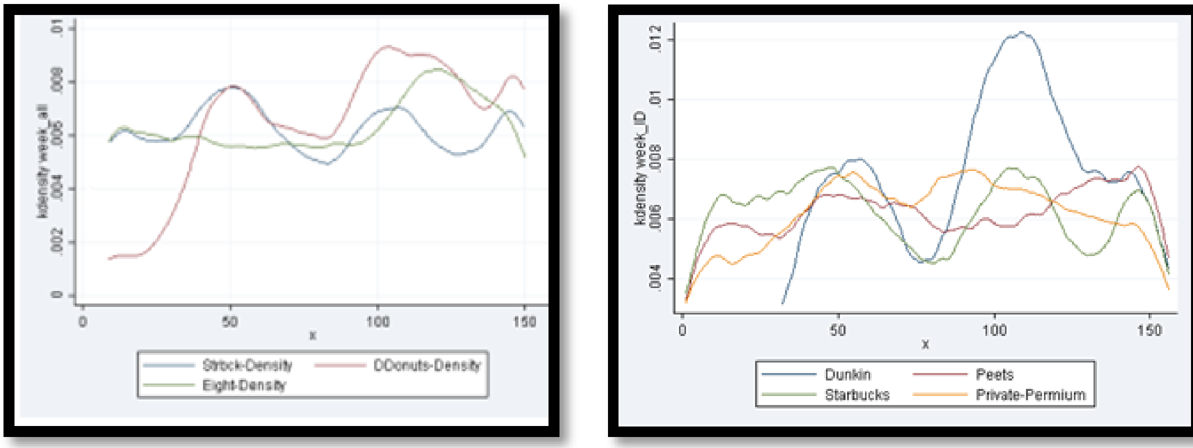
This case of Dunkin' Donuts delayed and lower ramp-up in thed data, raised the question of whether consumers in this market did not like the new products or maybe the trying-effect with the forgetting assumptions makes a better fit.

One perspective that can be considered is the assortment size of Dunkin compared to the

[5] Packaged Facts, the Coffee in the US

[6] The first Dunkin Donuts coffeehouses in California opened in 2015

Figure 1.2: The diffusion of Dunkin Donuts compared to the other brands in its category 2007-2009



(a) The US Market

Note: This figures are a summary of the AC Nielsen scanner data on the coffee category from between the years 2007 and 2009

(b) The Data

Note: This figure is based on the aggregated sales data from the individual level transactions that is used for the estimation in this chapter.

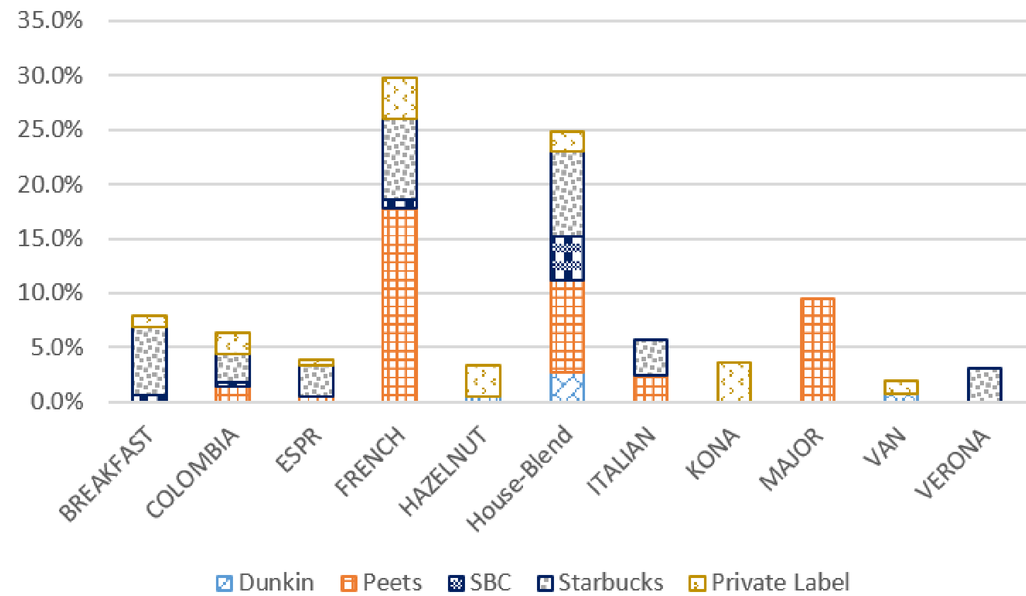
other premium brands. The grained and whole bean coffee has many sub-categories with varied names. The data refining process of the data analysis part has identified 11 large sub-categories, such as House-Blend, French, and Columbian beans. The tiny sub-categories, such as holiday special, were defined in others. Some coffees should be defined as House-Blend, such as Starbuck's Pike, and this was handled by the data clearing process. Two categories had the highest market share, the French coffee, with 30% and house-blend with 25% of the major sub-categories in the premium-coffee market. Figure 1.3 presents the major sub-categories available in the premium grounded and whole bean coffee category. It also shows that each brand had different distribution of its market shares among the brands. As can be seen, Dunkin Donuts' market share distribution was concentrated on the House-blend category.

The market and the data that are used were natural settings for analyzing this question. The retail supermarkets chain that gave the data was among the largest in the US. The retail coffee was growing in these years even through the recession years of mid-2008 till end of 2009. The chain had a strong loyalty club of over 75% of the customers and the customers were using it for their shopping to get large discounts in their baskets. Consumers were buying frequently at the store. In the chapter, I used the customers that shopped at least twice at the store and at least once for premium coffee. [7]

70% of the consumers purchased premium coffees only and about 10% mixed between regular

[7]The reason for that was to observe a panel data of frequent consumers and remove random consumers who may have random choices. Moreover, a single choice of coffee is an evidence that they are interested in packaged coffee and it is in their consideration set. If they had not chosen any packaged coffee throughout

Figure 1.3: Large Assortment of Sub-Categories in the Premium Coffees



and premium coffees. The rest purchased regular coffee throughout the panel data. The households that purchased packaged grained or whole bean coffee along the 7 quarters of the data had purchased it in 10% of their baskets, they had 48 shopping events, where the minimum was 2 and the maximum was 96. They shopped every 2 weeks at the channel and purchased premium coffee every 19 weeks, which is about 4 months, with a large standard deviation, which means that many of them were in the market between the 9th week (after two months) and the 30th week, post the prior purchase. Most of them were in the panel throughout the 7 quarters. The median tenure in the panel was 93 weeks, which is almost the length of the full panel of 96 weeks. The minimum tenure in the panel is 6 weeks. The consumers were mostly loyal to one brand, however they slightly varied the products that they purchased every now and then. The median number of brands and sub-categories that were purchased at least once are 2 from each. The income level depends on the zip codes. The mean level is between $50,000 and $75,000. Table 1.1 presents additional information.

The premium brands pricing were varied along the panel data as presented in **??**. The variation in pricing was through the varied sub-categories, varied brands, and through the quarters. The gap between the possible prices was pretty large and varied between 7.5 and 9.5 per package, which means that the difference is of over 20% between the most expensive product and the least expensive product at the premium coffees category. Moreover, within this 20% gap, the five leading brands were selling in varied pricing points, and the sub-categories themselves were priced differently from each other. Thus, the consumers could have varied their choices according to their tastes for brands, sub-category, taste for variety,

the long panel data then it is a good reason to believe that they are not considering it.

Table 1.1: Summary Statistics - Households in the Panel Data

	p50	p25	p75	Min	Max
# of shopping events	48	31	66	2	96
# of all coffee purchases	5	3	10	2	26
Tenure in Panel	93	88	94	6	95
Shopping Frequency	1.9 (2 wks)	2.8	1.4	3.0	1.0
Inside-Good Purchase Frequency	18.6 (4 mons)	29.3	9.4	3.0	3.7
Inside-Good Purchase % of the carts	10%	10%	15%		
# of brands tasted	2	1	3	1	9
# of sub-categories tasted	2	2	4	1	9
% Premium of all Coffee Purchases	100%	85%	100%		

Table 1.2: Share of new customers and returning customers

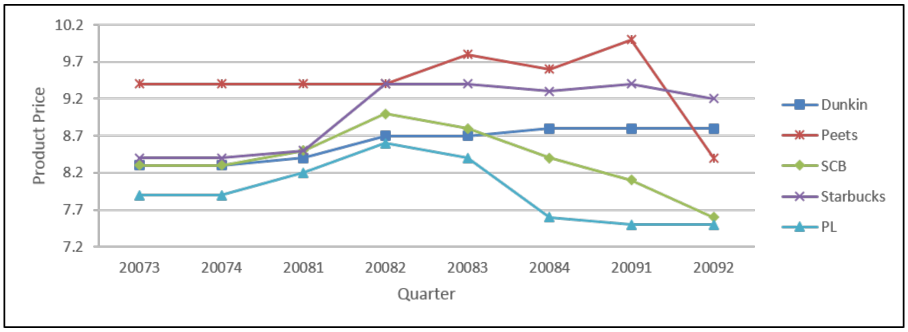
	Dunkin	Peets	Starbucks	Private Label
New Customers	56.5%	23.3%	27.1%	43.5%
Re-Purchases	43.5%	76.7%	72.9%	56.5%

sensitivity to price, and other individual or common demand shocks.

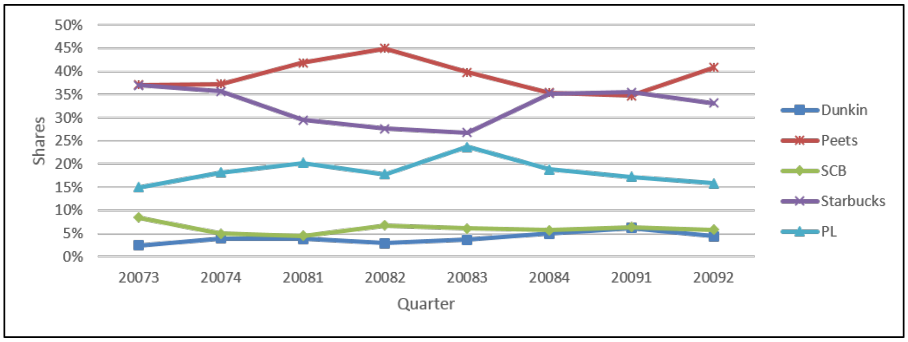
Another important aspects of the data are the ratios of of new customers to the brand and the returning customers per brand as presented in Table 1.2. For the returning customers themselves there is a difference of behavior between returning to the varied brands as can be seen in Table 1.3. Finally, in this chapter we are interested to learn about whether the entry of Dunkin Donuts to the market was optimized or could have been better optimized through an equilibrium between the brands, which according to the theory is should have a pattern of cyclic sales. Since the introduction of Dunkin was mainly done in the house-blend category together with the introduction of the private label brand and since the house-blend category is the second largest sub-category among the other sub-categories, it is of an interest to observe the house-blend category itself separately from the whole market.

Figure 1.5 presents the shares and the mean prices in the premium house-blend subcategory through the panel data. As can be seen, both Dunkin Donuts and the premium private label launched their house-blend products in Q3 of 2007. Thus, their shares start from 0 and ramp up along the first quarter. Dunkin Donuts launched with prices that were slightly lower than its common prices and gradually increased prices to a common level and kept them almost the same level throughout the years, even when other brands lowered the prices along the years of the economic recession. Clearly, these are general trends in pricing, while

Figure 1.4: Prices and Shares in the Premium Coffees Q3/2007 - Q2/2009



(a) Prices



(b) Shares

on a weekly basis prices have varied through temporary sales, which were also local to zip codes. There was a large price variation within each product, which will be discussed in the estimation section.

The shares of Dunkin Donuts and the private label products are consistent with the diffusion process that was presented in figure 1.2. Dunkin Donuts' shares were comparable to the ones of the private label, however there was a delayed gradual increase that peeked at the beginning of 2009.[8]

The estimation section elaborates on how consumers that tried Dunkin Donuts' product continued purchasing it more often than ones that tried the private label's product, which explains the differences in shares, even when prices of Dunkin Donuts' did not adjust to the competition.

To summarize, this section presents the benefits of using this unique settings as they are natural settings to study the question of the trying effect and how to strategically plan the pricing to encourage consumers to try products that their trying effect is significant, positive, and large.

[8]Eventually there was a drop, which may expected given to the not competitive prices of Dunkin Donuts compared to the rest of the products in the sub-category.

Table 1.3: Returning customers post purchase at t_0

# of purchases post t0	Dunkin	Peets	Starbucks	Private Label
1	67.0%	53.5%	46.8%	48.1%
2	19.6%	19.9%	20.8%	20.9%
3	7.2%	9.6%	10.6%	9.4%
4	4.3%	4.8%	6.0%	5.8%
5	1.0%	2.8%	4.3%	5.0%
6	0.5%	2.1%	3.0%	2.3%
7	0.5%	1.9%	1.6%	2.2%
8	0.0%	1.2%	1.3%	1.3%

1.3 The Model and Empirical Framework

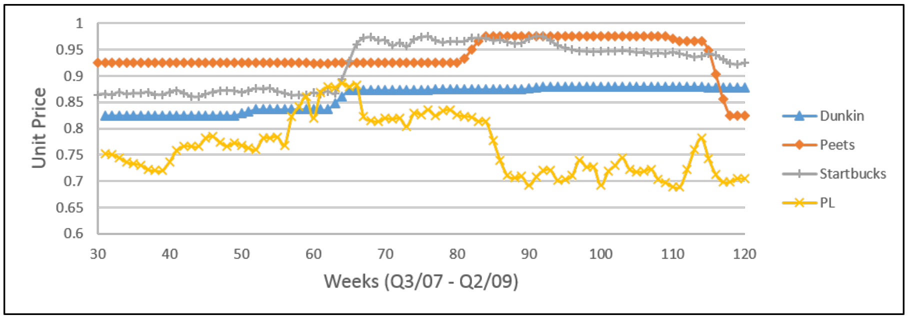
Consider a market in which there are J products (j = 1, ...,J) that belong to J brands. The J firms to whom the brands belong are the price setters and consumers are myopic. The researcher is able to observe the prices and choices of a representative panel of households. One of their possible choices in the choice menu is not to buy any product. Firms compete, week after week, using their pricing strategy. Time is discrete. Since the brands that are in the consideration set are among the biggest brands in the USA, the advertising strategies and advertising budgets are at an equilibrium that are meant to build the brands' equities and there is a strong know how of introductions of new products using advertising. Moreover, the advertising budget is planned at least on a quarterly basis. However the prices are highly variable and can change on a weekly basis according to demand and supply shocks, responses to competition, and other strategic motivations, such as introductory pricing, which I will elaborate on later here.

The model assumes that consumers have some baseline prior information about each product in the consideration set, which they collected from advertising, word-of-mouth, trying samples, or other means. However, when they purchase the product, they have a real opportunity to try it for a while at home and gather more information about its fit to their tastes, preparation capabilities, and usage at home.

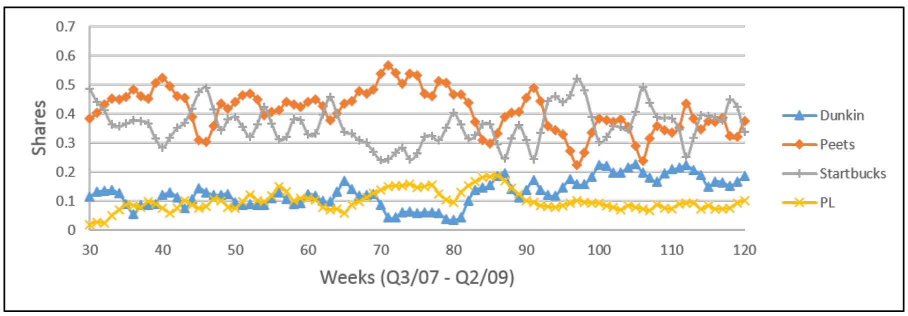
The dynamics of the model stem from the trying effect's carry-over, i.e., the long term effects of trying that encourage re-purchasing a product and carry over to periods post the ones where consumers had incomplete information about the fit and tried. The trying effects are relevant both to the demand and to the supply sides of the model. However, they gradually decrease over time, thus its effect is not guaranteed forever.

In the following, I firstly describe how trying-effect influence product demand in this model. Then I discuss how firms choose their optimal product prices levels in order to maximize

Figure 1.5: Prices and Shares in the Premium House-Blend sub-category 2007-2009



(a) Prices



(b) Shares

their objective function, which is the expected present discounted value of profits. Finally, I define the Oblivious Equilibrium solution concept that captures the strategic interactions between firms.

Demand

The primitives of the model are the product characteristics and the consumer preferences.
It assumed that households buy one product from a choice set S with J + 1 alternatives in each time period. NP denotes the number of households in the panel and NM as the number of markets in which the households made choices. The consumers in a household derive utility from one product in that choice set (buying alternative j = 0 means that the consumer made no purchase of the available brands.)
The household buys the product for which the perceived utility is the greatest, but will make no purchase from the category if the utility of each of the brands is less than the utility of not purchasing.
The utility that consumers in household i at market t [9] obtain from product $j \in 1..J$ depends

[9]for ease of notation and description, the model uses t, which can stand for time if the markets are defined by time only, however in this case, they are defined by both time and store.

on observed and unobserved attributes of the product, where for j=0, $U_{i0m} = 0$. Assume that utility takes the form

$$U_{ijt} = -\alpha_i p_{jt} + \mathbf{X_j} \beta_{\mathbf{ijt}}(\mathbf{s_{ijt}}) + \xi_{jt} + \epsilon_{ijt} \tag{1.1}$$

Where α_i is the mean price sensitivity per market t, p_{jt} is the price of product j at market t, $\mathbf{X_j}$ are product j's attributes, which don't change along the time, β_{ijt} is a vector of tastes per attribute of household i per attribute of product, j.
ξ_{jt} is the product and market fixed effects, and ϵ_{ijt} is the household, product, and market, iid demand shock, at time t.

While in the standard model, the individual taste for a product is known prior to when the choice is made. As discussed earlier, this does not have to be the case. Thus, this model defines the individual taste for the product attributes as follows:

$$\beta_{ijt}(s_{ijt}) = \gamma_{ij}(s_{ijt}) - (t - t_{0ij})\zeta_{ij}(s_{ijt}) \tag{1.2}$$

s_{ijt} indicates whether household i purchased product j before time t, and if so, it tells what was the closest time to time t, which is defined as t_{0ij}. ζ_{ij} is the additional information about the attribute post the first few trials of the product (which are not observed in the data and are assumed given the purchasing event.)

To ease notation, denote:

$$V_{ijt} = -\alpha_i p_{jt} + \mathbf{X_j} \beta_{\mathbf{ij}} + \xi_{jt} \tag{1.3}$$

Then:

$$U_{ijt} = V_{ijt} + \epsilon_{ijt} \tag{1.4}$$

A simplifying assumption commonly made (see McFadden, 1981, BLP 1995, and more) is that ϵ_{ijt} is distributed i.i.d. with a Type I extreme value distribution. Moreover, in this model, the mean utility of the outside good is not identified, and it normalized to 0.
Given these and the previous definitions, the likelihood that a household will purchase product j in market t is defined as follows

$$Prob(y_{ijt} = 1|\theta) = \int \frac{\exp(V_{ijt})}{1 + \sum_{k \in J} \exp(V_{ikt})} df(\cdot|\theta) d\eta_{ijt} \tag{1.5}$$

Where $y_{ijt} = 1$ if household i has chosen product j in market t given the parameters θ, which are the parameters of distribution F, from which the house-level tastes and preferences are drawn.
As in McFadden and Train, 2000, and as in the other chapters that use SMML, tastes are drawn for all the attributes and the interactions of them for products, 1..J, when the taste is for a product. Suppose the distribution F is the normal distribution, then parameters to

be estimated for the distribution are the mean for the tastes and the variation matrix. This is described by the following

$$\begin{aligned} & let\ \eta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \ \ and\ \tilde{\eta_{\mathbf{ijt}}} = \begin{bmatrix} \tilde{\alpha_i} \\ \tilde{\beta_{\mathbf{ijt}}} \end{bmatrix} \\ & Then \\ & \eta_{\mathbf{ijt}} = \eta + \Sigma^{1/2}\tilde{\eta_{ijt}} + \Pi D_i \ \ s.t.\ \ \tilde{\eta_{\mathbf{ijt}}} \ \sim N(0, I_K) \end{aligned} \tag{1.6}$$

where Σ is the matrix of standard error parameters [10], we assume that the parameters are distributed normally, which should be fine for the goal of the estimation, and K equals to the number of random variables, including interaction variables, I is the identity matrix, Π is a $K \times d$ matrix of the coefficients that measure how the taste characteristics (suppose there are K of variables, including the interaction variables) vary with demographics (d demographic variables).

We use the notation t for *beta* and for *eta* as these are functions of t when they are put into the model, however the draws from the distribution are not affected by t and neither the coefficients that *beta* includes as in equation 1.2.

Supply

Firms Decisions

As discussed in the introduction section, the firms in this market have developed strong brands over many years till time t_0. Thus, the firms assume that they will be able to maintain their brand equity through the common advertising know-hows and can predict the long term brand time fixed effects. Each firm control a different brand in the category in which they are competing. Some of the firms hold products of the same brand that are considered outside good. Thus, they may have some expected benefits from consumers buying the outside good.

