Essays on Quantity Surcharges and Consumer Heterogeneity

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Hyewon Kim, M.A.

Graduate Program in Economics

The Ohio State University
2018

Dissertation Committee
Javier Donna, Advisor
Jason Blevins
Maryam Saeedi

Abstract

This dissertation examines quantity surcharges and heterogeneity in consumer attention. Quantity surcharges exist for packaged goods when a smaller package size is cheaper than its larger size counterpart per unit. Even though consumers face quantity surcharges in their daily lives, minimal research has been done on this topic.

Chapter 1 documents the relevance of quantity surcharges for households who shop at grocery stores. I use rich scanner data in the peanut butter category for this analysis and throughout the other chapters. The data show that quantity surcharges are highly frequent: they exist in 62% of weeks. Quantity surcharges also exist consistently over time, rather than being a side effect of occasional sales on small size items. Households have heterogeneous purchasing behavior during quantity surcharge periods: some households take advantage of the surcharge and purchase multiple small size jars, but others pay extra and buy large jars ("miss" purchases). I analyze the characteristics of those households who make miss purchases using negative binomial regressions, and find that they have large expenditures per shopping trip and small variety of peanut butter purchases.

Chapter 2 takes a deeper look at the purchasing behavior in response to quantity surcharges and welfare effects. It begins by comparing explanations for purchasing behavior through estimated demand models. I compare three models: (1) a standard discrete choice model for packaged goods, (2) a model that allows preferences for package size, and (3) a model with inattentive consumers. In the consumer inattention model, some households are not aware of the existence of quantity surcharges when they shop due to inattention, and make miss purchases. I use Bayesian methods to estimate the models. Among the three models I consider, the consumer inattention model explains the miss purchases best: it has the smallest prediction error for expected demand. I apply the estimation results to calculate the costs of consumer inattention. The simulation results show that a household loses $0.69 per a miss purchase on average.

Chapter 3 introduces an alternative demand model for packaged goods that considers households' dynamic decisions to consume and purchase. The model features heterogeneity of consumers in storing and attentiveness, and allows four household types: attentive non-storer, inattentive non-storer, attentive storer, and inattentive storer. I identify the household types using the method of *kmeans* and estimate the parameters using maximum likelihood. The estimation results show that attentive storer type households are most price sensitive and inattentive non-storer type households are least price sensitive.

Acknowledgements

I am grateful to Jason Blevins, Maryam Saeedi, Bruce Weinberg, and especially Javier Donna for their guidance and support. I would like to say special thanks to my husband Andrew Chen for all the support and encouragement. I would also like to thank SymphonyIRI Group, Inc. for making the data available. All estimates and analyses in this paper, based on data provided by SymphonyIRI Group, Inc. are by the author and not by SymphonyIRI Group, Inc. Any errors are my own.

Vita

2010	....	B.A. Economics, Yonsei University
2010	....	B.A. Applied Statistics, Yonsei University
2012	....	M.A. Economics, Yonsei University
2013	....	M.A. Economics, The Ohio State University

Fields of Study

Major Field: Economics

Area of Specialization: Industrial Organization

Contents

List of Figures

List of Tables

Chapter 1

Quantity Surcharges and Consumer Heterogeneity

1.1 Introduction

Quantity surcharges exist when a small package size item is cheaper than its larger size counterpart per unit (e.g., roll, ounce). The opposite of quantity surcharges, quantity discounts, may be the more intuitive pricing strategy. Quantity surcharges, however, are frequently observed at grocery stores (Widrick (1979)). Data plans for smart phones that allow customers to use a certain amount of data at a fixed fee and then charge a lot more per byte once they pass the limit are also an example of quantity surcharges.

In this chapter, I study quantity surcharges at grocery stores, focusing on heterogeneous consumer behavior. I use rich scanner data in the peanut butter category for the study. Quantity surcharges are highly frequent for peanut butter products. Consumers show heterogeneous purchasing behavior when quantity surcharges exist: some purchase multiple small size jars while some others purchase large size jars. The households who purchase large size jars in quantity surcharge periods show distinctive characteristics.

I identify quantity surcharges between small and large size jars, controlling for all the other observable characteristics. Quantity surcharges are more frequent than quantity discounts in the peanut butter category, and the price gap between the two sizes is also bigger in quantity surcharge weeks than quantity discount weeks.

I prove that quantity surcharges are consistent phenomena rather than being fully determined by promotional activities. Sale on small size jars increases the frequency of quantity surcharges. However, in many other occasions quantity surcharges occur without small size items being on sale. The non-price promotional activities, feature and display, have no significant effect on quantity surcharges.

Households show two interesting purchasing patterns when quantity surcharges exist. First, multiple small size jar purchases are four times more frequent in quantity surcharge weeks than in quantity discount weeks. This is not a surprise once we understand the substitution opportunities in quantity surcharge periods: consumers who demand a large quantity can save money by purchasing multiple small size items instead of one large size item. However, 19.43% of households purchased large size jars in quantity surcharge weeks. Clerides and Courty (2017) argue that those households who miss substitution opportunities are "inattentive".

I further study the households who make miss purchases. I design a negative binomial regression model that explains the number of misses each household make using the three groups of household characteristics: demographics, grocery shopping, and peanut butter purchases. Estimation results suggest that households with big family size and large expenditure on each grocery shopping make more miss purchases in quantity surcharge weeks. Households who purchase a variety of peanut butters make less miss purchases.

The literature provides two explanations for quantity surcharges. One explanation is that retailers want to price discriminate against consumers with high demand (Agrawal, Grimm, and Srinivasan (1993)). The other argues that quantity surcharges exist as a side effect of temporary price promotions when only the small package is on sale (Sprott, Manning, and Miyazaki (2003) and Clerides and Courty (2017)). The welfare effects of quantity surcharges to consumers are ambiguous in both cases. It is critical to identify inattentive consumers and understand their behavior to analyze the welfare effects.

The peanut butter category is ideal for studying quantity surcharges for the following reasons. The dominance of three national brands shortens the list of brands to consider. Also, peanut butter products are relatively homogeneous, which makes it easier to compare apple to apple. Lastly, both quantity discounts and quantity surcharges are widely observed at grocery stores.

Even though consumers face quantity surcharges in their daily lives, minimal research has been done on this topic. Also, most of this research is limited to use store level data and fails to catch heterogeneity in consumer behavior (Agrawal, Grimm, and Srinivasan (1993), Sprott, Manning, and Miyazaki (2003), and Clerides and Courty (2017)). Understanding heterogeneity in attention is important, especially for packaged goods in the presence of nonlinear pricing.