The market is such that the firms are considered the manufacturers and they sell their products through a retailer. The relationships are such that the firms should recommend a long term pricing plan and direct the retailers according to it. The firms have limited amount of opportunities to request a price decrease or sale along each year.

Since the firms will not be able to immediately react to every promotion that a competitor is launching and prefer to save their unique opportunities to ask for a promotion to when there is a supply or a demand shock, they prefer to initially plan a long term strategy and commit to it. These settings makes the oblivious equilibrium as discussed in Weintraub et al., 2008, a natural equilibrium concept for simulating the equilibrium for the market and reducing the curse of dimensionality that the common dynamic games direct MPE methods

[10] $\Sigma^{1/2}$ is the Cholesky factorization

are suffering from. [11]
Given these, each firm defines a long term pricing strategy and optimizes it given its own state and knowledge of the long run equilibrium distribution of states. This is, at the starting period, t_0 when the firm plans its long-term strategy, the information and last choices states of the market are observed by the firm and so do the preference of the consumers at the varied information states. As discussed earlier, some of them have incomplete information about the fit of some of the products and this is taken into the firms consideration in defining their strategies. The firms who hold the strong brands have valid expectations about the brand-time fixed effects, which for example can affect how choices per brand if consumers prefer to buy super-premium brands which hold coffeehouses around the household area in holiday seasons. They also assume an i.i.d. demand shocks for the consumers. These variables define the elasticities of demand for the products within the category and for the outside good as well. Thus, the expected states of the firm and for the other firms can be calculated, given the firm strategy and the other firms' expected states and expected strategies.

The Dynamic Game

This section formulates the non-zero sum stochastic dynamic game that operates through the firms decisions. [12]
The system evolves over discrete time periods and an infinite horizon. Time is indexed by $t \in T$. There are n firms indexed by $S \in \{1, ..n\}$. The state of each firm is defined by the brand-time fixed effects, the consumers' preferences, and the consumers' levels of information (e.g., per brand or per brands' product-attribute) about the fit of the products to their preferences.
The firms' states evolve through the periods, following the pricing decisions that they make. The action space is continuous. Each strategy, $\sigma_{\mathbf{j}}$, is a sequence of prices, $p_{jt} \in \mathbb{R}$. Each p_t influence consumers' choices at time t. Consumers' choices affect information level about each product fit and create state dependency effects, which affect the choices in the next periods.
At time t, the state of each product-line in the market $j \in J$ is denoted by x_{jt}, the mutual state of all firms together at time t is x_t. The *system state* s_t is defined to be a vector over possible states that specifies for each state what the probability of it to occur is. The probabilities are defined by how many consumers from each segment made choices in time

[11] While this cannot be the case for every market, it is most relevant to markets that consist of strong manufacturers and strong retailers. As mentioned in the introduction, applying this equilibrium concept is not clear for this case and hence, it was not yet applied to these kind of problems. Thus, it is the hope that this chapter will provide a route for more literature that explore similar problems.

[12] This section follows Weintraub et al., 2008 in some ways, but has some important differences as well. It is also highly related to the single agent dynamic problem that is solved through the Nested Fixed Point algorithm (NFXP), Rust (1987,1988).

t-1 that led them to have certain information and state dependency states, which other than seasonality and i.i.d. demand shocks, define the states of j.
At the starting time t_0 the firms define their long-term strategy. The system state is at s_0, and each firm, simultaneously defines its strategy, σ_j. The other agents strategies, σ_{-j}, are assumed to be fixed by each product-line, j, when it defines its strategy.
Each consumer, has specific preferences towards the varied available products in the category and towards the outside good, and each consumer also has a specific level of information per product, which depends on two dynamics, one is whether consumer i has tried attributes of product j, such as another product of product j's brand, and the other is how many consumption events have occurred between the last purchasing event of j interacted with how fast consumer i forgets the trying experience of product j.
The information at time t_0 about the consumers, defines the share of consumers in each state, x_t, which eventually define the weight of the state in each firm's strategy decision.
Each agent's value function is defined as follows

$$\begin{aligned} &V_{jt}(\mathbf{s_t}|\theta, \sigma_{-\mathbf{j}}) = \sup \left\{ \pi_{jt}(\mathbf{s_t}, \mathbf{p_t}) + \beta \int V_{jt}(\mathbf{s}'|\theta, \sigma_{-\mathbf{j}}) f(\mathbf{s}'|\mathbf{s_t}, \mathbf{p_t}) df(\mathbf{s}|\theta) \right\} \\ &Where \\ &t \in 0, .., T \ and \ \forall j \in J \ \sigma_j = (p_{j0}, .., p_{jT}) \end{aligned} \tag{1.7}$$

The supremum is taken with respect to p_{jt} the price set by firm j. The Bellman equation is defined conditional on a specific competitive strategy profile p_{-j}, i.e. a specific assumption by firm j about the behavior of the firms competitors according to the equilibrium concept and the market conditions. The right-hand side of the Bellman equation defines the best response to $\sigma_{-\mathbf{j}}$. β is the common depreciation level of profits as they are viewed in period t.
Where the stationary profit function is defined using the common stationary profit function $\pi_{jt} = (p_{jt} - c_{jt})Q_{jt}$, in which p_{jt} is as defined above and c_{jt} is the variable cost for product j at time t.
$Q_{jt} = Q_{jt}(\mathbf{s_t}, \mathbf{p_t}) = M \times Prob(y_{ijt} = 1|\mathbf{s_t}, \mathbf{p_t}, \theta)$, where
$Prob(y_{ijt} = 1|\mathbf{s_t}, \mathbf{p_t}, \theta) = \int \frac{\exp(V_{ijt})}{1+\sum_{k\in J}\exp(V_{ikt})} df(\cdot|\theta) d\gamma_{ij}$
Starting at t_0 and given the strategies, σ_j and σ_{-j}, each agent, j, calculates the probabilities for transition of the consumers between the states, starting at time t_0 and ending when the depreciation factor, β makes the future smaller than a certain ϵ, and the profit function and the probabilities to be in each state at each time, the value function is calculated and the supermum accordingly.
The transition matrix is evaluated at each time, $t + 1$, based on t, as follows

$$\begin{aligned} &\forall x_{t-1} \in S, \ i \in N, \ j \in J, \ and \ t \in T \\ &Prob(y_{ijt} = 1|x_{t-1} = 1, \sigma) = Prob(y_{ijt} = 1|x_{t-1} = 1, \mathbf{p_t}) \times Prob(x_{t-1} = 1|\sigma) \end{aligned} \tag{1.8}$$

For the trying event there are two states and for the forgetting effect, there are t states, per product j, consumer i, and time t.

$Prob(y_{ijt} = 1|x_{t-1} = 1, \sigma)$ is the probability defined in 1.5, given the parameters of the state, the time, and the pricing strategies.
Given each of the probabilities, $Prob(y_{ijt} = 1|x_{t-1} = 1, \sigma)$, each state (e.g., the state that a consumer chose product j at time t) and the system state, s is updated. This is done through the process in which each choice probability adds a share of consumers into a bucket of a state (with some probability), in which each of the firms is at.
Since we are considering a model in which product attributes, such as the brand, are shared between some products. Thus, suppose r and j share an attribute, when product r is chosen in a certain probability, the system updates the state for the shared attribute for product j as well.

1.4 Demand Estimation - Explaining Consumers Preferences

This chapter strategy, as defined earlier, is to estimate the parameters of the demand model and to simulate the optimal pricing strategy based on the supply model, which integrates the demand model through the probability of consumers' choices at each state. Thus, this section starts with the estimation of the demand and continues with the simulation.

Data Organization for Demand Estimation

As discussed in the data and estimation strategy section, the data that this estimation uses is individual level data of all the transactions made by the members of the loyalty club of one of the largest supermarkets for groceries chain in the US and Coffee is a product that is commonly purchased in this supermarket by the consumers. [13]
To reduce the estimation time, but preserve the significance of the model, 6% of the households were randomly selected (out of 581,000 households that were initially in the data), and the zip codes starting with 945, which represent the data that were described earlier, and are 24% of the data, were selected. Then, out of the 8,366 households, 4,804 households were households that held loyalty card, purchased over once through the years, but not more than twice a week, and purchased at least once out of the houseblend sub-category, which makes them potential customers of that sub-category in which both Dunkin Donuts and the Private Label launched two new products.
After an additional data cleansing, a representative sample is of 2659 consumers, with 27,213 choice situations, including choices from the outside good [14]

[13] According to the market research presented in the introduction and following a discussion with the data scientists of the supermarket chain.

[14] choices from the outside good were randomly selected such that the probability to choose the inside good was about 0.5. The reason is that it is assumed that the consumers are not in the market for coffee in each of their purchases, however they are in the market every 4 weeks on average.

The maximal number of alternatives that one household might face is 26 and the maximal number of choices per household is 46, which mean that households that purchased more than once in two weeks were not included.
As discussed earlier, the names of the products do not clearly define the sub-category of the product. Thus, this is a work that was done for this chapter. The house-blend sub-category was clearly recognized in the data and so were the other categories. As discussed earlier, the inside-good for this chapter is the premium house-blend.
Each of products in the sub-category were defined by the brand, size, whether they are regular or decaf, and unit price. The model assumes that each consumer, i, is choosing one product at each purchase opportunity. Each time a consumer is visiting the supermarket, that consumer has the option to buy from the inside good or the outside good. When purchasing from the inside good, the consumer may purchase a few products at the same time, t, however the model and the estimation assume that one purchase is selected each time. In many cases two products that are defined the same according to the above are purchased (the incentive may be a quantity discount). In these cases, the two are combined to one purchase, the size is increased and the unit price is defined according to the total size.
One challenge was to develop the choice menu that the consumers were facing the shelves in each store. This is because the data did not contain the weekly product location on the shelves or the prices menu that was presented on the shelf per store. Thus, the way to overcome it was to develop a menu by looking at all the purchases in each store and develop the consideration menu. This is because there were many purchases per store and they should reflect the actual variety that is commonly purchased. The weekly unit prices per product (changes with size because of secondary price discrimination) was used for the menu because the discounts were applied to all the consumers that are in the dataset as the panel data that is used is of consumers that used their membership cards.

Possible Endogeneity

The prices in the model defined in 3.1 may be correlated with the error term as prices move with unobserved demand shocks, such as news reports. The estimation handles these possible problems through the control-function method as defined in Petrin and Train, 2009, as follows

$$p_{jt} = W(\tilde{\mathbf{Z}}_{\mathbf{j}}, \zeta_{\mathbf{j}}) + \nu_{jt} \tag{1.9}$$

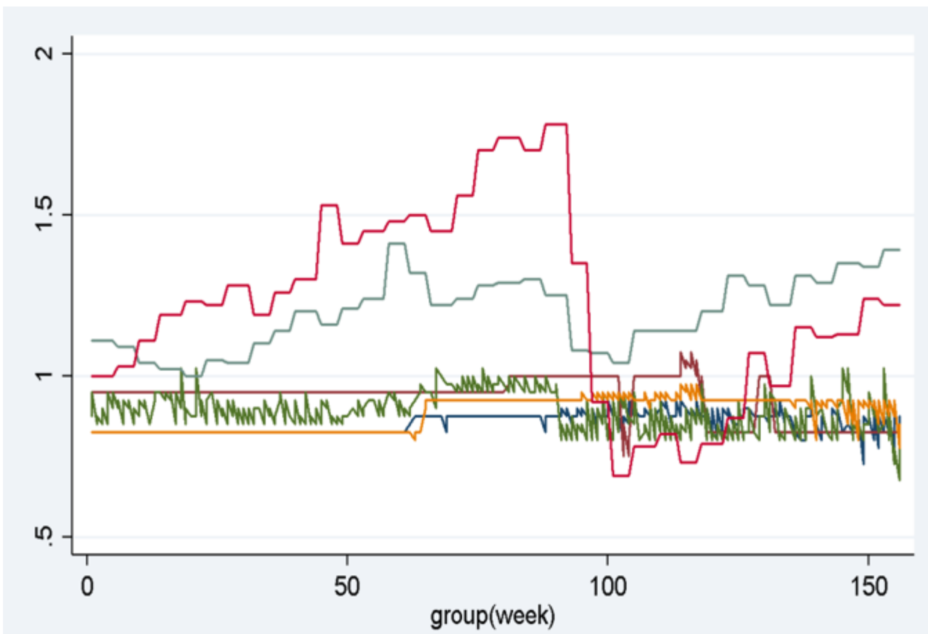
$$p_{jt} = \gamma_{\mathbf{jt}} \mathbf{Z}_{\mathbf{jt}} + \nu_{jt} \tag{1.10}$$

$$p_{jt} = \hat{p_{jt}} + \nu_{jt} \tag{1.11}$$

Different instrument variables which are correlated with price, but are not correlated with the individual error term were considered for equation 3.7, including the GDP and the

commodity prices of green coffee beans, however the variable that make a good fit to the weekly coffee prices are the weekly gas prices. Eventually, for the pricing, more instruments are used, which are all the observed characteristics of the product. In this chapter, W is assumed to be a linear function of the variables, as in equation 3.8, and to ease notation, as in equation 3.9.
The fit is of 96.7%, which is very high, hence does not require additional techniques for the control function in case it was not significantly explain the price. [15]

Figure 1.6: Instrumental Variables



Note: Gas prices in red, Arabica prices (green coffee - commodity prices) are in light-blue. The private label, Starbucks, Peet's coffee, and Dunkin Donut's house-blend unit prices are below. These provide a graphical example of the market and the correlation between the prices.

Demand Estimation

Built upon the model that was discussed in 1.3, and given the data and endogeneity treatments, the following empirical utility model, is a complete model that allows identification of the required parameters

$$U_{ijt} = -\alpha_i \hat{p_{jt}} + \mathbf{X_j}\beta_{\mathbf{ijt}}(\mathbf{s_{ijt}}) + \xi_{jt} + \lambda_{ij}\nu_{jt} + \epsilon_{ijt} \tag{1.12}$$

[15] In cases were the fit is not very high for the linear function as discussed in Petrin and Train 2009, another control function method that does not force a functional form is discussed in Villas-Boas and Winer, 1999, and another method that can be helpful is BLP, 2004.

Where other than the variables and parameters that were discussed in 1.4, $\mathbf{X_j}$ includes the brand dummies, dummies for the 11 major characteristic of the product, e.g., Houseblend and French coffee, and whether it was decaffeinated [16]
To estimate the model, we can write the likelihood function as the product of the likelihood of purchases, given the prices and the product attributes, and conditional on the parameters, which are estimated through the model, as follows

$$L = \prod_{t=1}^{NM} \prod_{j=1}^{J} \prod_{i=1}^{NP} Prob(y_{ijt} = 1|\theta) \tag{1.13}$$

This likelihood function, where NM is the number of markets, J is the number of products per market, and NP is the number of households in the panel. The probability function is as defined in 1.5. [17] There are a few possible techniques to estimate the model. The simulated maximum likelihood as discussed in McFadden and Train, 2000, using the control function, as was discussed earlier. [18]

Estimation Results

The focus of the empirical analysis is the trying and forgetting effects per attribute, which provide the empirical baseline for the supply side simulation that is discussed in section 1.5. The chapter has considered many partial models that are discussed in the appendix section 1.7. The main conclusion is that trying and forgetting effects are very significant in this case for all the models and they make better fit than the standard model that does not include these effects.
As discussed, we are interested in testing whether consumers gain information about products' fit through trying their attributes, but gradually forget the information gained through trying make a good fit to these data.
Thus, we particularly test how significant the trying and forgetting effects per brand are, whether they are positive or negative, and what their magnitudes are. Given the summary statistics that were presented, we expect that the trying effect will be positive and significant to all the brands, but we cannot tell what the forgetting will be.
The model that was described in 1.4 is tested and compared to partial models and to the standard model, as defined in Berry, 1994.
The main results of the estimation, are presented in table 1.4. As we can see, the ever-trying and the forgetting effects are significant and have the expected sign and magnitudes. A positive high ever-trying effect means that post purchasing the consumers' expected utility

[16] Other attributes that were considered included organic indication and roasting level, however both were not significant and were left out.

[17] each household made different number of choices and each brand was selling in a different number of market, which the matlab code overcomes.

[18] It is also possible to use Villas-Boas and Winer, 1999, or BLP, 2004 in case it was important not to use a functional form for the price's possible endogeneity. This was not important for this chapter's purposes.

Table 1.4: Main Effects

		Mean	*Standard Deviation*	*Mean*	*Standard Deviation*
Price		4.22	0.53		
		(2.305)	(0.0495)		
Residual		-0.20	-1.74		
		(0.3412)	(0.9501)		
PL	Ever / Frgt-Rate	1.29	0.78	-0.11	-0.03
		(0.0951)	(0.082)	(0.0157)	(0.021)
Starbucks	Ever / Frgt-Rate	1.24	0.76	-0.07	0.00
		(0.0677)	(0.0554)	(0.0097)	(0.012)
SBC	Ever / Frgt-Rate	2.31	0.98	-0.21	-0.02
		(0.1643)	(0.1395)	(0.0376)	(0.0454)
Peets	Ever / Frgt-Rate	1.23	1.08	-0.05	-0.02
		(0.0717)	(0.0596)	(0.0111)	(0.0144)
Dunkin	Ever / Frgt-Rate	2.54	0.92	-0.44	-0.31
		(0.3415)	(0.2806)	(0.1443)	(0.1139)
LLK	-35750.454				

Note: Standard errors in parentheses. When the second column is Ever/Frgt-Rate, the brand effect of ever trying is in the first two columns and the forgetting effect is in the second two columns from the left. The rest of the estimated effects are in Table 1.5 and in Table 1.6. The LLK is the value of the log likelihood including the fixed terms.

from the product that contains the brand is higher than before purchasing. The trying-effect gradually fades over time at the pace that is defined by the forgetting effect per brand.

The results of the estimation show that Dunkin Donuts and SBC that do not have a large representation on the shelves as the other products, have larger ever-trying effects, which means that inducing consumers to try should be more important to these brands' than to other brands in the category. However, the consumers that tried Dunkin Donuts' forgot the fit four times faster than the ones that tried the Private Label and nine times faster than the consumers of Peets coffee. The private label, which has a large assortment, but together with Dunkin Donuts and SBC, do not have external outlets that allow consumer to self-sample, also have a significantly larger TF effect than Peets and Starbucks.

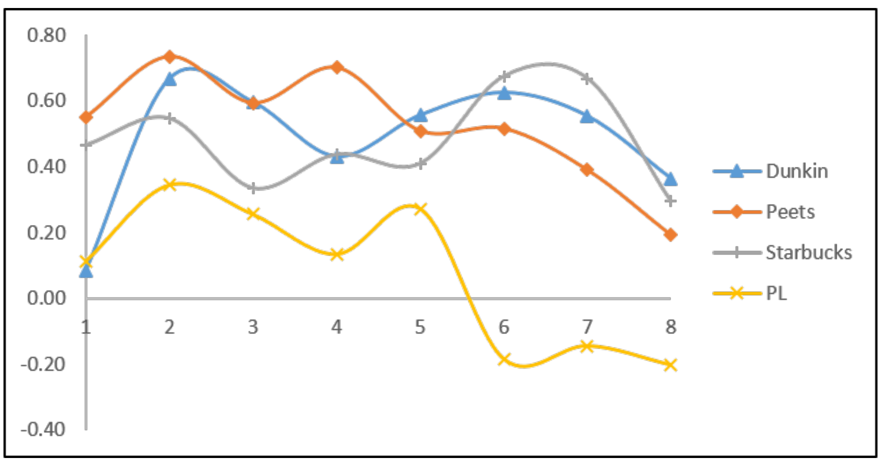
Identifying this effect allows the brands that have a lower brand equity, but large effect of the interaction of brand and the ever-trying or the forgetting effect to compete better in the market through pricing or other means. Moreover, the magnitude of the effect allows planning the budget that should be targeted at utilizing it as will be presented in the simulation part.

The log likelihood of this model is the maximal among the complete models and the estimation of the partial models that are presented below, however some of the effects, other than the interactions of the brands and the TF effects, are not significant in the complete model. In the appendix 1.7, the chapter considers other models, such as the standard model as it is in Berry, 1994, and a model that identifies trying and forgetting effect in an interaction

and uses a functional form. The latter model may be relevant to specific cases where the trying-event and the forgetting are not separately identifiable and is highly significant. [19]
The individual effects for the major attributes, which are summarized in table 1.5, are as expected. The mean of the decaf attribute coefficient is negative as most of the population preferred to buy regular coffee, however together with the standard deviation, we see that about 16% of the population prefers decaf, which is the market share of the decaf products. The major attributes are such that the mean is not significantly different than 0, however it does reflect the market share in the population. Hence, it seems that there is not enough variation between brand and main attribute for identifying the mean separately from the brand. Together with that, the standard deviation was significant.
Since the chapter uses this estimation as a baseline for the simulation, in which we calculate individual distributions as discussed in 1.5, the significance of the coefficients is not an obstacle. However, if the identification of the mean of the major attributes is of interest, a larger variation between brand and major attribute will be required.

Table 1.6 summarizes the fixed effects of the model and Figure 1.7 visualizes the results by graphing this estimation. We see that the quarterly fixed effects per brand generally explain the interactions of prices and market shares as was presented in Section 1.1.

Figure 1.7: Fixed Effects - Brand and Quarter - Trend Graph



1.5 Supply Simulation - Pricing Optimization

The discussion in 1.3 focused on the model and the strategy of the supply simulation. The demand estimation in 1.4, provided the parameters for sensitivity to pricing, ever-trying,

[19]This can probably be handled by increasing the sample size and number of simulation draws that are taken into the simulated multinomial maximum likelihood estimation that is used, which can be done through taking a larger sample of the dataset or using Monte Carlo method. In this case, both should and will be increased in the next versions of the chapter.

Table 1.5: Major Attributes Estimates

	Mean	Standard Deviation
Decaf	-1.24	2.05
	(0.14)	(0.08)
Major1	-1.15	1.57
	(296.52)	(0.08)
Major2	-1.06	1.41
	(296.51)	(0.08)
Major3	-1.63	1.79
	(296.51)	(0.11)
French	-0.56	1.65
	(296.51)	(0.06)
Major5	-1.99	2.1
	(296.52)	(0.14)
House-blend	-0.4	0.89
	(296.51)	(0.05)
Major7	-1.35	1.46
	(296.51)	(0.09)
Major8	-0.86	1.14
	(296.53)	(0.13)
Major9	-1.03	1.64
	(296.51)	(0.09)
Major10	-2.03	2.17
	(296.51)	(0.19)
Major11	-1.94	2.08
	(296.51)	(0.15)

Note: Standard errors in parentheses.

Table 1.6: Fixed Effects

	Q3 07	Q4 07	Q1 08	Q2 08	Q3 08	Q4 08	Q1 09	Q2 09
Dunkin	0.08	0.67	0.60	0.43	0.56	0.63	0.55	0.36
Peets	0.55	0.74	0.59	0.70	0.51	0.52	0.39	0.19
Starbucks	0.47	0.55	0.34	0.44	0.41	0.67	0.67	0.29
PL	0.11	0.35	0.26	0.13	0.27	-0.19	-0.15	-0.20

the forgetting effects, and others. These allowed separating major segments in the data and taking their specific distribution parameters and mean parameters as inputs to the supply simulation.

There are two reasons for defining segments in this case. The first is that it is not computationally possible to go through all the individuals in the dataset (and it is not generic to other problems to do so even if the dataset was small in this case), the other is that theory, such as Varian 1980, shows that to reach equilibrium, firms use a distribution that defines the pricing, which means that firms commonly target segments that gain highest utility from their products as this allow them to extract the highest surplus through avoiding Bertrand competition.

This is while the firms may launch a sale every few periods of time, to either remind, as discussed in Villas-Boas and Villas-Boas 2008, or to gain higher profits through higher demand, without other firms' punishment, as discussed in Varian 1980. More information why it is important to define segments is discussed in Perry, 2012 (based on the discussion in Revelt and Train, 2000).

δ is defined in Equation 1.14, however the demand estimation estimates, θ, that is the same for all the household in the data. The data and estimation provide the information to estimate the probability that a household made a specific set of choices $\mathbf{y_i}$ and the corresponding set of products' attributes per choice $\mathbf{x_i}$. This was defined by

$Prob(\mathbf{y_i}|\mathbf{x_i}, \theta) = \int Prob(\mathbf{y_i}|\mathbf{x_i}, \delta_\mathbf{i}) f(\delta_i|\theta) d\delta$

Applying the Byes rule we find the segments' distributions,

$h(\delta|\mathbf{y_i}, \mathbf{x_i}, \theta) \times Prob(\mathbf{y_i}|\mathbf{x_i}, \theta) = Prob(\mathbf{y_i}|\mathbf{x_i}, \delta) \times f(\delta|\theta)$

which is $h(\delta|\mathbf{y_i}|\mathbf{x_i}, \theta) = \frac{Prob(\mathbf{y_i}|\mathbf{x_i}, \delta) \times f(\delta|\theta)}{Prob(\mathbf{y_i}|\mathbf{x_i}, \theta)}$.

Finally the segments' distributions are estimated given equation 1.14.

$$\delta_{ij} = \int \delta h(\delta|\mathbf{y_i}, \mathbf{x_i}, \theta, \delta) d\delta \tag{1.14}$$

The simulation can be simulating a population with many segments as the computational requirements in this oblivious equilibrium solution are very low compared to the commonly used MPE. [20] For the purposes of the chapter, it was sufficient and efficient to select five major segments, following the concepts that were discussed in Varian, 1980 (and the chapters that followed it), were there are two target groups per brand, one is the most loyal customers who value the brand higher than the rest of the segments and the rest, who don't value any of the brands higher than the others. When defining the segments, full information was considered (TF-effects were added to the individual level brand effects) because in the simulation, the brands can take actions (reduce price) to induce purchase, which can increase the information level. The segments' attributes are summarized in Table 1.7. These segments have weights in the population, which are taken into consideration and are defined in the table as well.

As discussed in Section 1.5, the model assumes that the households in each segment are moving between discrete forgetting states. This is, if there are five forgetting levels, the forgetting that occurs over five periods of purchase by the segments is defined.

In each period, there is a probability that a consumer in a segment will purchase either of the products or the outside option. The probabilities are distributed on each segment grid according to the choices probabilities per segment. Each brand calculates the probability that each segment has chosen the brand in that period and the rest of the probabilities for

[20] In terms of complexity, assuming $T > n$, the OE is $O(T)$ and the MPE is $O(n^T)$

Table 1.7: Segments for Simulation

Segment		S1 – PL pref.		S2 – STBK pref.		S3 – Peets pref.		S4 – Dunkin pref.		S5 – General	
Weight		13.5%		15.3%		15.1%		13.6%		42.5%	
		mean	sd	mean	sd	mean	sd	mean	sd	mean	Sd
alpha		4.18	0.14	4.15	0.13	4.14	0.12	4.21	0.13	4.13	0.10
PL	Ever-Tried	1.44	0.12	1.22	0.09	1.22	0.09	1.23	0.08	1.29	0.13
	Frgt-Rate	-0.11	0.00	-0.11	0.00	-0.11	0.00	-0.11	0.00	-0.11	0.00
STBK	Ever-Tried	1.16	0.10	1.43	0.13	1.17	0.10	1.16	0.11	1.23	0.16
	Frgt-Rate	-0.07	0.00	-0.07	0.00	-0.07	0.00	-0.07	0.00	-0.07	0.00
Peets	Ever-Tried	1.10	0.19	1.12	0.16	1.61	0.27	1.09	0.20	1.22	0.30
	Frgt-Rate	-0.05	0.00	-0.05	0.00	-0.05	0.00	-0.05	0.00	-0.05	0.00
Dunkin	Ever-Tried	2.49	0.08	2.49	0.08	2.50	0.08	2.68	0.11	2.55	0.11
	Frgt-Rate	-0.44	0.03	-0.44	0.03	-0.44	0.03	-0.43	0.03	-0.44	0.03
House-blend	median	-0.60		0.00		-0.36		-0.49		-0.28	
alpha-normalize		-1		-1		-1		-1		-1	

the other brands, and it does that for all of the periods according to the assumed fixed pricing for the other brands. Accordingly, it optimizes the pricing strategy for all the periods. The states are not updated every period, but every four periods as an assumption.
Finally, for the simulation to be rational the variable costs per unit are assumed to be 0.6 of the average regular prices given the coffee-brands financial reports and the marketing research groups. This assumption is approximated and is not major in the model. Its main purpose is to reach similar pricing to the actual ones in the data. They do not make a great difference on the final results as long as they are around the ballpark of the actual costs. The regular pricing goals that are observed in the data and the assumed costs are summarized in table 1.8. [21]

Table 1.8: Average Pricing and Assumed Costs

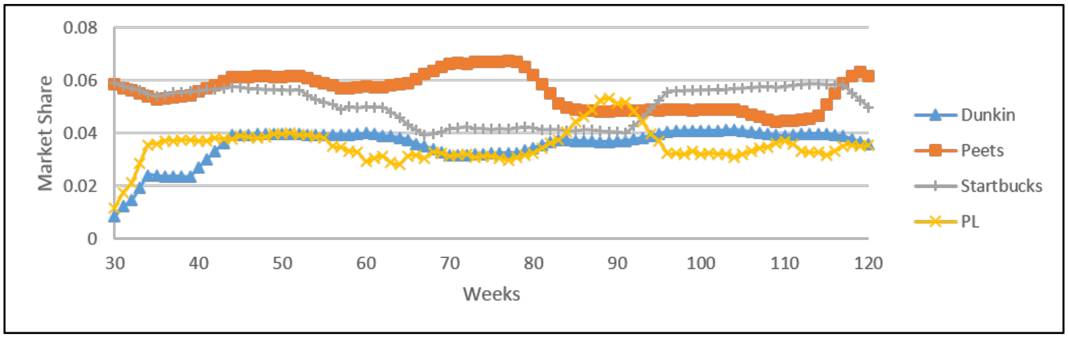
	Dunkin	Peets	Startbucks	PL
Regular Pricing Goal	0.85	0.925	0.925	0.775
Assumed Unit Costs	0.51	0.555	0.555	0.465

Simulation Results

As a baseline, the same pricing strategies of the data were used for the simulation. The result, as presented in figure **??** is that the market shares as they are in the data are predicted by the simulation (presented in figure 1.5b). This result provides a good indication that the demand model estimation makes a very good fit to the data and specifically that the simulation approximate the forgetting rate well. Given that the baseline for the data's

[21] Folgers financial report, 2008, directly presents gross margins per packaged coffee.

Figure 1.8: Shares Simulation given the Pricing in the Data



prices makes a good fit to the actual shares in the data, the pricing optimization was ran and the results are presented in Figure 1.9.

The optimized prices are continuous, however a pricing menu that is provided to the supermarkets is usually rounded to prices that can be used on the shelves. Figure 1.10 presents rounded pricing scheme, where unit prices are rounded to the 0.025 level. It shows that the unit prices suggested by the optimization range between 0.7 and 1, which are 12 price points. Together with that each brand's pricing strategy suggestion range differently and also change through the time.

The first quarter (third quarter of 2007) is the introduction quarter for the new products by Dunkin Donuts and the Private label brands, hence the pricing levels for the brands are not representative, but are meant to allow the stabilization in the post introduction period according to the equilibrium.

The results of the optimization show that the cyclic sales equilibrium predicted by the theory of trying, forgetting, and sales, e.g., Villas-Boas and Villas-Boas, 2008 holds for the estimated parameters, which is an important support to the validity of both the demand and supply models. As predicated the varied gaps between the prior and posterior together with the initial state and the forgetting rate, predict varied lengths and depths of sales cycles according to the ever-trying importance and the forgetting magnitude.

Figure 1.9: Pricing Simulation

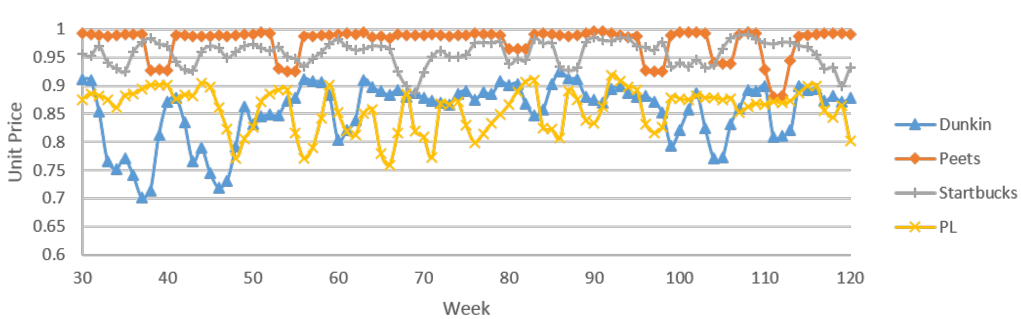


Figure 1.10: Pricing Simulation - Rounded Prices

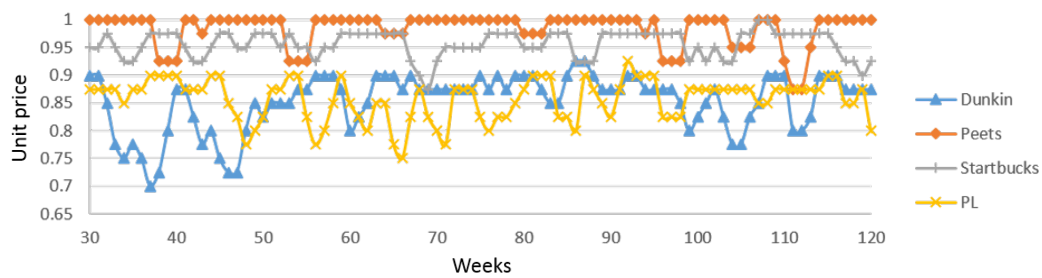
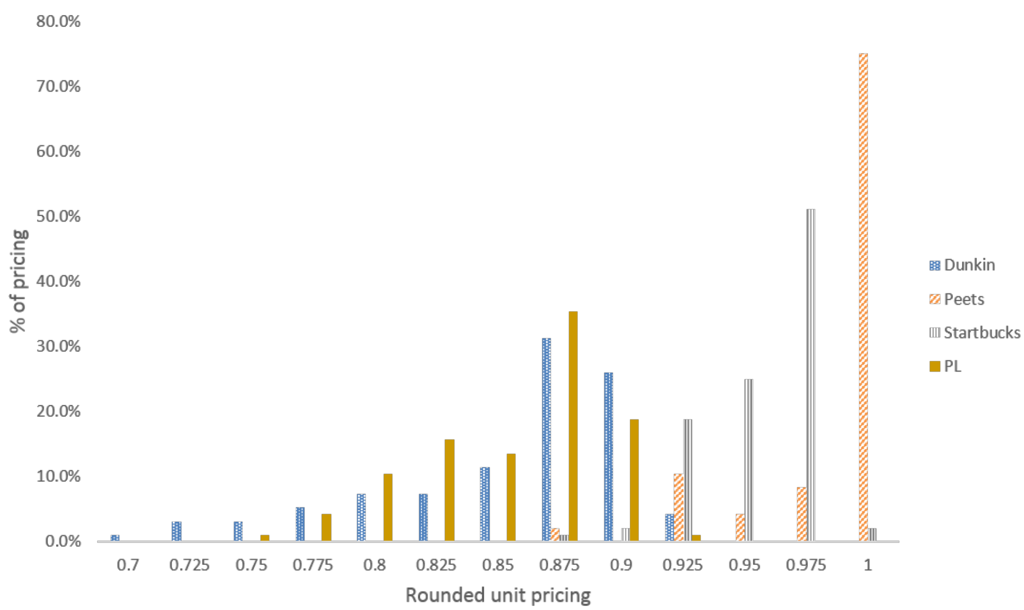


Figure 1.11: Pricing Levels

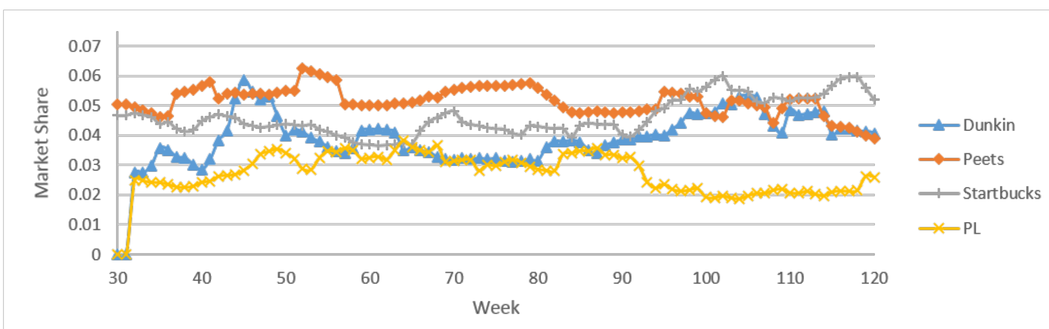


The optimized pricing strategy is aimed at increasing profits in the algorithm as the equilibrium is found when the firms do not want to deviate from their strategies as they cannot optimize their net present profits any further. This is achieved. As the theory predicts, this is done through the cyclic sales that overall increase market share. This is in line with the motivation of this chapter, which was to test whether Dunkin Donuts' shares could have been increased overall through optimized pricing scheme and whether the equilibrium can allow that to increase the shares to be comparable to the general share of the brand in the US market.

Figure 1.12 presents the predicted shares and table 1.9 shows that the pricing strategies that

are suggested raise the average market shares for Dunkin Donuts and the PL and decrease them for the bigger brands.
The assumption of the simulation is that if consumers do not buy from the sub-category their purchase is according to the market share of the rest of the sub-categories and the outside option. This is a rational assumption that solves the product-line optimization [22] without using the whole sub-categories data.
Overall the sub-category's shares are significantly increased. While this chapter is not aimed at extending the literature on market entry, it is nice to see that the result are generally in-line with the results that are discussed in Bresnahan, 1987 for the sub-category market share increase through the entrance of the two new leading brands to it.
Table 1.10 presents the change in profits through the varied leading products in the house-blend sub-category. It shows that overall the profits in the sub-category have increased, that three of the four brands are more profitable in this pricing strategy equilibrium and that Peet's coffee's profits have slightly decreased.
As discussed earlier, the owner of the prices' strategies is the category manager at the supermarket chain and the brands have strong influence on the decisions. This long-term pricing equilibrium, based on the oblivious equilibrium concept, which means that the brands are not calculating a Markov equilibrium, but commit to a long-term pricing strategy, given the results on the profits and shares at the sub-category, are possible to implement as they are overall very positive for the goals of the supermarket's category manager and the leading brands.

Figure 1.12: Predicated Shares for the Optimized Pricing Strategies



1.6 Conclusions and Discussion

This chapter estimates the demand side and simulates the optimized supply side of a novel structural model of trying and forgetting of products, finding evidence for both types of dynamics. By simulating the optimized pricing given the estimated effects, the chapter

[22]the full product line optimization equation as in Navo 2001 for example

Table 1.9: Average Shares - Data Compared to Optimization

		Dunkin	Peets	Startbucks	PL
Difference	Sample	$466	$758	$1,685	$323
	Northern California	$16,194	$26,316	$58,521	$11,226
Profit Change	Profit Change in the category	1.1%	1.1%	2.9%	0.9%

Table 1.10: Profits - Data Compared to Optimization

	Dunkin	Peets	Startbucks	PL	Total
Data	2.02%	6.37%	5.77%	1.47%	15.64%
Optimization	3.88%	5.10%	4.64%	2.78%	16.39%
Change	1.92	0.80	0.80	1.89	1.05

demonstrates that a model of trying and forgetting has important implications that cannot be obtained by the standard choice model, or models that include trying effects only.

The chapter contributes to the choice modeling literature, to the dynamic game simulation and estimation literature, and to the state dependence literature. The novel demand model has greatly increased the fit of the standard model, which was not significant without the trying-and-forgetting effects. It adds to the choice modeling literature on the one hand, on the other hand, it present a different perspective of inertia, which is that consumers are not irrational, but have varied information levels, which make them choose products that share major product attributes with previously chosen products.

Additionally, the chapter is among the first chapters in this literature to use the concept of oblivious equilibrium. Through a novel algorithm, it simulates and optimize prices in the market through 95 periods and finds the equilibrium solution to the dynamic games with multiple choices, multiple states per choice, and multiple segments. This was not feasible for previous models in this literature.

There are a number of extensions for this research that would be useful. First, this chapter shows that the forgetting coefficient is not zero and assumes that it is the same for all brands, estimating the trying effects only. It is interesting to estimate forgetting rates to the varied products. Additionally, it would be interesting to estimate the supply side in markets where cyclic sales are observed. The control function, following Revelt and Train, 2000, is used as a technique to treat the possible endogeneity, however it assumes a linear function. The BLP method or the control function method as in Villas-Boas and Winer, 1999, could be more robust. Finally, it would be highly interesting to study the causal relationship between the trying-forgetting rate and the presence of branded-stores around the supermarket as we see that there is a high correlation in the data and in the estimation.

1.7 Appendix

This chapter also considers various partial models, which continue to show that the TF effects are significant and make a better fit than the common model, which does not consider them. The next two models that were considered are a model that includes heterogeneity in tastes for brands, but does not include TF effect in Table 1.11, compared to a model that do not include heterogeneity in tastes for brands and includes heterogeneity in tastes for the interaction of brands and their TF-level in Table 1.12. The model that includes

Table 1.11: Heterogeneity in Liking Brands - Partial Information

	Mean	*Standard Deviation*
Price	0.04	0.55
	(2.18)	(0.05)
Residual	-0.26	-0.12
	(0.32)	(0.41)
Decaf-ind	-1.15	1.67
	(0.11)	(0.06)
Dunkin	-3.04	1.63
	(0)	(0.09)
Peets	-2.23	1.3
	(58.56)	(0.04)
SBC	-2.62	1.12
	(91.62)	(0.08)
Starbucks	-2.25	1.2
	(0)	(0.06)
PL	-2.52	1.13
	(0)	(0.05)
LLK	-38317.70	

Note: The LLK is the value of the log likelihood including the fixed terms. This table includes the major attributes' random effects and the fixed effects, which are presented in the appendix.

heterogeneity for the most important attributes in the products, which are the brands and the major attribute, such as House-blend, French, Colombian, and others, as defined in the products dictionary in Table **??**, in the appendix. This is the commonly used model, which does not take partial information on the brand equity due to TF effect into account. This model has the lowest log-likelihood and almost all the effects that are tested in it are insignificant. This gives a high motivation to consider the models that include TF effects in them for cases such as hedonic goods like coffee.