1.1.1 Literature Review

Few papers in the marketing literature study quantity surcharges. Widrick (1979), Nason and Della Bitta (1983), and Cude and Walker (1984) document the existence of quantity surcharges. Widrick (1979) focus on 10 product categories at 37 grocery stores in upper New York State using cross-sectional data. He found a high percentage of quantity surcharges across the categories (e.g., 84.4% for canned tuna fish and 33.3% for laundry detergent).

Agrawal, Grimm, and Srinivasan (1993) and Sprott, Manning, and Miyazaki (2003) study quantity surcharges on the seller side. Agrawal, Grimm, and Srinivasan (1993) interpret quantity surcharges as a practice of price discrimination against consumers with high demand. On the other hand, Sprott, Manning, and Miyazaki (2003) argue that stores practice quantity surcharges in order to build a low store-price image. Stores have incentive to lower the price of small size items when those are the items with a high sale volume. However, neither analyzes household purchasing behavior with the existence of quantity surcharges.

Kumar and Divakar (1999) is not directly related to quantity surcharges, but studies different package sizes offered within a product. The authors argue that competition exists not just at a brand level but across different brand-sizes. This implies that stores should practice promotional strategies at the brand-size level.

Clerides and Courty (2017) is the first attempt to analyze quantity surcharges on the consumer side, to the best of my knowledge. Using store level data in the laundry detergent category, the authors find that roughly 45%–75% of households are inattentive in the sense that they miss the substitution opportunity to switch from the large package size to the small size in quantity surcharge weeks. The authors suggest search costs as a rationale for consumer inattention. However, due to the limitation of store level data, the authors cannot analyze the behavior of the households who are attentive.

The chapter proceeds as follows. Section 2 describes the data sets used for the analyses and provides summary statistics. Section 3 studies quantity surcharges at grocery stores and Section 4 analyzes household behavior when quantity surcharge exists. Section 5 concludes the chapter.

1.2 Data

1.2.1 Description of Data

I use weekly panel scanner data collected by Information Resources Inc. (IRI) for the analyses. The data set covers a period from January 2001 to December 2011, and 31 categories of products, including beer, carbonated beverages, and laundry detergent (Bronnenberg et al. (2008)). IRI collected data from supermarkets in 50 regional markets defined by IRI, and also from household panels from two of the regional markets. I focus on the peanut butter category in the Eau Claire, Wisconsin, market, from January 2008 to December 2010. Eau Claire is one of the two markets that have both store level and household level data.[1]

The store level data consist of weekly peanut butter sales observations at store-UPC (Universal Product Code) level, including the number of jars sold, total sales in terms of dollar amount, and information on promotional activities. From now on, I use "grocery stores" and "stores" interchangeably. The household level data include peanut butter purchases, trips to grocery stores, and household demographic characteristics. There is also an additional data set available on the attributes of peanut butter items. Please see Appendix A.1.1 for more details.

I merged the three household level data sets and product attributes data in order to extract the maximum information on households' product purchases. Then I merged the store level data set to get additional information on price levels and promotional activities. I dropped the observations that were not matched during the merging processes. I also dropped the households who showed extreme purchase history: more than six jars on a single trip or more than 100 jars in total during the time period.

Appendix A.1.2 describes how I merged each data set, and Appendix A.1.3 shows how I further clean the data set in detail. The final data set contains 2,367 households with 23,287 purchase observations in total. Those households purchase 123 UPCs of peanut butter items from six grocery stores. The households in the sample are broadly representative of the national data. The household demographics in the sample are similar to the national level, except for race. Please see Appendix A.2 for details.

1.2.2 Peanut Butter Market in Eau Claire

The big three national brands of peanut butter are Jif, Skippy, and Peter Pan. Jif and Skippy have significantly large market shares in Eau Claire, at 31% and 28%, respectively. The market shares of the top 10 selling brands are listed in Table 1.1. J.M. Smucker Company, Jif's current parent company, acquired Jif from P&G in 2001. In addition to other Jif brands, such as Simply Jif, Jif to Go, and Jif Natural, J.M. Smucker Company also

[1] The other household panel market is Pittsfield, Massachusetts.

owns Santa Cruz Organic and its own peanut butter brand, Smucker's. Skippy belongs to Unilever, which also owns Skippy Natural and Skippy Super Chunk.

The third top selling brand, Private Label, is actually not a brand. It is known as a store brand or generic brand, and its market share increased over the period I analyzed (2008–2010).

Table 1.1: Top 10 Selling Brands

Rank	Brand	Parent Company	Market Share (%)	Cum Market Share (%)
1	Jif	J.M. Smucker Co.	31.02	31.02
2	Skippy	Unilever	28.25	59.27
3	Private Label	Private Label	16.39	75.66
4	Peter Pan	Conagra Foods, Inc.	7.87	83.53
5	Smucker's	J.M. Smucker Co.	5.17	88.71
6	Skippy Natural	Unilever	3.50	92.21
7	Skippy Super Chunk	Unilever	2.43	94.64
8	Smart Balance	Smart Balance, Inc.	1.13	95.77
9	Simply Jif	J.M. Smucker Co.	1.12	96.88
10	Holsum	Holsum Foods	0.51	97.39

Note: The data used are store level peanut butter product sales data from six grocery stores in the Eau Claire market, 2008–2010. The market share is based on the total number of peanut butter jars sold in the market.

1.2.3 Household Peanut Butter Purchases

2,367 households in the panel purchased at least one jar of peanut butter during the three year time period. Table 1.2 shows summary statistics of the total number of peanut butter jars each household purchased. Half of the panel purchased more than 11 jars, and 25% of them purchased more than 20 jars. The average number of jars purchased is 15.37. The table also shows summary statistics of the number of jars purchased on a single grocery shopping trip, conditional on purchase. Households frequently purchased multiple jars of peanut butters at a time. The mean is 1.57 jars, and more than 25% of the purchase occasions are multiple jar purchases.

Table 1.3 looks into multiple jar purchases in detail. 10,452 out of 22,973 purchase occasions, which is 46.88%, are multiple jar purchases. Most of the multiple jar pur-

Table 1.2: Summary Statistics of Number of Peanut Butter Jars Purchased

Variable	Mean	SD	Min	P25	P50	P75	Max
Num of Jars for Three Years	15.37	14.69	2	5	11	20	99
Num of Jars on a Trip	1.57	0.78	1	1	1	2	6

Note: The data used are household level panel data from the Eau Claire market between 2008 and 2010.

chases are of the exactly same peanut butter item, rather than two different items. The data show that households purchased two jars of peanut butter on a single trip 8,814 times, and 8,336 times out of that, the two jars have the same UPC. UPC is the finest way to define a product, which means two jars with the same UPC are identical. Similarly, 1,166 out of 1,955 times that households purchased three or more jars, the jars have the same UPC. These suggest that most of households purchased multiple jars at a time for large quantity rather than for variety.