The model in Table 1.12, considers the case of heterogeneous TF effects with no heterogeneity. This estimation has the highest log-likelihood (except from the complete model), while it is also significant for all of the effects that are considered - both the random and the fixed

Table 1.12: TF Model - No Heterogeneity in Brands

	Mean	*Standard Deviation*
Price	3.39	0.59
	(2.2)	(0.04)
Residual	-0.28	0.94
	(0.32)	(0.62)
Dunkin X TF	2.40	0.81
	(0.21)	(0.33)
Peets X TF	1.36	0.89
	(0.07)	(0.07)
SBC X TF	2.43	0.90
	(0.14)	(0.17)
Starbucks X TF	1.39	0.46
	(0.07)	(0.13)
PL X TF	1.57	0.60
	(0.08)	(0.12)
Decaf-ind	-1.31	2.02
	(0.13)	(0.08)
LLK	-36436.46	

Note: The rest of the table is in Table **??** and in Table 1.13. The LLK is the value of the log likelihood including the fixed terms. This table includes the major attributes' random effects and the fixed effects, which are presented in the appendix. All of the estimates in this model are significant at the 0.001 level.

effects.
The difference in between this model compared to the model is that γ_{ijt} are assumed to be zeros and their effects are captured by the interactions of TF and the brands.
As discussed in section 1.2, in this dataset, 11 major attributes were defined through the data-refinement process. These products' sub-categories, such as vanilla flavor or house-blend, define segments as the consumers have heterogeneous tastes for these, given high variation of product assortment.
The random coefficients on the sub-categories of the TF model with no heterogeneous brand effects were all significant, while for the complete model, which includes the heterogeneous brand effects, they were not, as can be seen in Table **??**, which presents both models' estimation results.

Other partial models were examined in the estimation. One of them that is interesting as well is the standard model with the brand and major effects heterogenous together with the price sensitivity, while the TF-effects fixed, which is summarized in Table **??**.

Table 1.14 presents the results of the estimation of the standard model with fixed TF-effects, which was briefly discussed in 1.4.

Table 1.13: Fixed-Effects for the TF-Model

	Q3_07	Q4_07	Q1_08	Q2_08	Q3_08	Q4_08	Q1_09	Q2_09
Dunkin	-0.65***	-0.06***	-0.11***	-0.24***	-0.14***	-0.01***	-0.15***	-0.24***
Peets	-0.01***	0.18***	0.04***	0.15***	0.04***	0.08***	0.01***	-0.21***
SBC	-0.36***	-0.42***	-0.52***	-0.27***	-0.31***	-0.56***	-0.61***	-0.81***
Starbucks	-0.13***	-0.04***	-0.19***	-0.11***	-0.09***	0.16***	0.14***	-0.2***
PL	-0.52***	-0.29***	-0.38***	-0.44***	-0.35***	-0.72***	-0.66***	-0.72***

Table 1.14: Standard Model with TF as Fixed Effects

	Mean	*Standard Deviation*
Price	1.8	0.24
	(2.29)	(0.09)
Residual	-0.15	1.06
	(0.34)	(0.73)
Decaf-ind	-1.34	2.01
	(0.13)	(0.08)
Dunkin	-2.3	1.32
	(0)	(0.19)
Peets	-1.7	0.97
	(184.19)	(0.05)
SBC	-2.31	-1.01
	(0)	(0.11)
Starbucks	-1.76	0.87
	(0)	(0.06)
PL	-2.09	0.78
LLK	-36099.57	

Extended consideration of Inertia

Previous chapters have identified a lagged-choice effect, which is often referred to as habit persistence or inertia (e.g., Osborne, 2011). Together with that, in this chapter, 36.6% of the choices are repeated even though the coffee category there is a great variety of choice. These gave a strong reasoning to consider the case of a significant effect to inertia. An IV that could be very good was the previous gas prices, which as discussed in 1.4 and makes a great fit for the prices could explain the habit persistence in case previous time lower prices made a consumer purchase a product and then the habit has persisted even though the consumer does not specifically like the product better at time t. However, the fit for that IV on the parameter of potential inertia was very low, at less than 3% of the data, which gives a reason to believe that consumers that choose a product that they have chosen previously

Table 1.15: Inertia and Trying-Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DUNKIN'-Try-Forget	0.385***		0.387***	0.147***		0.226***	0.160***
	(0.016)		(0.017)	(0.044)		(0.041)	(0.040)
PEET'S-Try-Forget	0.125***		0.108***	0.060***		0.110***	0.081***
	(0.003)		(0.003)	(0.012)		(0.007)	(0.009)
SBC-Try-Forget	0.250***		0.251***	0.191***		0.203***	0.223***
	(0.012)		(0.012)	(0.036)		(0.027)	(0.031)
STARBUCKS-Try-Forget	0.102***		0.095***	0.058***		0.078***	0.083***
	(0.003)		(0.003)	(0.012)		(0.008)	(0.009)
PL-Try-Forget	0.145***		0.148***	0.063***		0.101***	0.085***
	(0.005)		(0.005)	(0.016)		(0.012)	(0.013)
DUNKIN'-brnd		0.074***	0.049***				
		(0.005)	(0.005)				
PEET'S-brnd		0.104***	0.067***				
		(0.003)	(0.003)				
SBC-brnd		0.069***	0.050***				
		(0.004)	(0.004)				
STARBUCKS-brnd		0.084***	0.058***				
		(0.003)	(0.003)				
PL-brnd		0.069***	0.047***				
		(0.003)	(0.003)				
DUNKIN'-Ever-tried				0.033	0.348***	0.160***	
				(0.046)	(0.015)	(0.037)	
PEET'S-Ever-tried				0.019**	0.108***	0.016*	
				(0.007)	(0.003)	(0.007)	
SBC-Ever-tried				0.044	0.208***	0.047*	
				(0.024)	(0.010)	(0.024)	
STARBUCKS-Ever-tried				0.025***	0.091***	0.025***	
				(0.007)	(0.003)	(0.007)	
PL-Ever-tried				0.032**	0.128***	0.044***	
				(0.012)	(0.005)	(0.011)	
DUNKIN'-inrt				0.222***			0.242***
				(0.049)			(0.039)
PEET'S-inrt				0.049***			0.047***
				(0.009)			(0.009)
SBC-inrt				0.016			0.029
				(0.032)			(0.031)
STARBUCKS-inrt				0.021*			0.021*
				(0.009)			(0.009)
PL-inrt				0.056***			0.065***
				(0.014)			(0.013)
Constant	0.051***	0.000	0.000	0.050***	0.050***	0.050***	0.051***
	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
	0.039	0.010	0.043	0.040	0.035	0.040	0.040
	866.071	210.323	475.007	297.318	780.117	438.638	442.870

have selected it because the expected utility from it at time t was more than the other options and not because of a habit persistence. In order to verify that the process of trying and forgetting is stronger than "last touch", a simple and standard logit choice models, e.g., Train 2003, were tested on these data looking at lagged choice effects, specifically separating the effect of ever trying a product (with ever-lasting effects), trying and gradual forgetting process, and next term post-purchase lagged-effect, which is commonly referred to as inertia when it is positively affects the choices. Table 1.15 shows the results, which are that the most significant effect is the process of trying forgetting (model number 3).

Chapter 2

Introduction of Product Attributes through Targeted Promotions

2.1 Overview

Prior theoretical literature shows that there are important set of cases where targeted marketing can be very effective in increasing firms profits. Targeting individual consumers with tailored offers has been widely discussed (for instance, Blattberg and Deighton 1991, Chen et al. 2001, Lederer and Hurter 1986, Montgomery 1997, Zhang and Krishnamurthi 2004, Iyer et al. 2005).

The main emphasis of this literature has been on evaluating how targeted marketing can facilitate the practice of price discrimination by firms. Together with that empirical literature has developed techniques to find customer specific preferences based on purchase histories through panel data, e.g., Rossi and Allenby 1993, Rossi et al. 1996, and Revelt and Train 2000. Moreover, literature shows how to find heterogenous elasticity of demand to products, which allows firms to realize the optimal pricing for their products within their categories, e.g., Berry et al (1995) henceforth BLP, McFadden and Train, 2000, Nevo, 2000, 2001, and BLP, 2004 (MicroBLP).

However, marketers, retailers, and policy makers are often interested that consumers will vary their consumption and thus they are interested in identifying the willingness to pay for specific attributes rather than for products.

Moreover, some segments may have low willingness to pay for specific products because the consumers of that segment derive low utility from a specific attribute in these products. This chapter examines ways to complement on the low willingness to pay through effective and efficient targeted promotions.

It extends the literature by focusing on a model and method to identify the difference between the average willingness to pay (WTP) for specific attributes and the specific groups of interest WTP. It specifically looks at the case of confronting the over-weight epidemic, which is a great interest of policy makers that has not been solved by the literature yet.

It does that by realizing the WTP for fat by consumers who consistently buy the fattiest products in the liquid milk category and showing that it significantly higher than the populations one. Moreover, the chapter presents the counterfactuals of providing targeted price promotions for that specific segments. This is by complementing the WTP for fat up to the next best option for the consumers using a targeted price promotion. The counterfactuals show that this method is highly effective and efficient.

Prior literature on confronting the Over-Weight Epidemic

The World Health Organization (WHO) formally recognized obesity as a global epidemic in 1997. Since then and even beforehand, legislators and researchers have been searching for effective ways to reduce overweight in the population. Nevertheless, no consensus about actual efficacy of any method was reached thus far.

Meanwhile, overweight and obesity have continued growing rapidly, such that in 2008, 35% of adults aged 20+ were overweight [1] and about 205 million men and 297 million women over the age of 20 who were obese [2], a total of more than half a billion adults worldwide. Overweight has become the fifth leading risk for global deaths since a significant portion of the burdens of diabetes, heart diseases, and certain cancers are attributable to it.

The major economic impact on the society includes direct medical costs as well as productivity costs, transportation costs, and human capital costs. The OECD has estimated that an obese person incurs 25% higher health expenditures than a person of normal weight in any given year and that obesity is responsible for 1-3% of total health expenditures in most OECD countries (5-10% in the United States and the UK). Figures 1.1 and 1.2 present the estimations and projections for overweight and obesity rates in the OECD countries till the year 2012 present the global rates of overweight and obesity, and as can be seen in them, the USA is leading in the overweight and obesity rates, with more than 70% (2009-2010) of adults over age 20 and about 32% of children and adolescents aged 2 through 19 years who are classified as overweight and obese [3] [4] England is following the USA and is very similar to it in its alarming rates of overweight.

According to the 2013 statistics from the NHS Health and Social Care Information Centre (HSCIC), in 2013, 65 percent of men and 58 percent of women in England were classified as either overweight or obese. This means that only one third of British men are considered to be normal weighted. Overweight is also affecting the country's youth. Nearly 1 in 10 children attending reception class in 2011-12, aged 4-5 years, were classed as obese. Three quarters of men and women are not consuming the recommended daily amount of fresh fruit

[1] Overweight is defined for people with BMI 25 kg/m2.

[2] Obesity is defined for people with BMI 30 kg/m2.

[3] CDC/NCHS, Health, United States, 2011, Table 74. Data from the National Health and Nutrition Examination Survey (NHANES).

[4] Data are from the National Health and Nutrition Examination Survey.

and vegetables, these figures are even higher among children [5] [6].
Since the fundamental cause of obesity and overweight is greatly associated with large intake of energy-dense foods that are high in fat, a major part of prior research of this topic was focusing on finding effective ways for convincing consumers to make healthier food and drink choices. The main question of whether governments can achieve desirable dietary goals through food price interventions was analyzed by prior literature in a few settings. An approach that was extensively studied is the usage of taxes for unhealthy properties of products; alas, there is no consensus regarding its efficacy. For example, Schroeter et al. (2007) created a microeconomic model to estimate the effects of a tax on high-calorie food. They have conducted an empirical analysis by obtaining statistics for price and income elasticities and using energy accounting to come up with weight elasticities. One of their findings was that a tax on high-calorie soft drinks would decrease soft drink consumption, which can positively affect a decrease in weight. Other researchers who have focused their studies on soft drinks have found similar results (Gustavsen 2005, Tefft 2006). Richards et al. (2004) used household scanner data in a random coefficient (mixed) logit model to test if rational addiction to food nutrients may be a cause of obesity. They found that a rational addiction to carbohydrates, fat, protein, and sodium exists and concluded that fat taxes may be more effective than information-based policies.
In the marketing literature, Khan et al. (2012), used a very large milk sales panel data from the US they have investigated whether price incentives induce substitution to healthier alternatives. They have used an exogenous variation in retail prices of milk, which provided a natural quasi-experiment, and then analyzed the impact of small price differences on substitution across fat content. Since they were looking to understand whether taxes can have positive effect on substitution to lower fat milks, the aggregated data was sufficient. They showed that when prices are uniform, whole milk has the highest market share at 36.4%. Under non-uniform prices, 2% milk is on average 14 cents cheaper than whole milk, and whole milk share falls to 29.7%. Under uniform prices, whole milk share for the lowest income quartile exceeds the highest income quartile by 17 percentage points. As the whole milk premium over lower-fat alternatives increases, whole milk share for both income groups falls, but the response is stronger for low income consumers. The discrepancy in market shares between income groups disappears with a whole milk premium of 15-20%. They have concluded that within-category demand for milk is highly elastic suggesting that taxes, if reflected in shelf prices, can serve as an effective mechanism to shift choices toward healthier options, particularly amongst lower income consumers. Other researchers were not as hopeful looking at additional categories and considering caveats in the taxes approach.
Kuchler et al. (2004) simulated health outcomes of a fat tax by using reduction in weight as a measure of health. They calculated the effects of a tax on different levels of consumer responsiveness to price. For each elasticity scenario, four possible tax rates ranging from 0.4 to 30 percent were considered. They were able to calculate reduction in caloric intake for

[5] latest statistics from the NHS Health and Social Care Information Centre (HSCIC)
[6] http://www.cdc.gov/nchs/fastats/overwt.htm

each scenario, assuming that nothing was substituted for the salty snacks and that all food purchases were consumed. From this they calculated reduction in body weight (3500 kcal per pound of body weight). Their results show that a small tax of 0.4 or 1 percent would not significantly affect consumption or health outcomes. In later work, the same authors further estimated demand functions for potato chips, all chips, and other salty snacks. Using the resulting elasticity estimates, they explored the effects of a 1, 10, and 20 percent tax on each snack category. They found that a small tax on salty snacks would not impact diet very much and that even a relatively large tax would not appreciably affect the diet quality of the average consumer (Kuchler et al. 2005).

In addition, Smed et al. (2005) have found that a tax on fats would decrease fat intake but increase sugar intake, while a tax on sugar would decrease sugar intake but increase fat intake. These tax scenarios predict a decrease in energy intake, which is a big problem because it is rational to assume that the energy levels that are required to the consumer will be complemented. Clark and Levedahl (2006) have estimated a demand-characteristic system for beef, pork, and poultry. According to their estimates, a tax that would increase the price of pork would increase the consumption of fat from pork and may contribute to obesity. In other words, a general usage of taxes without a well tested design not only will not help, but can even harm. Cash and Lacanilao (2007) have summarized the critique of imposing general taxes of fat and concluded that very little can be said on how diets would change under a regime of broadly imposed tax increases. They believe that ineffective and even perverse outcomes of such programs are not just theoretical possibilities, one main concern is that increasing prices for high-fat and sugar products will push the lower-income population to buy products that are even worse than the ones that they have intended to buy due to these products even lower prices (even after imposing taxes), for example convert from buying full-fat milk to dietary soda.

Another approach that was studied is the opposite of unhealthy-tax, the Thin Subsidies. Although such subsidies would require government outlays, governments are supporting healthier consumption initiatives, which require government spending, because of their great potential to reduce expenses through Medicare and other obesity-type problems taxation through decreased incidence of diet-related diseases, lessening the burden on the health care system. For example, Cash et al. (2005) estimated the health potential of thin subsidies, using epidemiological evidence on the efficacy of fruits and vegetables in reducing heart disease and stroke. They concluded that a thin subsidy could be an effective way to provide health benefits, especially to disadvantaged consumers. Their estimates of the cost per statistical life saved compare favorably with the costs associated with other U.S. government programs. In this chapter, I take a similar approach and estimate the overall spending and saving using the individual level price promotions. In the public health and dietetics literatures, French et al. (2001) used an experimental design to determine the effects of decreasing the price of low-fat snacks relative to regular snacks in vending machines. Four levels of pricing were examined. They found that a 10 percent decrease in price of low-fat snacks increased the percentage of snacks sold that were low-fat without increasing sales volume, which suggests that customers may have been substituting low-fat snacks for regular snacks. This is a positive result from a public health perspective.

Their results are especially interesting because they also present evidence that decreasing the price of low-fat snacks by 25 or 50 percent caused an increase in sales volume, which suggests that consumers may be buying more snacks or more consumers buying snacks from the vending machine, which could imply a negative net health outcome. Environmental interventions in a restaurant setting have yielded similar positive results (Horgen and Brownell, 2002). To conclude, the studies about the Thin Subsidy approach raise some optimism about an effective way to introduce consumers to new products within their consideration set, without a threat of greatly hurting a portion of the consumers.

Both taxes and subsidies are not trivial for implementation by governments; taxes are strongly rejected by firms. For instance, in response to plans for a tax on sugared beverages, Coca-Cola threatened to suspend investments planned in France and subsidies can be very costly if they are not carefully designed, thus the previous literature considers them in a theoretical or small-lab-experiment setting only. However, as mentioned earlier, recently, not only governments are looking for efficient ways to reduce overweight, but also firms and retailers are looking for ways to promote healthier consumption, following the consumers health-trend, which can be.

An example is the Public Health Responsibility Deal (RD) in England, which is a publicprivate partnership organized around a series of voluntary agreements that aims to bring together government, academic experts, and commercial and voluntary organizations to meet public health objectives, which include food and alcohol consumption. Through the RD, businesses commit to voluntary pledges to undertake actions for a public health benefit. Since its launch in 2011, all the major grocery retailers, including Tesco, Sainsbury, Waitrose, and ASDA, have signed at least 7 pledges. An example for a pledge is the fruits and vegetables one, in signing which the retailers have committed to take actions to make fruit and or vegetables (including frozen, canned, dried) more affordable, for example through promotions or value ranges.

The Targeted Promotions Approach

Following the prior research, the main concern that is raised in this chapter is that general, macro, solutions are not very effective or efficient because there is heterogeneity of a few factors in the population. Firstly, trying to capture the willingness to pay for the unhealthy measures (UH-WTP), such as saturated-fat or sugar, and compensating for these using higher prices or discounts using the mean of the UH-WTP in the population, is not effective because it does not represent the UH-WTP of the group who is choosing the fattier products frequently and do not mix with healthier choices very often, which should be the group of consumers who are overweighed.

Thus, while fat-taxes may convert the healthier consumers to buy healthy products more often because their willingness to pay is around the populations mean, it may not create the required effect on the group who has a high willingness to pay for unhealthy products, and this is the group that should be targeted in order to increase overall welfare. Moreover, as was discussed earlier (e.g., Smed, Jensen, and Denver, 2005), there is also large heterogeneity in responses when consumers are affected to convert from one product to others.

Some consumers will continue purchasing the same basket, converting from the less-healthy product to a healthier product, but not making new unhealthy choices due to this conversion, but others may react differently, and the conversion might even make the overall choices to be less healthy. Finally, suppose one is converted to making healthier choices, there is a probability that the choice will be continued to next purchases because of change in taste thanks to the introduction of a new product or because of inertia (Dube, Histch, Rossi, 2010), recognizing change in preferences or inertia should be very useful to increase total welfare because taxes or subsidies can be relaxed if inertia exists.

This is something that is not addressed in the solutions that were discussed in the literature yet. Given the issues raised above, this chapter presents a new approach for confronting the overweight epidemic in a way that can be more efficient than the current macro solutions, using targeted promotions and messaging. This approach requires that consumers purchase at a brick and mortar or an online store and are identified through their membership cards or customer identification so that their historical baskets information is stored and can be analyzed by a system that will implement this suggested approach.

In fact, shopping using membership cards and online shopping is very common in most of the popular chains in the USA and in England, the countries which are most prone to overweight and obesity, according to eMarketer 65% of the consumers in the US belong to a rewards program and 90% of the cards holders are active members. The chapter is based on a few marketing streams that discuss targeted promotions. The theoretical baseline is discussed in chapters like Blattberg and Deighton 1991, Chen et al. 2001, Lederer and Hurter 1986, Montgomery 1997, Zhang and Krishnamurthi 2004, Iyer et al. 2005, which evaluates how targeted marketing can facilitate the practice of price discrimination by firms. The empirical literature on targeted marketing which uses purchase histories, panel data, or other types of consumer behavior as a basis for targeting includes, Blattberg and Deighton 1991, Bult and Wansbeek 1995, Rossi and Allenby 1993, Rossi et al. 1996, and Revelt and Train 2000.

This chapter mainly follows the concepts that are discussed in Revelt and Train 2000 and Train, 2003. Two streams of literature are relevant in the implementation of the approach. The baseline model is the structural static demand models with respect to consumer heterogeneity in preferences. In terms of static demand models, e.g., BLP (1995), McFadden and Train, 2000, Nevo (2000, 2001), and BLP (2004 - MicroBLP) who has modeled static consumer decisions for differentiated products systems with heterogeneous consumers. These chapters and those that have followed show that in order to obtain predictions for differentiated products, it is important to incorporate consumer heterogeneity in the demand systems. Petrin and Train (2010) have developed a methodology for estimating the baseline model is using the Control Function approach. Using the chosen baseline model, per category, depending on whether there is an endogeneity in prices and the assumptions that can be made on the distribution in tastes for the attributes of the products; we reveal the distribution of the willingness to pay per category. Following this baseline estimation, it is required to estimate the coefficients for the group of consumers who have consistently chosen the unhealthiest products. For doing so, the approach in this chapter follows Revelt and Train (2000) who use a Bayesian method to identify the individuals willingness to pay and the targeted groups distribution. The consideration set is category based. There are other ways to define the consideration set for a consumer, but there is no consensus about this (Stigler, 1961; Howard and Sheth, 1969; Hoyer, 1984; Mehta, Rajiv and Srinivasan, 2003; Pancras, 2010; Draganska and Klapper, 2010; Pires, 2012).

Taking the category based consideration set is very common in the literature of demand estimation (Berry, Lehvinshon, and Pakes, 1995; Nevo 2000; and more) because that it is rational to assume that a consumer derive a certain benefit from a set of attributes that are common to the products in a category and thus scanner-data information reveal choice of one of the products in the category, which is maximize the utility from the choice set for that consumer. For effectively confronting the overweight and obesity epidemic, it is required to implement the individual-based approach that is presented in this chapter in each of the main categories that are known to effect over-weight, at least to include dairy products categories, such as butter.

Together with that, in this chapter, I am focusing on one category to present the modeling and consideration process, which can be almost replicable to many other categories within a retailer store. Using individual level scanner data, which were collected on all consumers who have purchased from the three biggest online grocery retailers in the UK through a unique website, this chapter estimates the individual demand for the fresh-milk category.

The choice of the fresh milk category follows USDA guidelines 2000 and 2010 that suggest increased consumption of low fat milk, and generally due to the fact that when dairy products are consumed in their fattiest form they are known to be a major factor for obesity. In fact, dairy products are one of the largest sources of fat in U.S. consumers diets.

In addition, the type of fat contained in dairy products, the saturated fat, leads to specific health concerns. In fact, dairy products, including butter and margarine, contribute 16% of total fat, 28% of saturated fat, and 17% of cholesterol to the nations food supply (USDA 2007).

A prior empirical study of factors affecting demand for reduced-fat fluid-milk is the study that was established by Gould, 1996, which has compared choices for fluid-milk according to the fat-type on the individual level data using a panel data from the US, has shown that the whole-fat products price-elasticity is more elastic than the other types of fat milks. This notion makes much sense because low-income households tend to buy fattier products, but are more sensitive to price changes (John Kearney, 2010). Gould, 1996, methodology is different from this chapter by an important aspect, which is that it doesnt include consumers heterogeneity in tastes in the upcoming model. Moreover, the sample that was used is the survey is not precise as observed scanner data choices. Finally, I take the estimation a step further by estimating the individual demand of the consumers, so to be able to give strategy recommendations and understand the counterfactual implications.

The data that are used for the estimation is very unique and are firstly used by an academic chapter. The data was obtained from an online website, which has provided individual level data on each grocery transaction made by consumers that used the website for their grocery shopping. They include eight weeks of scanner data that were collected on 50,000 consumers who have purchased from the 3 biggest online grocery retailers in the UK, Tesco, Sainsburys, and ASDA. These online stores, according to BMI, hold more than 65% of the online grocery market in the UK. Overall there are 6,500 baskets that were analyzed in 22 markets.

Each market has offered a different variety of products, with some intersections in the products offering, but generally, most of the products that were offered per retailer were of their private labels. Using the analysis, I investigate what price incentives, promotion framing, and introduction of new product attribute can be applied to induce substitution to fewer alternatives. Through the sample period, 80% of the consumers in the analysis were loyal to one retailer. Thus, retailer effects did enter the model. Together with that, since the households shopped at three different weeks out of the 8 weeks, time effects can enter the model. In the estimation process, after showing that the prices in this category are exogenous given the brands and the other attributes, I follow Train (2003) and estimate a maximum likelihood model with random coefficients that accounts for the heterogeneity in tastes as a baseline and then I follow Revelt and Train (2010) to reveal the distribution of the willingness to pay for the whole fat milk compared to the other types of milk, for the group of consumers that have purchased full fat types of milks in more than 50% of their purchases.

The chapter makes a distinction regarding families with young babies, who are recommended to use the fattiest milk. Recognizing the exact difference in the willingness-to-pay between a product and the next best product and subsidizing the other product accordingly should convert an individual to buy the lower priced product, but should not expend consumption as the compensation is exact. This is if the consideration is only for a specific category. The small lab experiment by French et al. (2001), which was described above was discussing a main result, taking the snacks machine as the consideration set when certain healthier products were discounted (subsidized), consumers have converted to buy the healthier products, but the consumption did not increase, i.e., they did not buy more because of the discount.

Since subsidizing is costly, directing it to the specific group of unhealthy eaters should be much more efficient than directing to everyone. Personalized promotions are legal and are frequently used, even though consumers can opt out of these. For many years electronic distribution of personalized coupons is widespread under programs, such as Catalina Marketing Incorporated (CMI) checkout coupons and mobile app coupons. In this chapter I show that targeting individuals who make consistent choices of the unhealthiest products per category is very effective and also efficient as the effect is targeted at the specific problem that needs to be solved. The consumers that mix between healthier and less healthy products even when there are no pricing differences between the products are consumers that should not be approached using price manipulation because they dont have a preference for unhealthy products through their choices. There is no reason to manipulate their choices in order to introduce them to healthier consumption. On the other hand, consumers who consistently purchase the unhealthiest products should be more prone to overweight if they consume these products.

Table 2.1 presents a supportive evidence for this using analysis of the dataset, which is discussed below. It shows that the group that buy the fattiest milks more than 50% of the times

Table 2.1: The Correlation between Choosing more Fat Milk and more Snacks

	(1)	(2)
VARIABLES	Number of snack (consumer level)	Number of snack (basket level)
Fat Milk Purchase Ratio >50% (customer level)	0.789***	
	(-0.289)	
Fat Milk Purchase Ratio > 50% (basket level)		0.246***
		(-0.0755)
Constant	9.119***	2.517***
	(-0.12)	(-0.0321)
Observations	13,275	13,275
R-squared	0.001	0.001

buy significantly more snacks on average. For every additional 1.28% choosing fat-milk, the consumers bought an additional snack throughout the sample period.

The results emphasize the importance of recognizing the targeted group's distribution compared to the populations one. They show that there is a significant difference in willingness to pay for full-fat milk compared to semi-skimmed milk among the general population compared to the targeted group. Using the results for the frequently unhealthy products consumers group the chapter presents a precise strategy for introducing the consumers to the lower fat milks and estimates the impact of implementing the complete system in retailers stores on the overweight epidemic and the economic implications of it.

2.2 Model

A random utility function is set up, where each transaction at the grocery store is made by a household to maximize the utility from the consumption given the budget and storage constraints and both the product attributes as well as a random term are assumed to enter linearly (e.g., e.g. McFadden 1974, and Train 2002). Each household likes to buy some amount of a product from a category from different types of the product. In the fresh-milk, for example, 4 ml of skimmed milk of one brand and 2ml of whole-fat milk of another brand, the first is needed to add to a morning cereal and the second in order to add to the coffee. Another example is in yogurts, a Greek yogurt, 2% for common consumption, and an assorted yogurt with chocolate flakes to indulge ourselves once in a while.

When considering a set of products from a category, each household faces a set of bundles of brand and units from that brand. Following the previous example, if the household would like to purchase 4 ml skimmed milk, the representative consumers can buy 2 bottles of 2 ml, 1 bottle of 4 ml, or 4 bottles of 1ml. The households make a decision based on the available prices according to the promotions, taste for package size, preference for number of products in the basket, and preference for a brand, in order to maximize their utility from the purchase, and whether they have a baby at the house, which may mean that they need some amount of whole-fat milk.

Furthermore, it is assumed that when a household mixes a few brands or a few types of products, the only reason to do so is for answering different needs from the milk. If it is not for different usages or different consumers in the household (which I define as different usage), a household's representative should choose the one type of milk that maximizes the household utility and buy more of it instead of mixing.

If the budget is limited, then the households might purchase as many units as possible from the preferred type and the rest from the second preferred one. This is as long as they have enough units of milk for their needs. Hence, when building the consideration set, it is possible to assume that there are options for the consumer to buy a few units from each product, but an household is not considering options of type 1 unit of X and 1 unit of Y as a bundle versus 2 units of X. Thus, if a household has purchased 2 units of X and 1 unit of Y, it is assumed that this is for different needs that are maximized, and I separate this choice into two different choices between bundles. This specification and setup of the data keeps the single-unit purchase assumption of multinomial logit that I use in the following model. Other specifications are also plausible, such as the one that is described in Hendel (1999), Dube (2004), and Chan (2006), which provide models for multiple products purchasing with variety. They fit the snacks and soda category better as a baseline model. Even though these models were considered, since fluid milk is not a storable item, and since it doesnt have a taste that a consumer might want to vary, the single-unit purchase of a bundle seems more plausible.

The utility below represents the utility that the household gets from the consumption of the chosen bundle. This utility function is similar to the ones that were defined in Nevo (2000) and Train (2003) for discrete choice models with heterogeneity.

The primitives of the model are the product characteristics and the consumer preferences.

It assumed that households buy one product from a choice set S with J + 1 alternatives in each time period. NP denotes the number of households in the panel and NM as the total number of markets in which the panel data's households made choices.

The consumers in a household derive utility from one product in that choice set (buying alternative j = 0 means that the consumer made no purchase of the available brands.)

The household buys the product for which the perceived utility is the greatest, but will make no purchase from the category if the utility of each of the brands is less than the utility of not purchasing.

The utility that consumers in household i at market t [7] obtain from product $j \in 1..J$ depends on observed and unobserved attributes of the product, where for j=0, $U_{i0m} = 0$. Assume that utility takes the form

$$U_{ijt} = -\alpha_i p_{jt} + \mathbf{X_j} \beta_{\mathbf{ijt}} + \epsilon_{ijt} \tag{2.1}$$

Where α_i is the mean price sensitivity per market t, p_{jt} is the price of product j at market t, $\mathbf{X_j}$ are product j's attributes, which do not change along the time, β_{ijt} is a vector of tastes per attribute of household i per attribute of product, j, and ϵ_{ijt} is the household, product, and market, iid demand shock, at time t.
Eight weeks were sampled for the analysis in this chapter, which is a short time for changing individual tastes, thus it can be assumed that the tastes are constant over this period of time. The next steps depend on whether the price or other parts of the utility are endogenous to the shock term.
As will be shown in the data section below, the price in the fresh-milk category is almost fully explained by the observable attributes of the products, including the brands. Thus, in order to establish main results, the main model that was selected is the simulated maximum likelihood without handling endogeneity.
The individual level value is a measure of the interaction between the attribute of the product, such as price, the product of the sensitivity (taste) for the attribute with the individual taste for the attribute and finally, the interactions between the attribute and the consumers individual characteristics, denoted by D. Using the assumption that the distribution of the individual random shock is extreme value type 1 (the logit assumption).
To ease notation, denote: $V_{ijt} = -\alpha_i p_{jt} + \mathbf{X_j} \beta_{\mathbf{ij}} + \xi_{jt}$
Then: $U_{ijt} = V_{ijt} + \epsilon_{ijt}$
A simplifying assumption commonly made (see McFadden, 1981, BLP 1995, and more) is that ϵ_{ijt}, which is distributed i.i.d., is of Type I extreme value distribution. Moreover, in this model, the mean utility of the outside good is not identified, and it normalized to 0.
Given these and the previous definitions, the likelihood that a household will purchase product j in market t is defined as follows

$$Prob(y_{ijt} = 1|\theta) = \int \frac{\exp(V_{ijt})}{1 + \sum_{k \in J} \exp(V_{ikt})} df(\cdot|\theta) d\eta_{ijt} \tag{2.2}$$

Where $y_{ijt} = 1$ if household i has chosen product j in market t given the parameters θ, which are the parameters of distribution F, from which the house-level tastes and preferences are drawn.

[7] for ease of notation and description, the model uses t, which can stand for time if the markets are defined by time only, however in this case, they are defined by both time and store.

2.3 Individual Distribution

The baseline model reveals the standard errors in the population, and using these standard errors we can estimate the individuals' coefficients. But a more precise estimation can be made using past consumption characteristics because the actual individuals choices reveal something about her tastes, and this should be discovered and understood in order to introduce new products to consumers with high preference to fattier products.

In the next parts I discuss the results and show that generally, full-fat milk is not highly valued by consumers, hence taking the general population means and standard deviation in order to find the willingness to pay for full-fat milk is very imprecise. The willingness to pay will be very small (if not negative), and we will not find a way to convert the 23% consumers who buy the fattier products.

Thus, the chapter follows Revelt and Train's (2000), which was described earlier. This allows the understanding of how to address the households that consume the less healthy products that are major causes of overweight consistently. In other words, I want to find where in the distribution of tastes the households which consistently buy the less healthy products lay. The main goal is finding whether there is a significant difference in the distribution, which allows defining a specific treatment to this group of individuals.

I denote the individual random-coefficients as matrix δ. The distribution of δ in the population of all people is denoted by $g(\delta|\theta)$, where θ are the parameters of this distribution, such as the mean and variance. A choice situation consists of several alternatives described collectively by variables X.

Initially, everyone in the population faces the same choice situations described by the same variables X. Some portion of the population chooses each alternative. Consider the people who choose alternative j. The tastes of these people are not all the same: there is a distribution of coefficients among these people. Let $h(\gamma|y_i, X, \theta)$ denote the distribution of δ in the subpopulation of people who, when faced with the choice situation described by variables X, would choose the full-fat alternative through the sequence of choices. $h(\gamma|y_i, X, \theta)$ is the distribution of *delta* in the subpopulation of people who would choose alternative j when facing a choice situation described by X. Since we observe each household's choices y_i when facing the same alternatives, we know that that person's coefficients are in the distribution $h(\gamma|y_i, X, \theta)$ and since h is tighter than g, we get more useful information about the person's tastes by conditioning on his past choices.

The approach that is used, applies to discrete choice models with continuous or discrete distributions of coefficients and uses maximum likelihood for estimation. The models of Kamakura and Russell (1989) and DeSarbo et al. (1995) are special cases of this method.

The relation between h and g can be established precisely as follows, we consider again the choice among alternatives $j = 1..J$ in choice situations $t = 1..T$. The utility that person i obtains from alternative j in situation t is defined in equation 2.1.

Now, we use the Bayesian approach, as follows: Denote by $y_i =< y_{i1}..y_{iT}$, the households sequence of chosen alternatives, where T is the number of transactions that this household made in the category. Define γ_i as follows

$$\begin{aligned} &let\ \gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}\ and\ \tilde{\gamma_{\mathbf{ij}}} = \begin{bmatrix} \tilde{\alpha_i} \\ \tilde{\beta_{\mathbf{ij}}} \end{bmatrix} \\ &Then \\ &\gamma_{\mathbf{ij}} = \gamma + \Sigma^{1/2}\tilde{\gamma_{ij}} + \Pi D_i \ \ s.t.\ \ \tilde{\gamma_{\mathbf{ij}}}\ \sim N(0, I_K) \end{aligned} \tag{2.3}$$

Where $\gamma_i\ h(\gamma|y_i, X, \theta)$
If we could tell what γ_i, the individual parameters, then the probability of the persons sequence of choices would be a product of the individual likelihood function, as follows

$$P(y_i|X_i, \gamma_i) = \prod_{t=1}^{NM}\prod_{j=1}^{J}\prod_{i=1}^{NP} Prob(y_{ijt} = 1|\theta) \tag{2.4}$$

Where $Prob(y_{ijt} = 1|\theta)$ is as defined in 2.2. Given these, the probability of household i's sequence is :$P(y_i|X_i, \theta) = \int P(y_i|X_i, \gamma_i)g(\gamma|\theta)d\gamma$
Applying the Byes rule we find the segments' distributions

$$\begin{aligned} &h(\gamma|\mathbf{y_i}|\mathbf{x_i}, \theta) \times Prob(\mathbf{y_i}|\mathbf{x_i}, \theta) = Prob(\mathbf{y_i}|\mathbf{x_i}, \gamma) \times g(\gamma|\theta) \\ &Which\ \ is \\ &h(\gamma|\mathbf{y_i}|\mathbf{x_i}, \theta) = \frac{Prob(\mathbf{y_i}|\mathbf{x_i}, \gamma) \times f(\gamma|\theta)}{Prob(\mathbf{y_i}|\mathbf{x_i}, \theta)} \end{aligned} \tag{2.5}$$

2.4 Data Analysis

For learning about the context of the data, the following section describes the market of fresh-milk and mass grocery retailing in the UK.

The market of Fresh Milk and the Mass Retailing in the UK

The fresh-milk that comes from cows is one of the unique categories in food in which the products are hardly differentiated as the tastes and the looks of the products are essentially the same across products. Together with that, this product category has high demand across populations in western countries such as the USA and the UK. This may be a good reason for many researches that use scanner data to choose fresh-milk for their empirical testing, at least as a first stage of the testing. The different products tend to vary according to the way they are produced and their fat content. The fat content of milk varies depending on the product, e.g. whole milk has a fat content of 4% fat, semi skimmed milk contains 1.7% fat, skimmed milk contains about 0.1% fat, and in addition there is 1% fat milk.
Additionally, much of the milk in the market is now homogenized as well as pasteurized, where homogenization results in milk of uniform composition or consistency and palatability

without removing or adding any constituents, and pasteurization helps to protect against any foodborne illness that can occur through consumption of raw (unpasteurized) milk. In order to provide the consumer with a consistent product, most milk is standardized, which creates a skim portion and a cream portion of the milk. Separation produces a skim portion that is less than 0.01% fat and a cream portion that is usually 40% fat. The cream portion is then added back to the skim portion to yield the desired fat content for the product. Common products are whole milk (3.25% fat), 2% and 1% fat milk, and skim milk (¡ 0.1% fat).

Semi skimmed milk is the most popular type of milk in the UK with a fat content of 1.7%, compared to a minimum of 3.5% in whole standardized milk and 0.1% in skimmed milk. The data reflects this information, with about 59% market share to the semi-skimmed type, 25% to the whole fat, and the rest to the skimmed types. Other than the fat contents, some of the most popular alternatives have additional attributes, which are whether they are filtered or organic. Organic milk comes from cows that have been grazed on pasture that has no chemical fertilizers, pesticides or agrochemicals used on it. Filtered milk goes through an extra, fine filtration system, which prevents souring bacteria from passing through. The nutritional content of the milk is unaffected but the shelf life is increased up to 45 days when stored at temperatures of up to 7C and an average 7 days once opened.

The Retailers

The UK grocery market is highly concentrated, with the four leading chains, Tesco, Sainsburys, Asda (owned by Wal-Mart of the US) and Morrisons, accounting for about 75% of grocery sales. The market will continue to face scrutiny from competition authorities, although recent investigations have tended to endorse the status quo, despite grumbling from suppliers, smaller retailers and some local communities. Tesco holds a 30% share of UKs online grocery shopping, followed by Asda (17%), Sainsburys (16%) and Morrisons (12%). The two hard-discount heavy weights, Aldi and Lidl (both Germany), hold a combined 6% share [8].

As discussed earlier, I am looking at consumption choices that are done at the biggest retailers Tesco, ASDA, and Sainsburys in the dataset. Altogether they have 63% market share of the grocery market. In the online grocery channel, they are even stronger. At the end of 2012, when the sample was taken, Tesco was the biggest online grocery retailer, while Sainsbury operates UK's second-largest online grocery business. ASDA was also investing heavily in its online presence. According to surveys undertaken by the British Retail Consortium (BRC), customers are increasingly checking food and drink prices on their mobile phones prior to purchasing. Indeed, online sales showed a strong performance in the first six months of 2012, driving the retail sector in the UK. Total online sales in the period reached US$54.75bn, compared with US$48.64bn spent in the same period in 2011. The report also showed that shoppers spent around US$9.42bn in June 2012, a 13% increase compared with

[8] BMI -market Overview - Retail - United Kingdom - Q1 2013

June 2011.
In 2011 19% of the mass grocery retail sales came from the hypermarkets, and in 2012, the online sales share was expected to reach 21% [9].