Table 1.3: Household PB Multi-UPC and Multi-Jar Purchases on a Trip

	Number of UPCs			
Number of Jars	1	2	3	Total
1	12,521	0	0	12,521
2	8,336	478	0	8,814
3+	1,166	456	16	1,955
Total	22,023	934	16	22,973

Note: The data used are household level panel data from the Eau Claire market between 2008 and 2010.

1.3 Quantity Surcharge at Stores

In this section, I study quantity surcharges at stores. I first define a product and identify products from data. Then I use the definition of a product to define quantity surcharges and measure frequency and magnitude of quantity surcharges. Quantity surcharges are more frequently observed than quantity discounts. Also, the price gap between sizes is bigger in quantity surcharge weeks than in quantity discount weeks.

1.3.1 Identification of Products

Defining a product is an important pre-step to define quantity surcharges. I define a product as a group of items that are identical except for the package size. It is critical to control the product characteristics among the items except for the package size. Otherwise, we cannot separate quantity surcharges from price differences due to product differentiation.

I identify products in the peanut butter category using the definition. First I combine very similar brands together. As shown in Table 1.1, Skippy Natural and Skippy Super Chunk are sister brands to Skippy, and it is natural for consumers to consider them as Skippy brand with different characteristics. Hence I merge the three Skippy sister brands into one brand Skippy, and do the same for Jif and Peter Pan [2]. Then I focus on the four leading brands: Skippy, Jif, Private Label, and Peter Pan. Lastly, I group UPCs with the same observable characteristics (texture, flavor, salt contents, sugar contents, and process) except for the package size.

The list of 17 products identified is presented in Table 1.4. 12 of them have a single package size, and five of them have two package sizes. A small package size ranges from 15 to 18 oz, and the large one is 28 oz for all products. The market share of 17 products combined is 86.26%.

1.3.2 Identification of Quantity Surcharge

Suppose a product with two different package sizes, small (S) and large (L). The product is quantity surcharged if

$$p_S < p_L, \tag{1.3.1}$$

where p_S and p_L are prices of the small and large size packages, respectively, normalized to their package sizes. Quantity discounts exist if the opposite holds: $p_S > p_L$.

In order to study quantity surcharges, I focus on the five products that have multiple

[2] I combine Jif, Jif Natural, and Simply Jif as "Jif", and Peter Pan, Peter Pan Plus, and Peter Pan Smart Choice as "Peter Pan".

Table 1.4: Products Identified

Item ID	Prod ID	Brand ID	Oz	Jars	Share	Ave Price (per 16 oz)	Feature Freq.	Display Freq.	Creamy	Flavor	Salt	Sugar	Natural	Red Fat
1	1	1	16.3	6,433	0.1831	2.1784	0.0940	0.2062	1	0	1	1	0	0
2	1	1	28.0	983	0.0280	2.3241	0.0353	0.0203	1	0	1	1	0	0
3	2	1	16.3	2,225	0.0633	2.1521	0.0876	0.1859	0	0	1	1	0	0
4	2	1	28.0	387	0.0110	2.3385	0.0214	0.0160	0	0	1	1	0	0
5	3	1	16.3	397	0.0113	2.2404	0.0865	0.0171	1	1	1	1	0	0
6	4	1	16.3	356	0.0101	2.2097	0.0630	0.0160	0	1	1	1	0	0
7	5	1	16.3	1,416	0.0403	2.2723	0.0940	0.0128	1	0	1	1	0	1
8	6	1	16.3	498	0.0142	2.2802	0.0812	0.0118	0	0	1	1	0	1
9	7	1	15.0	708	0.0202	2.5034	0.0406	0.0064	1	0	1	1	1	0
10	8	1	15.0	445	0.0127	2.5302	0.0491	0.0043	0	0	1	1	1	0
11	9	2	18.0	4,933	0.1404	2.0600	0.0630	0.1827	1	0	1	1	0	0
12	9	2	28.0	1,362	0.0388	2.1085	0.0192	0.0107	1	0	1	1	0	0
13	10	2	18.0	1,342	0.0382	2.0755	0.0534	0.1026	0	0	1	1	0	0
14	10	2	28.0	240	0.0068	2.1162	0.0203	0.0118	0	0	1	1	0	0
15	11	2	18.0	1,213	0.0345	2.1160	0.0662	0.0214	1	0	1	1	0	1
16	12	2	17.3	430	0.0122	2.2638	0.0299	0.0021	1	0	0	0	0	0
17	13	3	18.0	2,204	0.0627	1.6610	0.0427	0.2746	1	0	0	1	0	0
18	14	3	18.0	308	0.0088	1.5364	0.0000	0.0064	1	0	1	1	0	0
19	14	3	28.0	350	0.0100	1.6844	0.0075	0.0011	1	0	1	1	0	0
20	15	3	18.0	663	0.0189	1.6880	0.0231	0.1244	0	0	1	1	0	0
21	16	4	16.3	2,556	0.0728	2.0596	0.0278	0.1357	1	0	1	1	0	0
22	17	4	16.3	852	0.0243	2.0970	0.0246	0.0534	0	0	1	1	0	0

Note: Brand ID is 1 if Skippy, 2 if Jif, 3 if Private Label, and 4 if Peter Pan.

package sizes as shown in Table 1.4. Both Skippy (Product 1 and 2) and Jif (Product 9 and 10) have two products with multiple package sizes, one with creamy texture and another with chunky texture. Private Label also has one product with two package sizes, Product 14. Another leading national brand, Peter Pan, is not identified as a product with multiple package sizes, as the large size items have minimal market shares.

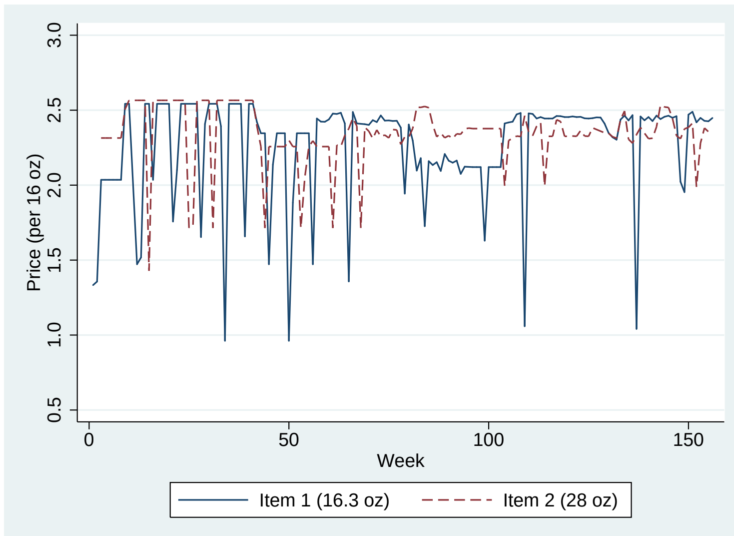
Quantity surcharges are determined by the price dynamics between the small and large size items. Figure 1.1 provides an illustrative example of how the price difference between sizes varies over time and how quantity surcharges are identified accordingly. In Figure 1.1a, the solid line represents the small size item (Item 1) and the dash line represents the large size item (Item 2). The two items belong to Product 1 and the prices are normalized to 16 oz.