Data Sample

The data consists of individual level data on online grocery shopping. These include individual consumption transactions, product dictionary data with the product characteristics, and individual general buying patterns and characteristics. This dataset is panel dataset because it tracks individual purchases of its participating households across all three retailers, giving the consumers the same identification (in the estimation they will also get the same simulated tastes). Another aspect of the data is that we observe complete baskets, so we have information about the outside option and about whether it seems that there is a baby in household. The data also provide information about the households shopping habits at the website. The information includes the number of baskets that were sent to delivery since the first basket was sent, the total length of subscription of the households (time since the consumer has sent the first basket), and more.
The panel data is of 8 consecutive weeks, beginning May 1, 2012 and ending in June 25, 2012. The transactional data is from the UK, and they were obtained from an online grocery shopping provider (website service) which was operating for 6 years by May 2012, with an average of 10,000 baskets sent on a daily basis. The website allows the consumers to replicate the consumption behavior at a brick-and-mortar store, which includes: browsing between aisles, adding products to a basket, searching for products, etc. The customers that use the website are assigned to a default store at the beginning of their purchase and they can switch to a different one. Once they have chosen a store and start shopping, the shopping process is as if they are at an atomic retailer store out of the three options. Overall, the data that were analyzed have included 24 markets, which are 8 weeks for 3 retailers; each had over 120 milk transactions in it.
The complete data contains 56,500 distinct consumers, 96,000 baskets, and a total of 3.5 million observations. 2% of the products that are bought are fresh-milk, these are 74,000 scanner transactions. 57% of the baskets contained fresh-milk. 55% of the consumers who have used the website have also purchased fresh milk. While looking at the complete choice set that is described here has many benefits, in order to get consistent estimators of the parameters of the model, the data were filtered to include consumers who have purchased at least twice along their subscription period. Moreover, a consideration set was defined to be the bundles that were chosen by a consumer at least 5 times per market. The consideration sets differed between markets. On average there were 22 products in each consideration set. The dataset that was used for the main estimation included 7933 baskets for 3361 unique households, these were baskets that have contained fresh-milk from the consideration set, and the rest were removed for the final analysis. In another analysis, not only baskets from

[9] BMI - Market Overview - Retail - United Kingdom - Q1 2013

Table 2.2: Summary Statistics - Transactions

Transactions Variables	Mean	Std. Dev.	Min	Max
Total Transactions per market	387.32	155.9	29	610
Total Transactions per product	241.81	204.16	5	701
Total Transactions per household	4.05	3.53	1	21
Total Transactions per product per market	33.3	27.09	5	106
Total Transactions per household per market	1.48	0.72	1	6
Bundles in consideration set per market	21.75	5.52	12	30

Households Variables	Mean	Std. Dev.	Min	Max
Basket Expenditure per household	81.68	45.2	3.1	729.9
Average number of products in the basket	63	35	3	604
Subscription tenure	490	550	7	2112
Total number of baskets	18.67	32.6	2	269
Baby Indicator	35.68%			

Products Variables	Mean	Std. Dev.	Min	Max
Price for a bundle	1.7	0.85	0.5	6
Units in a bundle	1.5	0.8	1	6

the consideration set were included, but also the outside option choices, then the data included 16550 transactions and 6572 households.

The general information about the households demographics includes location, which is defined by region and area, there are 11 regions. 60% of the households representatives have indicated their age. 26% of the consumers probably have young babies since they have purchased products that are dedicated to babies, such as milk formulas. This indicator is important because when there is a baby in the house there is a large probability that there is a need for full-fat milk for the baby. The data also gives information about the households consumption behavior, which includes the average expenditure on a basket, the average number of products that are included in a basket, number of transactions from the first transaction till the end date of the sample data, and the tenure of time since they have started shopping through the website.

The panel data cover over 55 product categories; however, as was discussed earlier, I restrict the analysis to the fresh-milk category, also excluding buttermilk, flavored milk, and non-dairy alternatives to ensure comparisons of fairly homogeneous products. Even though fresh-milk is purchased very frequently, as was discussed earlier, the information contains high portion of outside options, which gives raise to an estimation consideration of whether to include outside-option in the parametric estimation. This will be discussed later in this section.

The summary statistics of the data is given in Table 2.2. The table presents information about the milk transactions, the households characteristics, and the milk products that are

Table 2.3: Variation in Final Price per Bundle

rounded price	Total Size								
	0.7	1.05	2.1	3.5	4.2	6.3	7	8.4	12.6
0.5	V	0	0	0	0	0	0	0	0
1	0	V	V	0	0	0	0	0	0
1.5	0	V	V	V	0	0	0	0	0
2	0	0	V	0	V	0	0	0	0
2.5	0	0	0	0	V	0	0	0	0
3	0	0	0	0	V	V	0	0	0
3.5	0	0	0	0	V	V	V	0	0
4	0	0	0	0	V	0	0	V	0
4.5	0	0	0	0	0	V	0	V	0
5	0	0	0	0	0	0	0	V	0
5.5	0	0	0	0	0	V	0	0	0
6	0	0	0	0	0	0	0	0	V

in the consideration set. Each market consists of between 29 and 610 transactions, with about 400 transactions on average, each product that is in the consideration set was chosen at least 5 times and up to 701 times along the sample period. The households have purchased frequently from the website, the number of purchases ranges from 1 to 21 times, and 4 times on average. Per market, each product in the consideration set was chosen between 5 to 106 times and 33 times on average and each house hold was purchasing between 1 and 6 times and 1.5 times on average. The households in the sample were spending an average of about 80 British pounds on the baskets, they were purchasing about 60 products each time, they were subscribed to the website for an average of 1 year and 4 months, but theyre tenure was ranging from 7 days and 6 years, and they have purchased between 2 and 250 baskets overall, with an average of about 20 baskets along their subscription period. About 35% of them are households with babies. This fact shows that there is a selection bias to this group of consumers who use the online channel as a major channel.

The group that we are interested in for the targeting purposes are consumers who have purchased the fattiest products, the products which are of full-fat, in more than 50% of their baskets. 26% of the consumers are in this group. Table 2.3 presents the information about the groups. As we can see the differences indicate that the average expenditure and the

Table 2.4: Price-promotion usage

m_Offer	ASDA	Sainsbury	Tesco	Total
2 for 1.7	0	0	3,284	3,284
2 for 2	26,482	1,429	405	28,316
2 for 3	1,895	4,872	986	7,753
Other	1,889	575	394	2,858
was 1.88	0	0	8,923	8,923
Total	30,266	6,876	13,992	51,134

ratio of average expenditure-to-units are significantly lower among the high frequency (UH-H) group. When the UH-H group was considered to be from 50% of the purchases and not higher than that, the differences werent significantly lower. The results for the UH-H group show that the data are consistent with the previously described surveys. As was discussed, the consumers who purchase less healthy products tend to have a lower expenditure rate per product.

Finally, the analysis of the variation in prices and units shows that price for a bundle is 1.7 with a variation between 0.49 and 6. On average there were 1.5 units of a product in a bundle and it has varied between 1 unit and 6 units. Table 2.3 presents a schematic map, which summarizes the variation prices and is presented as a crosstab of total-sizes and rounded-prices (up to 0.5). While the full prices of fresh milk did not vary much across the eight weeks of data within the retailer, the prices varied across retailers. Moreover, each retailer has launched promotions to varied products along these weeks, which have created a variation within the retailer in the total price that was required to be paid and across the retailers. The promotions were of two types: a price decrease for some of the products and an offer to buy two products for a lower price. Table 2.4 presents a summary of the promotions offered and the frequency of them being in the assumed consideration set of the consumers.

2.5 Estimation

The parameters to be estimated are the models' parameters, including the means for each of the available product characteristics. In the estimation on fresh-fluid-milk, these included the total price that was paid, the total units purchased, and the size of the units, the brand, the fat-type, whether the product was filtered, and whether it is organic. Together with the product characteristics, four household characteristics were defined, including, a dummy for region, a dummy for whether there are indicators for baby at home, the log of the subscription-length, the log of number-of-sent-baskets till date, and the log of the ratio of average-total-expenditure and the average number-of-products in the basket, which is a proxy for the willingness to pay for products by the household.

In prior literature, demographics, such as income, were used as a proxy for sensitivity to pricing, but this does not need to be the case. More precise characteristics can be defined when individual observed choices are revealed. Behavior indicators can be used, such as the actual expenditure per product, which is the ratio that is used in this model's estimation. As discussed in the introduction, overweight and obesity tend to be more common in low income families.

Hence, there should be a negative correlation between average expenditure on a product and the consumption of whole fat products. I assume that this heterogeneity in income exists even in the selected households which consume online and the estimation of the model is consistent with this as will be described later in this chapter. Another example that allows revealing information about the household is the baby indicator, which controls for whether whole-fat products are required in the basket for a young baby in the household. Using the revealed choices, it is possible to identify whether there is an infant in the household. The indicator checks if the household has purchased a product that is categorized as a baby product throughout all the purchases that have been made by the same consumer in the sampled data. For example, in this dataset, 30% of the consumers have purchased a product that is dedicated to babies in the sampled period, and 33% of the consumers who have purchased milk had purchased products from the babies category, such as diapers and milk formulas.

As mentioned in the previous section, the first step in defining the estimation strategy was checking whether prices are endogenous in the utility model (equation 1). In order to learn about the endogeneity I have checked whether the brands of the products together with the rest of the observed characteristics have a very high explanatory power to the prices, and this was indeed the case as can be learned from 2.5. The table presents four models.

In the first one the final price for the bundle, given the promotions, was regressed on the number of units of the product and the brands and I find that the fit is almost perfect. It was not even necessary to add information about the other attributes of the products. The rest of the models, (2)-(4), were used for learning whether it is possible to use the fuel price as an instrumental variable. This has shown to be the case, and I will elaborate on this in the robustness checks part.

Thus, given the brands and the households' characteristics, the prices are exogenous to the shock term, thus it is possible to estimate the model using the maximum likelihood procedure.

The estimation is done using both simple multi-choice logit maximum likelihood procedure and a multi-choice simulated ML procedure, which simulates the households individual tastes. For this, a few distribution specifications were examined, and the normal distribution for all the parameters was the most reasonable. It is common to use log-normal to prices and to other parameters that are considered always negative, but this is not always the case. For products that have are low cost, such as fresh-milk, this may not be the case, as the price may be a signal for quality for the individual, a signal that cannot be instrumented through nothing but taste. As can be seen in the model, the parameters for the brands and for the

Table 2.5: The Bundle Price Explained by the Observed Characteristics

	(1)	(1)
	FullPricePaid	FullPricePaid
Units in BSKT		0.931***
		(-0.000786)
Size		0.364***
		(-0.000749)
Full Fat/Whole		0.182***
		(-0.00356)
Semi Skimmed		0.207***
		(-0.00351)
Skimmed		0.182***
		(-0.00363)
FilteredID		0.0248***
		(-0.00163)
Organic==Yes		-0.294***
		(-0.00796)
ASDA		-0.481***
		(-0.00591)
ASDA Organic		0.524***
		(-0.00661)
Cravendale		0.0798***
		(-0.00618)
Creamfields		-0.613***
		(-0.00603)
Sainsbury's		-0.320***
		(-0.00606)
Sainsbury's Organic		0.449***
		(-0.0067)
Tesco		-0.328***
		(-0.00593)
Tesco Organic		0.468***
		(-0.00613)
Exp. rtio times Bundle-price		0.0346***
		(-0.000667)
Tenure times Bundle-Price		0.0193***
		(-0.000174)
Total-Sent BSKTs times Bundle-Price		-0.00705***
		(-0.000219)
IV fuel 1	-0.0608**	0.0754***
	(-0.027)	(-0.00447)
IV fuel 3	-0.112***	0.0806***
	(-0.0258)	(-0.00428)
IV fuel 4	-0.136***	0.114***
	(-0.0242)	(-0.00406)
IV fuel 5	-0.240***	0.0910***
	(-0.0237)	(-0.00396)
IV fuel 6	-0.299***	0.0460***
	(-0.0237)	(-0.00397)
IV fuel 7	-0.0710***	0.0631***
	(-0.0237)	(-0.00398)
IV fuel 8	0.191***	0.123***
	(-0.0234)	(-0.00393)
Constant	1.946***	-0.465***
	(-0.00937)	(-0.00695)
Prob > F	0	0
Observations	187,897	187,897
R-squared	0.004	0.973

demographics do not depend on distribution in tastes according to the defined model (even though, theoretically, brands can have heterogeneity in tastes).
Given the previous analysis, the model is adjusted as follows,

$$Prob(y_{ijt} = 1|\theta) = \int \frac{\exp(V_{ijt})}{1 + \sum_{k \in J} \exp(V_{ikt})} d\phi(\cdot|\theta) d\eta_{ijt} \tag{2.6}$$

Where θ is a vector of the means and the standard errors. As a baseline, I assume that there is no covariance between the tastes. This is in-line with the model specification.
to numerically estimate the integral that expresses the individual probability in equation 2.6, draws were taken for the individual tastes for 8 product characteristics, which include the price that was requested for the bundle, the units in the bundle, the size of the packages, indicators for fat-type: full-fat, semi-skimmed, skimmed, 1%, no-fat, indicators for whether the milk was filtered and for whether the milk is organic. McFadden and Train, 2000, discuss the consistency of estimation using sampling. In this case, since the customers are returning customers who shopped at the website at least twice, the analogy principle applies for estimating the integrals in the models using the means of the expression that is used for the simulated log likelihood both per times and per consumers. The analogy principle will lead to consistent estimation since the number of repeated purchases, T (for times), is not bounded and eventually T can be assumed to be endless. We are interested in the willingness to pay for the fat compared to the next best fat option per individual. The difference individual parameter is defined as follows:

$$\Delta WTP_i = WTP_{FF,i} - WTP_{NB,i} \tag{2.7}$$

Where WTP_i is the willingness to pay for the product by household i, FF means full-fat milk and NB is the next best type of milk that was estimated for household i.
The final step of the estimation is finding the individual parameters, which is done using the results from the estimation of 2.6, which has estimated $\hat{\theta}$.
Given the repeated choices of the consumers, the individual γ_i were estimated. Their distribution is given by $\hat{h}(\gamma|y_i, X, \hat{\theta})$ as discussed earlier. Even though we are looking at individual level targeting, group level tastes are also of interest. This is achieved by a weighted average or an integral of the individual level tastes.

2.6 Results and Counterfactuals

By applying the estimation that was described above to the data, I was able to show that analyzing the general populations response to taxes and subsidies is not highly relevant to reducing overweight, which is how to introduce consumers who consistently make unhealthy choices to healthier choices. If the general populations elasticity of demand would have been considered, we could not have come up with any reasonable subsidy for the lower-fat products. The analysis that is described below proves the initial hypothesis, which is that using

Table 2.6: Main Parameters

		No-OOP	OOP
Price	Mean	-1.572***	-1.50***
		(-0.107)	(-0.102)
	Standard Dev.	0.268***	0.214***
		(-0.042)	(-0.031)
Unts In BSKT	Mean	1.162***	1.164***
		(-0.109)	(-0.103)
	Standard Dev.	0.353***	0.315***
		(-0.033)	(-0.026)
Size	Mean	0.788***	0.758***
		(-0.049)	(-0.046)
	Standard Dev.	0.012	-0.031
		(-0.039)	(-0.03)
Full-Fat Milk	Mean	-0.003***	0.362**
		(0)	(-0.017)
	Standard Dev.	0.692***	0.337**
		(-0.136)	(-0.113)
Semi-Skimmed Milk	Mean	1.053***	1.142***
		(0)	(-0.118)
	Standard Dev.	1.813***	1.419***
		(-0.077)	(-0.067)
Skimmed Milk	Mean	-0.219***	0.385**
		(0)	(-0.123)
	Standard Dev.	1.534***	1.066***
		(-0.097)	(-0.081)

individual targeting can create the required result while greatly increasing efficacy.
The main analysis used the data on choices from the consideration set only and was based on three estimations on the same set of variables. The estimations differed from each other in the number of simulation draws for tastes that were taken for estimate the integral that is defined in 2.7.
As was discussed in the estimation section, the baseline specification for the tastes distributions was the normal distribution for all of the variables, and this is the specification that was taken for these three estimations. The first three columns of Table 2.6 present the results of the main parameters.

We notice that as the number of simulations-draws is increased, the absolute value of the coefficient on price is slightly increasing and the standard coefficient estimator is increasing as well, which is consistent with the theory. The coefficient on price is negative and significant, as we would have expected. The parameters for units in the bundle are positive and almost exactly the opposite from the parameters for price.
This is consistent with the previous discussion about the correlation between price and the units in the bundle (see table 2.5 and the discussion about endogeneity). The rest of the coefficients' signs and magnitudes are as expected as well. A supportive analysis to the results is learning the share of the distribution that is above zero. Table 2.7 presents the summary of the estimation in terms of the distributions of the coefficients, given the normal

Table 2.7: Parameters Analysis

		Mean	StdDev	Share>0
Price	normal	-1.572***	0.268***	0
UntsInBSKT	normal	1.162***	0.353***	1
Size	normal	0.788***	0.012	1
Full-Fat Milk	normal	-0.003***	0.692***	0.5
Semi-Skimmed Milk	normal	1.053***	1.813***	0.79
Skimmed Milk	normal	-0.219***	1.534***	0.43
1% Fat Milk	normal	-0.831***	1.456***	0.21
FilteredID	normal	-2.477***	3.648***	0.16
OrganicXV3	normal	-8.241***	5.903***	0

assumption; the shares above 0 of the different milk types are consistent with the previous data analyses.

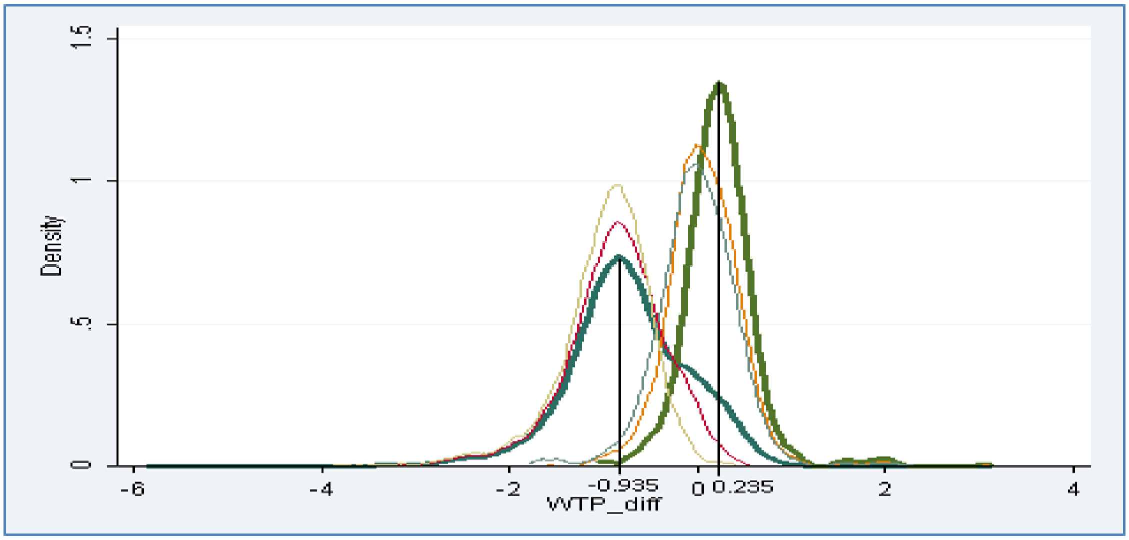
Following the baseline estimation, the individual coefficients were estimated using the procedures that were described in the previous section. This step uncovered the high frequency (UH-H) group's estimators, which have shown to be significantly different from the general population's ones. As described in Table 2.8, while the price coefficients do not differ between the groups, the fat-type coefficients significantly differ between them, such that the WTP-for-fat difference among them is of the magnitude 1.33, which is 78% of the common bundle's price. There is no analytic expression for the distributions of parameters of the groups (H and L), and thus the distributions were estimated using kernel estimations.

Eventually, looking at the results, we see that they have a normal distribution shape. Figure 2.1 presents the kernel distributions of the difference in WTP for fat and the WTP for the next best alternative for the individual. One line presents the distribution in the population and the other one presents the distribution in the high-frequency for fat, the H group. Two important findings are presented in this figure; the first is that while analyzing the general population's willingness to pay for fat was not useful, the high-frequency group's distribution of willingness to pay is very informative. The second is that the individuals' simulations of tastes are important in the estimation analysis and that convergence is not achieved immediately, but requires enough simulations. The main difference between the figures is that the number of simulations in the estimation has differed between them. In figure 1.13 we see the result that was described above. Essentially, if it is possible to target on an individual level, realizing what the next best option is, the difference in WTP is on average 0.26 British pounds, which is 15% of the average bundle price.

Table 2.8: Individual Parameters

	Ratio>0.5		Ratio<=0.5		Differences	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Price	-1.576	0.129	-1.573	0.135	-0.003	0.007
Units	1.161	0.173	1.153	0.186	0.008	0.009
Size	0.788	0.005	0.788	0.005	0	0
Fullfat	0.296	0.336	-0.054	0.311	0.35***	0.016
Semi skimmed	-0.467	0.708	1.359	1.087	-1.825***	0.053
Skimmed	-0.591	0.582	-0.151	0.869	-0.44***	0.043
1% milk	-0.892	0.719	-0.834	0.662	-0.058	0.034
Filtered	-2.472	2.555	-2.626	2.587	0.154	0.131
Organic	-8.628	2.697	-8.567	3.153	-0.061	0.157
Total expenditure	74.95	41.20	82.85	46.49	-7.90***	2.34
# of products	60.04	35.97	62.52	34.72	-2.48***	1.79
Subscription Tenure	403.57	493.48	497.01	554.85	-93.44***	27.99
Age	36.017	9.877	38.27	10.364	-2.253	0.524
Ln Expend. Tatio	0.24	0.313	0.281	0.304	-0.041	0.016
Ln tot sent bskts	1.853	1.171	2.004	1.24	-0.151	0.063
Fat milk ratio	0.964	0.107	0.068	0.164	0.896	0.008
WTP-Fat Milk (WTP_FM)	0.189	0.216	-0.035	0.202	0.224	0.01
WTP –Semi Skim.	-0.302	0.458	0.869	0.709	-1.171	0.035
WTP-Skim	-0.377	0.373	-0.096	0.567	-0.281	0.028
WTP-VLow-Skim	-0.568	0.453	-0.532	0.433	-0.035	0.022
Max – next best (NB)	-0.074	0.332	1.03	0.538	-1.104	0.026
WTP-diff (WTP_FM - NB)	0.235	0.39	-0.935	0.486	1.17	0.02

Figure 2.1: Chap2 Kernal Distributions of ΔWTP_i



Counterfactuals

As discussed earlier, 12% of the transactions were made by households which have purchased more than 50% full-fat milk in their baskets. Given the estimates, I have recalculated a quasi-counterfactual to learn whether the average discount per household in the UH-H group, given to the individual estimates, will indeed introduce these households to a new lower-fat product. The results were very supportive of the approach. Out of these 930 transactions, 501 were of household with WTP-diff that is higher than zero. Thus, the next-best products in the consideration sets of these transactions were discounted according to the estimated WTP-for-fat. The overall potential subsidy was very low, and it was 223 (British Pounds) only compared to the revenues from milk of over 13,500. Out of the consumers who have received a discount, 53% of the targeted transactions have converted to a discounted product as was expected because we have reached the indifferent condition, which has led to a final subsidy of 103 only.

2.7 Conclusions and Discussion

Policy makers, retailers, and marketers are often interested in familiarizing certain consumers with product-attributes that these consumers do not often try. Previous literature in targeted marketing have studied the willingness to pay (WTP) for varied products as bundles, and this chapter extends this literature by identifying it for certain attributes of products and specific segments.

Using a unique panel dataset, with individual transactions, it estimates the standard demand model. However, it also finds the individual distributions of households according to their consumption patterns looking at specific attributes. By that, this chapter presents a

new approach of how to target specific groups of interest according to their general liking of specific attributes across products.
In this chapter, the group of interest is the group that often purchases the fattiest products in the diary category and seldom purchase lower fat products. By allowing heterogeneity in tastes for varied attributes and then applying a Bayesian model on the actual choices, individual level coefficients are found.
The estimation results show that there are significant differences in the distribution of WTP for higher fat levels between the average in the population and specific groups of consumers who often purchase these fattiest products. The counterfactuals show that the required price promotions are very efficient and effective.
This chapter also presents a new approach for confronting the overweight epidemic using targeted subsidies. The results show that it may rather efficient to introduce consumers from the high-frequency group (UH-H) to healthier choices using promotions for the next best choice of fat-type. The average discount that is required to be made for the next-best-option is of 13.5% decrease in the bundles' prices, which is a 0.235 British Pounds decrease. The average expenditure per household, according to the data, we can assume that these households make a purchase once a week on average, thus the total yearly expenditure is around $5260 US dollars, but the data shows that only 60% of it is related to categories that contain saturated fat, sugars, and other unhealthy ingredients. Thus, the upper bound for the subsidies expenditure is Total*0.6*0.135, which is We are looking at Hence, the subsidy will be summed at $425 US dollars.
As was discussed in the introduction, an obese person incurs 25% higher health expenditures than a person of normal weight, taking the information about the public expenditure on health and finding the health expenditure on obese we find that in the UK, the extra expenditure on obese per capita is $692 US dollars. If this strategy is implemented as discussed above, the model suggests that it will greatly influence the households and thus, reduce overweight. But, since the model suggests that about 50% of the consumers will convert, even if it is possible to assume that this will be enough to reduce their weight over time, it means that the extra public expenditure will be summed at $350, which leaves $75 extra public spending for the subsidies. Together with this, as was discussed before, reducing overweight in the population have additional public health and economics advantages, which firstly include better lives for people, especially for children who do not have much choice and a led by their parents choices. The benefits also include higher productivity per capita, which will be reflected in income from higher taxes.
Alas, the analysis from above uses very rough estimations of the costs and saving, which should be greatly elaborated and improved. The main limitation of these data was that it was not very long and included 8 weeks only. It will be easy to extend these data and even add more information from brick-and-mortar stores that use membership cards. Then, as a first step, it will be possible to simulate an individual price reduction and check that the estimated parameters will indeed introduce the targeted consumers to the next best option. Secondly, it is possible to add correlation between choices to the model and estimate inertia parameters, in a similar way to Dube, Hitsch, and Rossi (2010). Using these with longer

dataset than the current 8 week, it is possible to estimate the individual inertia parameters, and adjust the promotions to this; meaning that if a household was introduced to a healthier option in the previous purchase, with a certain probability this household will buy the same product in the next purchase instance. This information can greatly reduce costs of subsidies, while bring to almost the same result overall. Lastly, for the robustness checks, even though the generalized methods of moments was built for this case, it was not fully estimated using the consideration set that was defined and another category should be added to the estimation to prove that this is not a private case.

Nevertheless, this chapter motivates continuing studying the new approach for reducing overweight and obesity as it shows that in implementing the plan the governments will not only treat the problem of overweight with higher probability than the previous macro solutions, but will also improve overall social welfare with a very low monetary cost or even overall saving, which gives an incentive for governments to subsidize discounts given by retailers according to the approach that was described in this chapter.

Thus, on the one hand, this chapter presents a complete method for introducing products to specific consumer groups, which is relevant to many marketers, and on the other hand it shows how targeted marketing can be used by policy makers for increasing consumers' and social welfare.

Chapter 3

Explaining the Sudden Price Loving Behavior in the US Recession

3.1 Overview

It is rare that the demand for durable goods' products is increased together with an increase of the index price of the products. However, when these cases are apparent from the data, they can point out on major effects that can highly affect lower-income populations. An example is the familiar case of demand for potetoes during the Irish Great Famine. [1] [2]
As the sub-prime crisis of 2008 hit the US and recession of 2008-2009 began, over 20% of the consumers (Segment 2) at a major supermarket chain has significantly increased the average unit price that they used to spend in the liquid laundry detergent category. [3] This is even though there was a significant price decrease at the category. This is exactly the opposite of the expectations. In fact, the rest of the consumers (segment 1) behaved according to the expectations and have purchased products with a significantly lower unit price, even if they had to change the products that they used to buy. As expected, using the standard demand model and adding treatment for taste heterogeneity, without treating endogeneity, it seemed that Segment 2 is not sensitive to price, and is even price loving, while the model explained the behavior of Segment 1 as expected.
Adding the summary statistics information that Segment 2 also significantly decreased the average package size and did not enjoy the quantity discounts that were offered in the category, while segment 1 did the exact opposite raised the question whether segment 2 became more limited on its borrowing ability and thus decrease quantities, so that the overall payment will be lower.
There is a very limited evidence for these situations occurring, however for the cases that

[1] Potatoes during the Irish Great Famine were long considered the most known example of a Giffen good.

[2] For some products, which are called Veblen goods, price paid has a positive utility to consumers, thus when price is higher the targeted consumers' demand increase. This chapter does not refer to these cases.

[3] This will be extended to more categories. For example, the coffee market went through a similar situation.

they do occur, the literature was highly interested to learn what were the effects that were driving the phenomena. While researched in micro settings, they can have large implications on the population's consumption patterns.
Sherwin Rosen, 1999, discusses the "Potato Paradox", when the increase of potatoes price during the Irish Great Famine actually increased demand of the lower income and more price sensitive population, which could have also be defined by a researcher that does not have the complete information a population that suddenly became less sensitive to price. While the prior literature to Rosen has discussed the potatoes as Giffen goods, Rosen, 1999, shows that the phenomenon could be explained by a standard demand model.
This chapter is extending this literature by presenting another evidence for a consumption situation that can initially be very confusing. In this case, it shows that in fact, the price decrease of the recession by the leading brand allowed segment 2 to be able to finally buy Tide's products as the product were less expensive than they used to be. [4]
Thus, the chapter extended the initial demand model to include treatment for prices endogeneity in the estimation, and using MLE and the control function approach, the model's estimation results show that the latter explanation for both segments' behavior is probably most accurate. [5]
This chapter concludes by a discussion of how to handle changes in consumer behavior for finding the exact preferences. It particularly highlights that when it seems that consumers' preferences have greatly changed post a major change in the market, it may be misleading. It emphasizes how price fluctuations and consumers behaviors significant changes provide a great opportunity for marketers to see choices that they could not observe in the data before and analyze consumer behavior with more information.

3.2 Data

The sample data is of 6% of all the transactions that took place between February 2008 and March 2009 at the Northern California stores of one of the biggest supermarket chains in the US in the laundry liquid detergent.
At the end of 2008, this category market was estimated at $3.6 billion sales for that year [6].
Given the sample size and the total purchase size per household in the retail stores that are considered in the sample, there were about 620 thousand regular customers who purchased every 2 months on average at an average total price of $12. Thus, the expected total yearly revenues for these stores is $7.4 million.
The households that are included in the sample were ones that purchased at least twice at

[4] we can assume that the shelves'-promotional tags have indicated the large prices decrease, which were probably captured in the significant effect coming from the residual in the control function.

[5] Another level of accuracy will be added when the exact price sensitivity for before and after the beginning of the recession will be added

[6] IRI market research, 2008

Table 3.1: Market Share by Main Attributes

(a) Sub-Brand

Sub-Brand	Frequency	Market Share
OOP	11,727	29.93
TIDE	10,542	26.91
PRIVATE LABEL	6,363	16.24
ALL	3,332	8.5
GAIN	2,468	6.3
ALL FREE CLEAR	2,380	6.07
ARM & HAMMER	2,366	6.04

(b) Loads-Category

Size Category	Freq.	Percent
size2 (32 loads)	8,203	29.88
size4 (64 loads)	5,997	21.85
size3 (36,51,52 loads)	5,846	21.3
size6 (96 loads)	5,404	19.69
size1 (24,26,30 loads)	1,824	6.64
size5 (78 loads)	177	0.64

the store, at least once at the liquid laundry detergent category (from the inside or outside option goods), and not more than once in a two weeks. Thus overall the sample includes 6210 households and 39,178 choice situations. [7]

The type of products in this category are durable goods that are required product for households. The category includes six leading sub-brands (of five brands) that held over 70% of the market share in the category. Each sub-brand offered a large variety of per unit prices and a large variety of package sizes that changed over time as can be seen in Table 3.1a. [8]

In the laundry detergent category, the main attributes that affect the value of a package are sub-brand and number of loads in a package. It is important to notice that the package size may vary, however when testing the demand model, the size itself is not significant to consumers (even when including heterogeneity), but the number of loads is. [9]

The positive value is compared to the package's price. Hence, the main attributes of a product that affect the choice in the utility model are price, sub-brand, and number-of-loads.

Table 3.1b presents the market share of each loads-size category and Figure 3.1 summarizes the per product and per size pricing. Altogether the inside good's market share is about 70%. As we can see, Tide is the leading sub-brand with a big gap compared to the rest of the sub-brands. Then comes the private label product, which also has double the market shares of the rest of the sub-brands.

There were significant quantity discounts as were seen in Figures 3.1 and **??**. The market

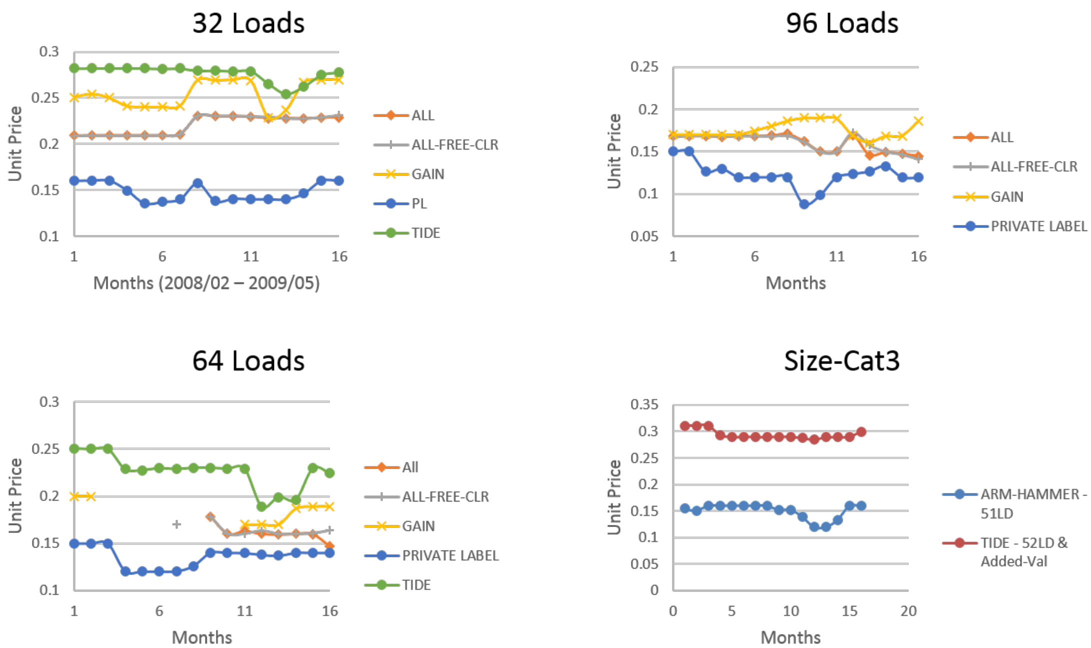
[7] The filtering conditions are give because of a few reasons: 1. a household-id that buys more than once in two weeks - it probably buys for some different households. 2. a household that purchased in the supermarket less than twice in the period is not a real customer. Maybe they just had to buy once due to an urgency and in urgency conditions, it is hard to identify real preferences. 3. If they have never purchased from the category, most chances are that they prefer to buy from this category in another retail store. This is because this category is required.

[8] These market shares are in-line with the US market shares according to IRI, 2008

[9] This notice is important for papers that are interested in studying quantity discounts as the size of the package is usually considered.

share by loads-category show that consumers generally prefer the packages of the 32 or 64 loads. As we could expect, consumers prefer smaller packages at this retail grocery store, however they prefer to use the quantity discounts.

Figure 3.1: Pricing of Sub-Brands per Loads' Category through the Panel Data



The household summary statistics is given in Table 3.3. This summary validates that the panel data is reliable given the variation in chosen products by brand and by size per household. In fact, 75% varied their choices.

As discussed in 3.1, the market could be divided to two main segments that have reacted differently to the recession. Table 3.4 presents the significant changes that have occurred in correlation with the indicator of period 2, which is the first eight months of the recession in the US. It is commonly not very apparent from the data how to divide consumers to segments. However, in this case, the differences between each segments behaviors in period 1 compared to period 2 were very salient.

Segment 1 was of 78.1% of the consumers' population and segment 2 included the rest, which was 21.9%. In period 2, Segment 2 has purchased products with significantly higher unit price per product and also decreased the number of loads per package at the same time.

On top of these summary statistics, there were significant changes in brands' consumption between the segments and between the periods as presented in Figure 3.2. Segment 1 has significantly increased choices of the PL and Arm & Hammer on top of the outside option (OOP) options and Tide, while Segment 2 has significantly increased for Tide on top of the PL and OOP.

Table 3.2: Unit Price on Product Attributes

	Mean
ALL	0.219
	(0.001)
ALL FREE CLEAR	0.219
	(0.001)
ARM & HAMMER	0.15
	(0)
GAIN	0.245
	(0.001)
PRIVATE LABEL	0.171
	(0.001)
TIDE	0.278
	(0.001)
size1 (24-30 loads)	0.004
	(0)
size2 (32 loads)	-0.002
	(0.001)
size4 (64 loads)	-0.048
	(0.001)
size5 (78 loads)	-0.035
	(0)
size6 (96 loads)	-0.058
	(0.001)
Added Value	0.016
	(0.001)
N	573779
r2	0.993
F	6.96E+06
Standard Errors in parenthesis	

Note: These are the results of the OLS regression of the unit price on the attributes as listed in the table. Standard errors are in parentheses. The regression shows a very large fit of R2 = 99%.

3.3 Demand Model and Estimation Strategy

The primitives of the model are the product characteristics, consumer preferences, and the equilibrium notion. I assume that households buy one product from a choice set S with J + 1 alternatives. I denote N as the number of households in the panel and Tn as the number of weeks for which I observe choices for every household. The consumers in a household derive utility from one product in that choice set (buying alternative j = 0 means that the consumer made no purchase of the available brands).

The household buys the product for which the perceived utility is the greatest, but will make no purchase from the category if the utility of each of the brands is less than the utility if it makes no purchase.

Table 3.3: Households Summary Statistics

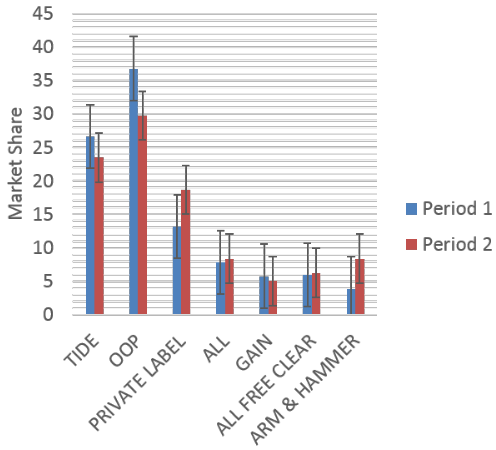
	Mean	sd	Min	Max
Product Variation	2.93	1.48	1.00	9.00
# of loads in a package	55.27	18.49	26.00	96.00
# of packages per product	1.08	0.22	1.00	4.00
Mean unit price (per FZ)	0.21	0.05	0.05	0.35
# of puchases	6.31	3.33	1.00	17.00
Total Payment per Package	11.33	3.39	4.49	35.98
Total Spend in the Category	55.13	39.82	4.48	352.80
Tenure in Panel	10.99	4.36	1.00	16.00
Min month	3.53	3.12	1.00	16.00
Min month	13.52	3.17	1.00	16.00
Mean loads in package	39.64	24.18	0.00	96.00
# of packages per product in purchase	1.10	0.24	1.00	4.00

Table 3.4: Two Segments changed Behaviors between the Periods

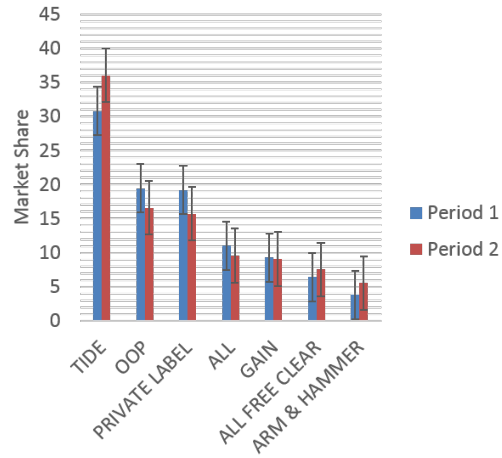
Groups	Loads Category Period1	Loads Category Period2	Unit Price Period1	Unit Price Period2	Total Price Period1	Total Price Period2	# of Choices Period1	# of Choices Period2
Segment1								
# of households	4502							
% of population	78.1%							
Mean	2.92	3.64	0.07	0.07	10.96	11.43	3.43	3.39
SD	1.22	1.26	0.11	0.10	4.77	4.77	2.00	1.97
Segment2								
# of households	1163							
% of population	**21.9%**							
Mean	4.03	2.99	0.08	0.09	12.37	11.18	3.59	3.77
SD	1.29	1.09	0.11	0.12	4.61	4.51	1.93	2.02

Figure 3.2: Market Share by Main Attributes

(a) Segment 1 - Brands period1 vs. period2



(b) Segment 2 - Brands period1 vs. period2



The utility that consumers in household i at time t obtains from product $j \in 1..J$ depends on observed and unobserved attributes of the product, where for j=0, $U_{i0t} = 0$. Assume that utility takes the form

$$U_{ijt} = -\alpha p_{jt} + \mathbf{X_j}\beta_{\mathbf{ij}} + \psi_{ijt} + \epsilon_{ijt} \tag{3.1}$$

This model follows the standard model that is defined in Berry, 1994. Where α_i is the mean price sensitivity per market t, p_{jt} is the price of product j at market t, $\mathbf{X_j}$ are product j's attributes, which don't change along the time, β_{ijt} is a vector of tastes per attribute of household i per attribute of product, j. ψ_{ijt} is the unobserved heterogeneity in tastes, and ϵ_{ijt} is the household, product, and market, i.i.d demand shock, at time t.
To ease notation, denote:

$$V_{ijt} = -\alpha_i p_{jm} + \mathbf{X_j}\beta_{\mathbf{ij}} + \xi_{jm} + \psi_{ijt} \tag{3.2}$$

Then:

$$U_{ijt} = V_{ijt} + \epsilon_{ijt} \tag{3.3}$$

A simplifying assumption commonly made (see McFadden, 1981, BLP 1995, and more) is that ϵ_{ijt} is distributed i.i.d. with a Type I extreme value distribution. Moreover, in this model, the mean utility of the outside good is not identified, and it normalized to 0.
Given these and the previous definitions, the likelihood that a household will purchase product j in market t is defined as follows

$$Prob(y_{ijt} = 1|\theta) = \int \frac{\exp(V_{ijt})}{1 + \sum_{k \in J} \exp(V_{ikt})} df(\cdot|\theta) d\eta_{ijt} \tag{3.4}$$

Where $y_{ijt} = 1$ if household i has chosen product j in market t given the parameters θ, which are the parameters of distribution F, from which the house-level tastes and preferences

are drawn.
As in McFadden and Train, 2000, and as in the other papers that use SMML, tastes are drawn for all the attributes and the interactions of them for products, 1..J, when the taste is for a product. Suppose the distribution F is the normal distribution, then parameters to be estimated for the distribution are the mean for the tastes and the variation matrix. This is described by the following

$$\begin{aligned}
&let\ \eta = \begin{bmatrix}\alpha \\ \beta\end{bmatrix}\ and\ \tilde{\eta_{\mathbf{ijt}}} = \begin{bmatrix}\tilde{\alpha_i} \\ \tilde{\beta_{\mathbf{ijt}}}\end{bmatrix} \\
&Then \\
&\eta_{\mathbf{ijt}} = \eta + \Sigma^{1/2}\tilde{\eta_{ijt}} + \Pi D_i \quad s.t. \quad \tilde{\eta_{\mathbf{ijt}}} \ \sim N(0, I_K)
\end{aligned} \tag{3.5}$$

where Σ is the matrix of standard error parameters [10], we assume that the parameters are distributed normally, which should be fine for the goal of the estimation, and K equals to the number of random variables, including interaction variables, I is the identity matrix, Π is a $K \times d$ matrix of the coefficients that measure how the taste characteristics (suppose there are K of variables, including the interaction variables) vary with demographics (d demographic variables).
We use the notation t for *beta* and for *eta* as these are functions of t when they are put into the model, however the draws from the distribution are not affected by t and neither the coefficients that *beta* includes as in equation 1.2.