Quantity surcharges exist whenever the dashed line stays above the solid line. That is equivalent to those weeks where the price difference is positive in Figure 1.1b. Price difference is defined as price of the small size item subtracted from price of the large size item. Quantity discounts exist whenever the dashed line is below the solid line, or equivalently, the price difference is negative. The pricing alternates frequently between quantity surcharges and quantity discounts in the earlier weeks. Quantity surcharges last for a while afterwards, and then mostly quantity discounts in the later weeks.

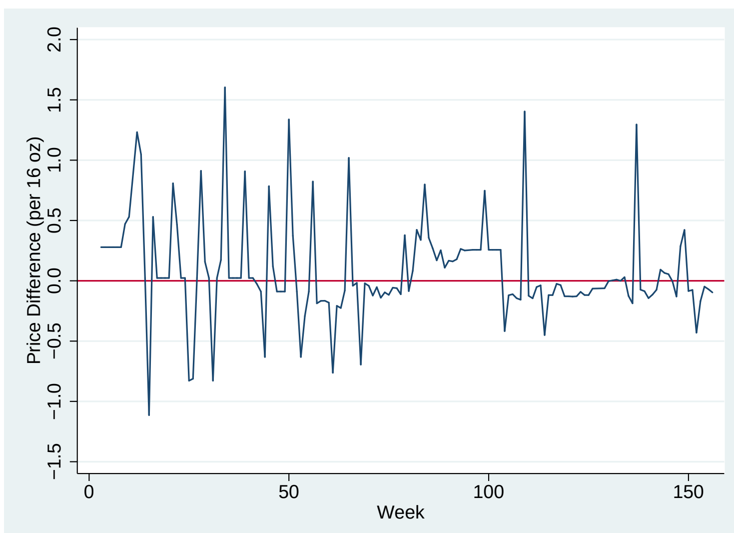
The price gap between the two sizes, which determines the magnitude of quantity surcharges, fluctuates over time. The small size item shows more frequent and deeper price drops, especially during the first 70 weeks, and those drops increase the price gap. Around week 70 to 100, the price gap stays at a somewhat moderate level. In later weeks, the gap becomes slim, except for the two big spikes of quantity surcharges.

For each of the five products, I obtain the frequency of quantity surcharges and quantity discounts as follows: 1) at each store, I determine whether the product is quantity surcharged or quantity discounted in each week; 2) I sum up the number of quantity surcharge weeks and quantity discount weeks across stores; 3) I divide both numbers by the total number of weeks. Note that I exclude the weeks that sales information of either

Figure 1.1: Price of Product 1 at Store 2



(a) Price per 16 oz of Product 1 at Store 2



(b) Price Difference Between Two Items of Product 1 at Store 2

Note: Panel (a) shows weekly prices (per 16 oz) of the two items, Item 1 and Item 2 and panel (b) shows the price difference between the two items (price of Item 1 subtracted from price of Item 2). The two items belong to Product 1 identified in Table 1.4. The price information is obtained from Store 2, one of the six grocery stores in the market, from 2008 to 2010.

size is missing at store level. The results are presented in Table 1.5.

Table 1.5: Frequency and Magnitude of Quantity Discounts and Quantity Surcharges

	QD ($p_S > p_L$)		QS ($p_S < p_L$)	
Prod ID	% of weeks	Ave $p_L - p_S$ ($)	% of weeks	Ave $p_L - p_S$ ($)
1	36.38	-0.1807	63.64	0.3105
2	36.62	-0.1754	63.38	0.3577
9	44.95	-0.1387	55.05	0.1998
10	47.99	-0.1440	52.01	0.2109
14	13.29	-0.4364	86.71	0.1785
Total	38.06	-0.1680	61.94	0.2589

Note: A product is defined as a group of UPCs with the same observable characteristics but package size. The five products are the products identified in Table 1.4 that has two different package sizes, small and large; p_S and p_L are the price per 16 oz of the small and large size item, respectively. A quantity discount (QD) exists for a product at a certain week at a store if $p_S > p_L$ holds, and a quantity surcharge (QS) exists if the opposite ($p_S < p_L$) holds. % of weeks is the percentage of weeks at the six grocery stores combined when quantity discounts or quantity surcharge existed; Ave $p_L - p_S$ is the average price difference per 16 oz between the small and the large items across weeks and stores. It measures the average magnitude of quantity discounts and quantity surcharges. The difference is negative in quantity discount weeks and positive in quantity surcharge weeks.

It shows that quantity surcharges are more frequent than quantity discounts across products. The two Skippy products (Product 1 and 2) have quantity surcharges for around 63% of the weeks. Quantity surcharges are less frequent for the two Jif products (Product 9 and 10), compared to the Skippy products, but still more frequent than quantity discounts.

Product 14 shows the highest frequency of quantity surcharges among the five products, which goes up to 87%. It can be explained by the fact that Product 14 belongs to Private Label. Private Label is a store brand, and stores might have more direct influence on prices of their own brands. Using this ability, stores might have wanted to keep the price of the small size item, which is more popular than the large size item, at a low level.

Large size items are more expensive by $0.20 to $0.36 per 16 oz than small size items in quantity surcharge weeks. The third and the last columns in Table 1.5 show the average magnitudes of quantity discounts and quantity surcharges of each product. I subtract the price per 16 oz of a small item from the price per 16 oz of a larger counterpart to calculate the magnitudes.

The magnitude of quantity surcharges is bigger than that of quantity discounts, except for Product 14. It is also true when the five products are combined together. Again it can be explained by the fact that Product 14 is owned by stores. Among the four national brand products, the two Skippy products (Product 1 and 2) show bigger magnitudes of quantity surcharges than the two Jif products (Product 9 and 10).

1.3.3 Quantity Surcharge and Promotions

I analyze the relationship between quantity surcharges and promotional activities stores practice in this section. There are two types of promotional activities. The first type is a price promotion, also known as sale or temporary price reduction. The second type is a non-price related promotion, such as feature and display.

I follow Hendel and Nevo (2003)'s approach and define a sale for an item as follows: I first define the regular price as the modal price, which is the most frequent price for an item at a store. Then sale is defined as any price at least 5% lower than the regular price.

Since there are two different package sizes for each product, there are four possible sale states: only the small size item is on sale, only the large size item is on sale, both items are on sale, and none of them is on sale. Table 1.6 shows the frequency of each sale state across the five products. Roughly a half of the time at least one item was on sale, and the three sale states (small only, large only, and both) show almost an equal share.

Table 1.6: Frequency of Promotion Activities

Promotion Activity	Small Only	Large Only	Both	None
Sale	14.62	16.62	14.90	53.86
Feature	6.05	1.87	0.44	91.64
Display	14.77	1.38	0.03	83.82

Note: A sale is defined as a price at least 5% less than the modal price at each grocery store for the period analyzed.

Quantity surcharges are somewhat related to a temporary price reduction on small size items, but that is not the main cause. Table 1.7 describes the relationship between quantity surcharges and sale. It shows that quantity surcharges are most common when neither package size is on sale. The second most frequent state is sale on the small size item only, except for Product 10. It takes 23.05% when all five products are combined together. Therefore quantity surcharges are more of consistent phenomena for peanut butter products than side effects of sale on small size items.

Table 1.7: Quantity Surcharges and Sale on Each Package Size

Prod ID	Small Only	Large Only	Both	None
1	33.09	7.01	12.95	46.94
2	27.49	11.75	13.35	47.41
9	24.26	3.55	23.47	48.72
10	19.03	20.13	29.20	31.64
14	8.29	0.00	1.51	90.20
Total	23.35	8.57	16.40	51.68

Note: A sale is defined as a price at least 5% less than the modal price at each grocery store for the period analyzed.

Another way to analyze the relationship between quantity surcharges and sale is to compare the frequency of quantity surcharges between the weeks when a small size item is on sale (sale week) and the weeks when a small size item is not on sale (regular week). Table 1.8 displays the comparison. During the sale weeks, quantity surcharges strictly dominate quantity discounts, ranging from 74% of the weeks for Product 1 to 100% for Product 14. During the regular weeks, depending on products, quantity surcharges are more or less frequent than quantity discounts. When you combine the five products together, quantity surcharges are slightly more frequent than quantity discounts.

Feature and display are non-price activities stores practice to promote sales. An item is featured when a store lists an advertisement for the item on a flyer. Display means placing an item at a specific location of a store, such as a lobby and end of an aisle, to make it easier for consumers to notice. As shown in Table 1.6, these two promotions are far less frequent than sale and mostly focused on small size items.

It is somewhat challenging to analyze the relationship between quantity surcharges and feature, as the number of feature observations is small. Even when a product is featured, it is usually the small size item. This pattern is also shown in Table 1.9. During

Table 1.8: Quantity Surcharges in Regular Weeks and Sale Weeks

	Regular Weeks		Sale Weeks		Total	
Prod ID	QD	QS	QD	QS	QD	QS
1	42.53	57.47	26.22	73.78	36.02	63.75
2	42.88	57.12	21.46	78.54	35.72	63.64
9	60.21	39.79	5.10	94.90	44.95	55.05
10	62.26	37.74	12.45	87.55	47.99	52.01
14	14.52	85.48	0.00	100.00	13.29	86.71
Total	47.05	52.95	16.59	83.41	38.06	61.94

Note: Sale weeks are the weeks when a small size item of a product is on sale, and regular weeks are the weeks when the small size item is not on sale.

85 to 91% of the quantity surcharge weeks, neither size is featured. On top of that, there is no feature activity during quantity surcharge weeks for Product 14. When five products are combined, the frequency of each feature event during quantity surcharge weeks is very similar to the frequency of each event in overall, as shown in Table 1.6.

Table 1.9: Quantity Surcharges and Feature on Each Package Size

Prod ID	Small Only	Large Only	Both	None
1	13.85	0.54	0.72	84.89
2	12.75	0.60	0.60	86.06
9	10.45	0.00	0.00	89.55
10	8.63	0.00	0.00	91.37
14	0.00	0.00	0.00	100.00
Total	9.65	0.25	0.29	89.81

Note: Feature is a non-price promotional activity listing advertisements of items in the flyer.

Table 1.6 shows that display activities are still less frequent than sale but more frequent than feature. Table 1.10 describes the relationship between quantity surcharges and display. Similar to feature, the majority of the quantity surcharge weeks do not overlap with any display activities. The proportion of each display event in quantity surcharge weeks also looks very similar to the proportion of each event in overall weeks as shown in Table 1.6. Thus, unlike sale, feature and display have no clear relationship with quantity surcharges.

Table 1.10: Quantity Surcharges and Display on Each Package Size

Prod ID	Small Only	Large Only	Both	None
1	26.98	0.18	0.18	72.66
2	22.91	0.60	0.00	76.49
9	26.04	0.20	0.00	73.77
10	16.81	0.22	0.00	82.96
14	1.26	0.00	0.00	98.74
Total	19.79	0.25	0.04	79.92

Note: Display is a non-price promotional activity placing items at specific locations of the store to make them more noticeable.

1.4 Quantity Surcharges and Heterogeneous Household Behavior

In this section I study household purchasing behavior when quantity surcharges exist. A small size jar is cheaper than its large size counterpart in quantity surcharge weeks. If all households were aware of the substitution opportunity, the two following purchasing patterns would be expected during quantity surcharge weeks: 1) no large size jar purchase, and 2) more frequent multiple small size jar purchases compared to quantity discount weeks.

In other words, households would purchase small size jars only, and all the households who would demand a large quantity must have purchased multiple small size jars. Comparing the actual household purchasing behavior observed in the data to those hypothetical purchasing behavior can tell us how much households are aware of quantity surcharges. Here I assume no preference towards large size packages and no stock-outs of small size items. Those two possibilities are discussed later in this section.

1.4.1 Household Multiple Jar Purchases

Table 1.3 shows how often households purchase multiple jars of the same UPCs. In order to test households' awareness to quantity surcharges, I analyze how multiple jar purchases of each size vary during quantity discount and surcharge weeks. The results are presented in Table 1.11. A purchase is "single" when a household purchases a single jar on a trip, and "multi" when it is more than one jar of the same item.

Table 1.11: Household Multiple Jar Purchases in QD and QS Weeks

	Small			Large		
	Single	Multi	Total	Single	Multi	Total
QD($p_S > p_L$)	571	922	1,493	513	560	1,073
QS($p_S < p_L$)	2,981	3,786	6,767	598	261	859
Total	3,552	4,708	8,260	1,111	821	1,932

Note: "single" represents single jar purchases and "multi" represents multiple jar purchases of the same item on a trip.

Household purchasing patterns suggest that there is heterogeneity in households' awareness to quantity surcharges. Multiple small size jar purchases are roughly four times more frequent in quantity surcharge weeks. This implies that some households are aware of the existence of quantity surcharges and choose to buy multiple small size jars rather than a large size jar.