3.4 Estimation

Each purchase may include more than 1 package from each product. In such case, the quantity will be higher, but the purchase will be defined as one purchase and the unit price will be calculated for the total number of loads in the number of packages of the product.
To estimate the model, we can write the likelihood function as the product of the likelihood of purchases, given the prices and the product attributes, and conditional on the parameters, which are estimated through the model, as follows

$$L = \prod_{t=1}^{NM}\prod_{j=1}^{J}\prod_{i=1}^{NP} Prob(y_{ijt} = 1|\theta) \tag{3.6}$$

This likelihood function, where NM is the number of markets, J is the number of products per market, and NP is the number of households in the panel. The probability function is as defined in 3.4. [11] There are a few possible techniques to estimate the model. Since the data is individual level data, the simulated maximum likelihood as discussed in McFadden

[10] $\Sigma^{1/2}$ is the Cholesky factorization

[11] each household made different number of choices and each brand was selling in a different number of market, which the matlab code overcomes.

and Train, 2000, was chosen as the estimation method. Table 3.5 presents the results of the estimation. The results are that segment 2 is price loving. This is not consistent with

Table 3.5: Model Estimation with no treatment for Endogeneity

	All data		Segment 1		Segment 2	
	Mean	SD	Mean	SD	Mean	SD
rnd_unit_prc_menu	0.399		-0.745		5.017	
	(0.439)		(0.516)		(0.87)	
ALL	-3.953	1.721	-3.652	1.791	-5.694	1.536
	(0.418)	(0.052)	(0.461)	(0.06)	(1.293)	(0.089)
ALL FREE CLEAR	-5.415	2.665	-4.926	2.551	-7.341	2.751
	(0.425)	(0.077)	(0.47)	(0.088)	(1.304)	(0.175)
ARM & HAMMER	-3.064	2.215	-3.117	2.335	-2.574	1.429
	(0.099)	(0.063)	(0.119)	(0.074)	(0.173)	(0.099)
GAIN	-4.756	2.253	-4.563	2.405	-6.104	1.982
	(0.425)	(0.064)	(0.469)	(0.082)	(1.301)	(0.104)
PRIVATE LABEL	-3.651	2.131	-3.379	2.136	-5.296	1.941
	(0.413)	(0.05)	(0.455)	(0.058)	(1.287)	(0.096)
TIDE	-3.658	2.688	-3.414	2.859	-5.115	2.148
	(0.425)	(0.051)	(0.47)	(0.063)	(1.302)	(0.096)
size1 (24-30 loads)	-1.038	1.368	-0.934	1.158	-0.99	1.208
	(0.053)	(0.058)	(0.059)	(0.069)	(0.096)	(0.108)
size2 (32 loads)	0.992	0.778	0.745	0.887	2.324	0.915
	(0.405)	(0.04)	(0.444)	(0.043)	(1.277)	(0.069)
size4 (64 loads)	1.111	0.786	0.667	0.899	2.932	0.621
	(0.406)	(0.042)	(0.445)	(0.047)	(1.278)	(0.094)
size5 (78 loads)	-2.24	-0.629	-2.523	0.872	-1.732	0.031
	(0.121)	(0.173)	(0.172)	(0.192)	(0.139)	(0.312)
size6 (96 loads)	0.261	2.056	-0.39	2.316	2.698	1.192
	(0.407)	(0.047)	(0.448)	(0.066)	(1.277)	(0.071)
added_val_menu	1.195	1.345	0.883	1.389	2.751	1.237
	(0.405)	(0.047)	(0.445)	(0.063)	(1.277)	(0.084)
N	573779		448102		125677	
chi2	39020.4		31399.53		7160.549	
ll	-74183.7		-55620.9		-18229.4	

Note: Standard errors in parentheses. Size3 represents the products that have added values, such as softeners. Thus, size3 effects are captured by the brands and the added value attribute.

the period of the recession, however it is consistent with the summary statistics that show that segment 2 increased the unit price index in period 2. Given the inconsistency with the product category and with the recession period, treatment of endogeneity was considered

Possible Endogeneity

The prices in the model defined in 3.1 may be correlated with the error term as prices move with unobserved demand shocks, such as news reports or other factors that affect segment 2 and are unobserved in the data.
The estimation handles these possible problems through the control-function method as defined in Petrin and Train, 2009, as follows

$$p_{jt} = W(\mathbf{\tilde{Z}_j}, \zeta_\mathbf{j}) + \nu_{jt} \tag{3.7}$$

$$p_{jt} = \gamma_{\mathbf{jt}}\mathbf{Z_{jt}} + \nu_{jt} \tag{3.8}$$

$$p_{jt} = \hat{p_{jt}} + \nu_{jt} \tag{3.9}$$

Using the linear model with the observed attributes only has a fit of 99.3% as can be seen in Table 3.2. This is a very high fit, hence does not require additional techniques for the control function in case it was not significantly explain the price. [12]

Thus, built upon the model that was discussed in 3.3, and given the data and endogeneity treatments, the following empirical utility model, is a complete model that allows identification of the required parameters

$$U_{ijt} = -\alpha_i \hat{p_{jt}} + \mathbf{X_j}\beta_{\mathbf{ijt}} + \psi_{ijt} + \lambda_{ij}\nu_{jt} + \epsilon_{ijt} \tag{3.10}$$

The likelihood function 3.6 is updated according to 3.10, and the model is estimated in the same manner discussed prior to adding the treatment for endogeneity. [13] The results are presented in table 3.6. As we can see, including the endogeneity treatment has cleared the picture, so that in fact both segments are very sensitive to price, however segment 2 appreciates the market leader's (Tide's) products more than segment 1. Moreover, while the residual for the control function has zero mean for segment 1, it is not the case for segment 2. There are significant unobserved characteristic effects that affet segment 2's utility.

3.5 Discussion

This chapter considers a case where an exogenous effect has changed the demand and supply behaviors in the market. While beforehand the leading brand was holding price almost constant, in period 2, it has significantly deceased the pricing level for a few months. It is also rational to assume that consumers have become more sensitive to price or more limited

[12] In cases were the fit is not very high for the linear function as discussed in Petrin and Train 2009, another control function method that does not force a functional form is discussed in Villas-Boas and Winer, 1999, and another method that can be helpful is BLP, 2004.

[13] It is also possible to use Villas-Boas and Winer, 1999, or BLP, 2004 in case it was important not to use a functional form for the price's possible endogeneity. This was not important for this chapter's purposes.

Table 3.6: Model Estimation

	All		Segment 1		Segment 2	
	Mean	SD	Mean	SD	Mean	SD
Pred Unit Price	-27.957		-28.382		-29.736	
	(3.654)		(4.098)		(8.579)	
Residual	-1.167		0.214		-5.627	
	(0.442)		(0.521)		(0.877)	
ALL	2.332	1.758	2.41	1.797	2.455	1.574
	(0.894)	(0.055)	(1.004)	(0.06)	(2.18)	(0.091)
ALL FREE CLEAR	0.81	2.734	1.127	2.562	0.842	2.577
	(0.896)	(0.075)	(1.006)	(0.088)	(2.184)	(0.154)
ARM & HAMMER	1.159	2.269	1.023	2.335	2.636	1.425
	(0.552)	(0.062)	(0.62)	(0.074)	(1.29)	(0.094)
GAIN	2.335	2.181	2.215	2.4	3.006	1.987
	(0.981)	(0.068)	(1.102)	(0.082)	(2.375)	(0.104)
PRIVATE LABEL	1.256	2.144	1.362	2.138	1.137	2.001
	(0.741)	(0.048)	(0.832)	(0.057)	(1.837)	(0.095)
TIDE	4.238	2.775	4.252	2.853	5.049	2.192
	(1.088)	(0.053)	(1.222)	(0.063)	(2.62)	(0.087)
size1 (24-30 loads)	-0.827	1.298	-0.812	1.158	-0.839	1.275
	(0.054)	(0.063)	(0.062)	(0.069)	(0.101)	(0.113)
size2 (32 loads)	0.81	0.967	0.693	0.886	1.695	0.941
	(0.395)	(0.036)	(0.444)	(0.043)	(1.098)	(0.065)
size4 (64 loads)	-0.212	0.259	-0.641	0.903	0.744	0.693
	(0.431)	(0.069)	(0.485)	(0.047)	(1.171)	(0.078)
size5 (78 loads)	-3.263	0.682	-3.494	0.873	-2.954	0.101
	(0.171)	(0.147)	(0.224)	(0.193)	(0.331)	(0.305)
size6 (96 loads)	-1.47	2.042	-1.991	2.309	0.2	1.058
	(0.449)	(0.047)	(0.506)	(0.066)	(1.205)	(0.065)
Added Value	1.473	1.518	1.342	1.39	2.781	1.199
	(0.4)	(0.045)	(0.45)	(0.063)	(1.108)	(0.083)
N	573779		448102		125677	
chi2	38933.74		31377.43		7177.323	
ll	-74181.3		-55597.7		-18212.5	

Note: Standard errors in parentheses. Results for segment 1 and segment 2 are separated and on the left most columns there are the results for all the segments together. Size3 represents the products that have added values, such as softeners. Thus, size3 effects are captured by the brands and the added value attribute.

on borrowing ability given the situation in the market. However, unlike the expectations, a large portion of the market increased payment per unit for the products.
A few models were ran to explain the data, two of them are presented in this chapter. The first is a model that does not include treatment for endogeneity. It estimates that Segment 2 is price loving at period 2. The second model includes treatment for endogeneity and the

results are that, as expected, the sensitivity to pricing of both segments is very high. Thus, clearly, the first model did not identify the true sensitivity to pricing of Segment 2.

This clean result through the standard demand model, on a market that initially seemed to behave in an unexplained way is another important example that adds to the literature that discusses behaviors where products seem to be Giffen goood and it is an helpful tool for marketers to realize how to explore their markets when a huge exogenous event, like a recession occurs.

In this case, the leading brand's marketers could have noticed that a large portion of their target market has significantly changed. In fact, their target market was now of the consumers who are more sensitive to price than before, however these consumers greatly appreciate the chance to finally purchase the brand' products, once the prices have decreased. Together with that, it is possible that they could not allow themselves to buy bigger products giving their limited borrowing ability, and hence had to decrease number of loads per package, so that the total price of the product will not be too high for them.

While the limited borrowing ability option is of high interest. These data did not have the required variation to identify the effect. However, a follow-up work in this direction could use similar data analysis methods as in this chapter, and if there are categories where consumers reduced the sizes and did not significantly change their brand preferences, a limited borrowing ability could be found.

Another perspective of Segment 2's behavior is that the estimation has found a significant effect to the unobserved characteristics in its utility model. Marketers, who should have complete information about promotions and other sources for correlated demand shocks, should notice this effect and complement the information of what have changed the utility for the segment through varied marketing tools.

The model that made the best fit to the data and identified consumers preferences, assumes that the sensitivity to pricing has not change between the periods. However, the expectation is that it has changed. The reason is that these data did not allow finding the per period sensitivity to pricing. The main reason is that the prices did not fluctuate much in period 1 and consumers did not varied their choices in a way that allowed realizing Segment 2's high preference to the leading brand.

On the other hand, adding the two periods together revealed new consumers to new brands. This model and estimation, for example allowed marketers of the leading brand, Tide, to observe the high preference to Tide of Segment 2.

To conclude, this paper presents how a unique events, such as a recession, can change the market. While in such cases an immediate belief is that consumers might change their preference and design models accordingly, it shows that in some cases it is better to add the information of before and after together.

Finally, this paper joins the previous chapters by emphasizing the need to identify segments within the consumer population and analyzing them separately.

Bibliography

[1] Greg M Allenby and Peter E Rossi. Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89(1):57–78, 1998.

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

[3] Steve Berry, Oliver B Linton, and Ariel Pakes. Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654, 2004.

[4] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

[5] Steven Berry, James Levinsohn, and Ariel Pakes. Differentiated products demand systems from a combination of micro and macro data: The new car market. Technical report, National bureau of economic research, 1998.

[6] Steven T Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

[7] Timothy F Bresnahan. Departures from marginal-cost pricing in the american automobile industry: Estimates for 1977–1978. *Journal of Econometrics*, 17(2):201–227, 1981.

[8] Timothy F Bresnahan. Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics*, pages 457–482, 1987.

[9] Benjamin Caballero. The global epidemic of obesity: an overview. *Epidemiologic reviews*, 29(1):1–5, 2007.

[10] Pradeep K Chintagunta. Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science*, 17(3):253–270, 1998.

[11] Hayley H Chouinard, David E Davis, Jeffrey T LaFrance, and Jeffrey M Perloff. Fat taxes: big money for small change. In *Forum for Health Economics & Policy*, volume 10, 2007.

[12] Gregory S Crawford and Matthew Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

[13] Adam Drewnowski and Nicole Darmon. Food choices and diet costs: an economic analysis. *The Journal of nutrition*, 135(4):900–904, 2005.

[14] Jean-Pierre Dubé. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science*, 23(1):66–81, 2004.

[15] Dennis Epple, Richard Romano, and Holger Sieg. Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928, 2006.

[16] Tülin Erdem. A dynamic analysis of market structure based on panel data. *Marketing science*, 15(4):359–378, 1996.

[17] Tülin Erdem and Michael P Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1):1–20, 1996.

[18] Joseph Farrell and Paul Klemperer. Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, 3:1967–2072, 2007.

[19] Marshall Freimer and Dan Horsky. Periodic advertising pulsing in a competitive market. *Marketing Science*, 31(4):637–648, 2012.

[20] Lewrene K Glaser, Gary D Thompson, et al. Demand for organic and conventional beverage milk. In *Western Agricultural Economics Association Annual Meeting, June*, 2000.

[21] Brian W Gould. Factors affecting us demand for reduced-fat fluid milk. *Journal of Agricultural and Resource Economics*, pages 68–81, 1996.

[22] Amir Heiman, Bruce McWilliams, Zhihua Shen, and David Zilberman. Learning and forgetting: Modeling optimal product sampling over time. *Management Science*, 47(4):532–546, 2001.

[23] Amir Heiman and Eitan Muller. Using demonstration to increase new product acceptance: Controlling demonstration time. *Journal of Marketing Research*, pages 422–430, 1996.

[24] Dipak Jain, Vijay Mahajan, and Eitan Muller. An approach for determining optimal product sampling for the diffusion of a new product. *Journal of Product Innovation Management*, 12(2):124–135, 1995.

[25] Wagner A Kamakura and Gary J Russell. Measuring brand value with scanner data. *International journal of Research in Marketing*, 10(1):9–22, 1993.

[26] John Kearney. Food consumption trends and drivers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1554):2793–2807, 2010.

[27] Romana Khan, Kanishka Misra, and Vishal Singh. Will a fat tax work? *Marketing Science*, 2015.

[28] Kristin Kiesel and Sofia B Villas-Boas. Got organic milk? consumer valuations of milk labels after the implementation of the usda organic seal. *Journal of agricultural & food industrial organization*, 5(1), 2007.

[29] Jun B Kim, Paulo Albuquerque, and Bart J Bronnenberg. Online demand under limited consumer search. *Marketing science*, 29(6):1001–1023, 2010.

[30] Fred Kuchler, Abebayehu Tegene, and J Michael Harris. Taxing snack foods: Manipulating diet quality or financing information programs? *Applied Economic Perspectives and Policy*, 27(1):4–20, 2005.

[31] Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

[32] Daniel McFadden. The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328, 1974.

[33] Daniel McFadden. Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J.W. Wiebull, editors, *Spatial Interaction Theory and Planning Models*, pages 75–96. North Holland, Amsterdam, 1978.

[34] Daniel McFadden. Econometric models of probabilistic choice. In C.F. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. MIT Press, Cambridge, MA, 1981.

[35] Thomas T Nagle, John E Hogan, and Joseph Zale. *The strategy and tactics of pricing: A guide to growing more profitably.* Pearson/Prentice Hall Upper Saddle River, NJ, 2006.

[36] Aviv Nevo. Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, pages 395–421, 2000.

[37] Matthew Osborne. Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics*, 9(1):25–70, 2011.

[38] Amil Petrin and Kenneth Train. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13, 2010.

[39] Lisa M Powell and Frank J Chaloupka. Food prices and obesity: evidence and policy implications for taxes and subsidies. *Milbank Quarterly*, 87(1):229–257, 2009.

[40] David Revelt and Kenneth Train. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, 80(4):647–657, 1998.

[41] David Revelt and Kenneth Train. Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. *Department of Economics, UCB*, 2000.

[42] Sherwin Rosen. Potato paradoxes. *Journal of Political Economy*, 107(S6):S294–S313, 1999.

[43] Peter E Rossi, Robert E McCulloch, and Greg M Allenby. The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340, 1996.

[44] John Rust. Estimation of dynamic structural models, problems and prospects: Discrete decision processes. In *Advances in Econometrics*, chapter 4, pages 119–170. 1994.

[45] Trenton G Smith. Obesity and the economics of prevention. *American Journal of Agricultural Economics*, page aar141, 2011.

[46] Raphael Thomadsen and Pradeep Bhardwaj. Cooperation in games with forgetfulness. *Management Science*, 57(2):363–375, 2011.

[47] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[48] E Train Kenneth. Discrete choice methods with simulation, 2003.

[49] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

[50] Hal R Varian. A model of sales. *The American Economic Review*, 70(4):651–659, 1980.

[51] J Miguel Villas-Boas. Consumer learning, brand loyalty, and competition. *Marketing Science*, 23(1):134–145, 2004.

[52] J Miguel Villas-Boas. A note on limited versus full information estimation in non-linear models. *work-ing paper, Haas School of Business, University of California, Berkeley*, 2007.

[53] J Miguel Villas-Boas and Russell S Winer. Endogeneity in brand choice models. *Management Science*, 45(10):1324–1338, 1999.

[54] Sofia Berto Villas-Boas and J Miguel Villas-Boas. Learning, forgetting, and sales. *Management Science*, 54(11):1951–1960, 2008.

[55] E Wang. Estimating the distributional impact of taxes on storable goods: A dynamic demand model with random coefficients. Technical report, Working Paper, Duke University, 2010.

[56] Michel Wedel and Wayne S DeSarbo. A mixture likelihood approach for generalized linear models. *Journal of Classification*, 12(1):21–55, 1995.

[57] Gabriel Y Weintraub, C Lanier Benkard, Przemyslaw Jeziorski, and Benjamin Van Roy. Nonstationary oblivious equilibrium. *Manuscript, Columbia University*, pages 1375–1411, 2008.

[58] Gabriel Y Weintraub, C Lanier Benkard, and Benjamin Van Roy. Oblivious equilibrium: A mean field approximation for large-scale dynamic games. In *NIPS*, pages 1489–1496, 2005.

[59] Gabriel Y Weintraub, C Lanier Benkard, and Benjamin Van Roy. Computational methods for oblivious equilibrium. *Operations research*, 58(4-part-2):1247–1265, 2010.