The number of single small size jar purchases jumps up in quantity surcharge weeks compared to quantity discount weeks also. This could be caused by the fact that small size items are on sale roughly 40% of the times (including sale on small size item only and sale on both sizes) during quantity surcharge weeks, as shown in Table 1.7.

1.4.2 Household Miss Purchases

On the other hand, there is a positive number of purchase observations of large size items during quantity surcharge weeks in Table 1.11. This indicates that some households are not aware of quantity surcharges and miss the substitution opportunities. This finding is consistent to the concept of consumer inattention that Clerides and Courty (2017) argue: there exist inattentive households who are unaware of the existence of quantity surcharges and miss the opportunities to substitute. From now on I call purchasing large size items in quantity surcharge weeks as a "miss".

Almost 20% of households make at least one miss. Table 1.12 shows the number of misses each household makes. There are 2,013 households who purchase at least one item of the multiple package size products. 460 households make at least one miss purchase during the time period analyzed. This number takes 22.85% of the households who purchase at least one item of the five multiple size products, and 19.43% of the total households. The average number of misses is less than 1, as a large proportion of households makes zero misses.

Table 1.12: Number of Misses Each Household Make

Num of Misses	0	1	2	3	4	5
Frequency	1,553	272	101	38	22	9
Percent	77.15	13.51	5.02	1.89	1.09	0.45
Cum. Percent	77.15	90.66	95.68	97.57	98.66	99.11
Num of Misses	6	7	8	9	10	12
Frequency	6	6	1	2	1	2
Percent	0.30	0.30	0.05	0.10	0.05	0.10
Cum. Percent	99.40	99.70	99.75	99.85	99.90	100.00

Note: The summary statistics are based on the peanut butter purchases 2,013 households made for the three years. A miss is defined as an incident of purchasing a large size item in quantity surcharge weeks. The sample mean of the number of misses is 0.4267 and the standard deviation is 1.081.

What makes inattentive households inattentive? There could be several reasons. Some households might have a limited ability to calculate price per ounce at stores. High search costs could hurry some households and keep them from comparing prices between jar sizes. Clerides and Courty (2017) call this case rational inattention. Some households might have a strong belief that quantity discounts always exist. Those reasons are not mutually exclusive.

There could be alternative explanations for those miss purchases other than inattention. One possible explanation is a strong preference on large size jars. However, it is hard to argue so when small size jars have relative advantages to the large ones: a smaller jar can keep peanut butter more fresh than a larger jar does, and it is also more convenient to carry around. I further discuss this possibility in Chapter 2.

Another alternative explanation could be a stock-out. That is, some households purchase large size items just because small size items are not available upon their visits.

Conlon and Mortimer (2013) argue that ignoring incomplete product availability may bias demand estimates. I cannot precisely verify this argument, as the store level data are recorded at weekly level.[3] However, the five small size items that I analyze are more popular than most of the other UPCs on the market, which makes them relatively less likely to be stocked out at stores. Hence, it is hard to argue that strong preference on large size jars or stock-outs of small size items is the main reason for those miss purchases.

1.4.3 Miss Purchases and Non-Price Promotions

Some people might argue that households make miss purchases because the large size items are specially featured at advertisements or displayed at very noticeable locations at stores such as the end aisle. Table 1.13 presents how the frequency of miss purchases changes when the large size items are on non-price promotions or not.

Table 1.13: Miss Purchases and Non-Price Promotions on Large Size Jars in QS weeks

	Feature on Large Size		Display on Large Size	
	No	Yes	No	Yes
Misses (%)	11.33	6.32	11.14	46.15
Purchase Obs	7,531	95	7,600	26

Note: A miss is defined as an incident of purchasing a large size item in quantity surcharge weeks. Feature is a non-price promotional activity listing advertisements of items in the flyer. Display is a non-price promotional activity placing items at specific locations of the store to make them more noticeable.

Households make misses less frequently when large size items are featured compared to those weeks when large size items are not featured. Thus, the conjecture that features on large size items encourage households to make miss purchases is not supported. However feature is fairly rare as shown in Table 1.6. The number of purchase observations during the weeks when large size items are featured is just 95, compared to 7,531 purchase observations during the weeks when large size items are not featured. Hence it is hard to argue that the frequency difference is significant.

[3] The household level data have information on checkout time. However, this information allows me to observe product availability only when there exists a household who purchased the item of interest.

The frequency of misses rises up to 45.15% when large size items are displayed compared to 11.14% when large size items are not displayed. However, There are only 26 observations of large size purchases when large size items are displayed. Thus it is hard to argue that display on large size items make households more likely to miss either.

In this section, I study households' heterogeneous behavior during quantity surcharge weeks. I explain the heterogeneity using the concept of consumer inattention. Inattentive households are unaware of the existence of quantity surcharges and purchase large size jars. On the other hand, attentive households are aware of the existence of quantity surcharges and purchase multiple small size jars instead of large size jars when they want to purchase a large quantity.

1.5 Characteristics of the Households Who Make Miss Purchases

In this section, I study the characteristics of households who miss the substitution opportunities in quantity surcharge periods. I design a negative binomial regression model where the number of misses each household makes is explained by the three groups of household characteristics: demographics, grocery shopping, and peanut butter purchases. I estimate the model using the method of maximum likelihood.

1.5.1 Model

Let y_i be the number of miss purchases household i makes and x_i be a vector of the household's characteristics. Assume y_i given x_i follows the negative binomial distribution with conditional mean and variance

$$\begin{aligned} \mathrm{E}[y_i|x_i] &= \mu_i = \exp(x_i'\beta) \\ \mathrm{V}[y_i|x_i] &= \mu_i + \alpha\mu_i^2, \end{aligned} \tag{1.5.1}$$

where β is a vector of mean parameters and α is a dispersion parameter. The marginal effect of the j^{th} characteristic of household i, x_{ij}, to the conditional mean is obtained as

$$\frac{\partial \mathrm{E}[y_i|x_i]}{\partial x_{ij}} = \beta_j \exp(x_i'\beta). \tag{1.5.2}$$

Negative binomial regression model is commonly used for count data. The model is well suited to the household miss purchase data, as the number of misses is a non-negative integer ranging from 0 to 12. The model allows the two sets of parameters, mean parameters and a dispersion parameter. Poisson model is also widely used for count data for its simplicity. The model assumes Poisson distribution, which restricts the mean and variance of the data to be the same. This equidispersion assumption allows only one set of parameters to estimate, the mean parameters.

However, overdispersion is common for count data (Cameron and Trivedi (2013)). The number of misses data also show strong evidence of overdispersion. The unconditional mean is 0.4269 and the variance is 1.081, as shown in Table 1.12. The mean is smaller than the variance as roughly 77% of the households have zero misses. The variance is actually more than twice as big as the unconditional mean, and it is likely to remain overdispersed even after controlling the household characteristics. Thus, negative binomial model is better suited for the data.

1.5.2 Estimation

The conditional mean of household i's number of misses, $\mathrm{E}[y_i|x_i]$, is specified as

$$\mathrm{E}[y_i|x_i] = \exp(x_i'\beta) = \exp(\beta_0 + \beta_D \mathrm{Demo}_i + \beta_G \mathrm{GS}_i + \beta_P \mathrm{Pur}_i), \tag{1.5.3}$$

where Demo_i represents a vector of the household's demographic characteristics, GS_i represents a vector of the grocery shopping characteristics, and Pur_i represents a vector of the peanut butter purchase characteristics. Note that the log of the conditional mean is linear in the household characteristics.

The data used for the estimation include 2,013 households who purchase at least one item of the five products with multiple package sizes. The detailed information on demographics, grocery shopping trips, and peanut butter purchases of those households are presented in Appendix A.2.

The list of household characteristics variables and summary statistics are presented in Table 1.14. The first group of characteristics is demographics. The data provide information on each household's income, family size, number of children, and the head of household's age, education level, and race. I drop race since more than 95% of the households belong to the same racial group. Family size and number of children are highly correlated. In order to avoid multicollinearity, I drop the number of children.

Table 1.14: Variable Definitions and Summary Statistics

Variable	Definition	Obs	Mean	SD
INCOME	Combined pre-tax income of HH	2,013	7.37	3.16
FAMSIZE	Number of family members in a household	2,013	2.40	1.24
HHAGE	Age of the head of household	2,008	4.78	1.14
HHEDUC	Education level reached by HH	1,979	4.20	1.33
ANTRIPS	Average number of trips to stores in a week	2,013	1.45	0.84
AEXPD	Average expenditure on a grocery trip	2,013	42.01	22.66
NJARS	Number of PB jars purchased for the three years	2,013	13.33	12.62
RVAR	Number of UPCs purchased divided by the number of brands purchased for the three years	2,013	0.74	0.22

Note: HH stands for the head of household.

The second group of variables is the grocery shopping characteristics. ANTRIPS is to test whether the households who go to grocery shopping more often make more misses. Households might collect better information on price distribution among peanut butter products when they go visit stores more often, and become less likely to miss the substitution opportunities. The households who spend more money per grocery shopping trip might purchase a larger number of items at a time. This could lengthen the total time spent at a store, and the households might pay little attention on each item, including peanut butter. AEXPD is included to test that conjecture.

The last group consists of the peanut butter purchase characteristics. NJARS tests a conjecture that is very similar to the one with ANTRIPS: the more peanut butter jars households purchase, the better knowledge they have on pricing and less likely to make miss purchases. The number of UPCs and the number of brands purchased in Table A.4 together show how much variety each household seeks. However, they are closely correlated by nature. I take a ratio of the two variables, which represents the relative variety, in order to avoid multicollinearity.

There are two sets of parameters to estimate in the model, the dispersion parameter α, and the mean parameter $\beta = (\beta_0, \beta_D, \beta_G, \beta_P)$. I estimate the model using the method of Maximum Likelihood. The negative binomial model is robust to distributional misspecification and hence the maximum likelihood estimator is consistent for β.

1.5.3 Estimation Results

The maximum likelihood estimation results are presented in Table 1.15. The NB Model 1 includes all the explanatory variables presented in Table 1.14. Note that some households have missing information in the head of the household's age and/or education level. Hence the estimation is based on the subsample of 1,979 households who have complete demographic information.

However, the estimation results suggest that those two variables, HHAGE and HHEDUC, are not statistically significant. The likelihood ratio test also suggests that the two variables are not informative[4]. Therefore I exclude the two variables in NB Model 2 to utilize the larger size sample of 2,013 households.

The estimation results of NB Model 2 suggest that, both of the demographics variables are not significant. INCOME and FAMSIZE have positive coefficients of 0.0230 and 0.0670, respectively, but none of them are statistically significant. Thus, once the other household characteristics are controlled for, income nor family size has significant effect on the number of miss purchases each household makes.

There are two grocery shopping characteristics variables, ANTRIPS and AEXPD.

[4]The null hypothesis that the two likelihoods with or without the two variables are the same cannot be rejected with the p-value 0.4849

Table 1.15: Negative Binomial Maximum Likelihood Estimation Results

	NB Model 1		NB Model 2		Mean Effect	
Variable	Coeff	SE	Coeff	SE	Ave Resp	OLS
INCOME	0.0229	0.0174	0.0230	0.0172	0.0102	0.0065
FAMSIZE	0.0787	0.0528	0.0670	0.0447	0.0298	0.0161
HHAGE	0.0378	0.0585				
HHEDUC	0.0419	0.0392				
ANTRIPS	0.1102	0.0719	0.1205	0.0717	0.0536	0.0549
AEXPD	0.0167	0.0027	0.0176	0.0027	0.0078	0.0087
NJARS	0.0045	0.0046	0.0049	0.0046	0.0022	0.0033
RVAR	-2.4817	0.2693	-2.4583	0.2678	-1.0945	-0.9433
N	1979		2013			

Note: Mean effects are calculated based on NB Model 2.

ANTRIPS is not statistically significant, and hence, the conjecture that the frequent shoppers make less miss purchases is not supported. AEXPD has a positive coefficient of 0.0176 that is statistically significant. Hence, it supports the conjecture that the households with bigger baskets make more miss purchases. However, this coefficient estimate doesn't directly mean the marginal effect of average expenditure on the number of misses. In the negative binomial regression model, the log of the conditional mean is linear in the explanatory variables, not the conditional mean itself. Thus, the point estimates do not represent the marginal effects.

The marginal effect of the j^{th} characteristic of household i on the conditional mean is presented in the equation (1.5.2). The average response across households can be obtained by taking the average of the marginal effects. The Ave Resp column in Table 1.15 shows the average response for each household characteristic, calculated based on the NB Model 2 estimation results. It shows that one unit increase in the average expenditure on a grocery shopping increases the number of misses by 0.0078.

The OLS column shows the OLS coefficient estimates, assuming that the conditional mean of the number of misses is linear in the household characteristics. The coefficient estimate directly implies the effect of the one unit change in an explanatory variable on the conditional mean. The OLS coefficients are similar to the average responses, but still not the same. For example, the OLS model predicts that one unit change in average expenditure increases the number of misses by 0.0087, instead of 0.0078.

The last group of variables are the peanut butter purchase characteristics. NJARS is not statistically significant and the information conjecture is not supported again, as AEXPD failed. On the other hand, the relative variety measure, RVAR, is highly significant and has a negative effect on the number of miss purchases. One unit increase in the relative variety decreases the number of misses by 1.0945.

1.6 Concluding Comments

I study quantity surcharges at grocery stores and heterogeneous consumer behavior in this chapter. Quantity surcharges are often forgotten when packaged goods and nonlinear pricing are of interest. I reconfirm the existence of quantity surcharges using the peanut butter category scanner data. Quantity surcharges are more frequent than quantity discounts, and the price gap between the two sizes is also bigger in quantity surcharge weeks than in quantity discount weeks.

This chapter has three main contributions. The first contribution is to prove the existence of quantity surcharges as more of a consistent phenomenon than a mere side effect of sale. Quantity surcharges are more frequent when small size items are on sale compared to those weeks when small size items are not on sale. However, quantity surcharges triggered by sale on small size jars only take less than a quarter of the total quantity surcharge occasions.

The second contribution is to identify heterogeneity in consumer attention to quantity surcharges. The current literature focuses on the existence of inattentive consumers who purchase large size items during quantity surcharge weeks. However, taking advantage of the rich household panel data available, I also identify attentive consumers, who actively respond to quantity surcharges and purchase multiple small size jars.

The last contribution is to identify the characteristics of the households who make miss purchases. I use the property of the number of misses as a count variable and design a negative binomial regression model. Three groups of household characteristics are considered to explain the number of misses: demographics, grocery shopping, and peanut butter purchases. I estimate the model using the method of maximum likelihood.

Estimation results suggest that households with large expenditure on each grocery shopping make more miss purchases in quantity surcharge weeks. Households who purchase a variety of peanut butters make less miss purchases. Households who purchase peanut butters often don't necessarily make more misses, once the other variables are controlled for.

Chapter 2

Why Do Consumers Pay More for Less?

2.1 Introduction

In this chapter, I study demand models for packaged goods when quantity surcharges exist. Clerides and Courty (2017) argues that consumer inattention explains why consumers purchase large size items and pay the surcharges in quantity surcharge periods. Consumers don't pay attention when they shop, and as a result, they miss the substitution opportunities. In Chapter 1, I label the purchasing behavior as a "miss" purchase. Chapter 1 also finds that some consumers take advantage of the substitution opportunities and purchase multiple small size jars in quantity surcharge weeks. Please see Section 1.1 for the definition and examples of quantity surcharge.

Based on the heterogeneous purchasing behavior found in Chapter 1, I develop a series of structural demand models and estimate the models using Bayesian methods. Then I simulate demand using the estimation results and calculate the welfare loss caused by inattention. I use the rich panel data set from grocery stores in the peanut butter category for the analyses. Please see Section 1.2 for further details on the data.

To understand what causes miss purchases, I study a series of demand models. The first model assumes standard preferences and rationality. Then I explore two extensions of this basic model. First extension explains preference on package sizes as the motivation. Second extension considers consumer inattention. It restricts the choice set where consumers choose the optimal quantities from in case of miss purchases. All three models yield a complex optimization problem as they feature multiple discrete choices. I solve this problem using Allenby, Shively, Yang, and Garratt (2004)'s two stage optimization approach.

I estimate the models using the Markov Chain Monte Carlo (MCMC) methods. The likelihood function to maximize is complicated as the models include two stage optimizations, and the number of elements in the choice set is large. The Bayesian methods have been shown to be more robust in this setting.

The estimation results support the model with consumer inattention. The model provides the predicted demand that is closest to the actual demand observed, compared to the other two model specifications. This result suggests that consumer inattention explains households' heterogeneous purchasing behavior well, and omitting it will lead to biased predictions.

An effective consumer awareness program, such as a clear per unit price display requirement, can prevent consumers from making miss purchases. I measure the effect of the success of such a program, as a welfare loss coming from consumer inattention. I design a hypothetical scenario where all households are attentive to approximate a successful consumer awareness program. The consumer inattention model is used as a benchmark.

The structural model makes it possible to simulate purchase outcomes that households would choose if they were attentive. Then I measure the welfare loss as the additional income needed to compensate the utility loss from a miss purchase. The average welfare loss of a miss purchase is \$0.69, which is substantial considering the average price of a small size jar is roughly \$2.

Clerides and Courty (2017) explain rational inattention as the main reason why consumers become inattentive. That is, it can be too costly for some consumers to pay attention when they shop, and as a result, it is rational for them not to pay attention. The notion of rational inattention is developed by Sims (1998), Sims (2003), Reis (2006), Mackowiak and Wiederholt (2009), and Mondria et al. (2010).

Behavioral studies of information opacity find empirical evidence of limited attention. Chetty, Looney, and Kroft (2009) and Hossain and Morgan (2006) consider consumer goods and show that it is difficult for consumers to process the information when it is opaque. However, there is no opacity of price information when consumers shop peanut butter items at grocery stores. Cognitive research on decision-making in grocery purchases also provides evidence of limited attention. Monroe and Lee (1999) and Dickson and Sawyer (1990) show that consumers do not remember the prices of items they recently purchased, or don't even check the exact price of the item selected at the time of purchase.

One of the key purchasing patterns in the peanut butter category is frequent multiple jar purchases. Hence incorporating the multiple discrete choice in the model is crucial for the analysis. Several papers have tried to handle the multiple discrete choice problem. Manchanda et al. (1999), Wygant et al. (2000), Chib et al. (2002), and Lee et al. (2013) studies multiple category purchase decisions. Harlam and Lodish (1995) proposes a choice of model of multiple brands but not the purchase quantity decisions.

Hendel (1999) suggests a model which allows both multiple units and multiple brands. Kim, Allenby, and Rossi (2002), Dubé (2004), and Dubé (2005) also take similar approaches to handle the multiple discrete choice. All four papers mentioned above consider that consumers buy complementary products to cater for their various needs.[1] Hence the authors cannot handle a decision whether or not to purchase multiple small size jars for a large quantity.

Chintagunta (1998), Chang et al. (1999), Silva-Risso et al. (1999), and Andrews and Manrai (1999) deal with packaged goods when nonlinear pricing exists, but the au-

[1] Especially Dubé (2005) considers items with the same characteristics but different package sizes as complements, whereas I consider them as perfect substitutes.

thors consider different brand/size combinations as different choice alternatives. Allenby, Shively, Yang, and Garratt (2004) introduces a multiple discrete choice model for packaged goods. In the model, consumers consider different package size combinations within brand as substitutable alternative, and purchase multiple items for a larger quantity, not for variety. However, the authors assume quantity discounts only and fail to analyze consumer inattention when quantity surcharges exist.

The chapter proceeds as follows. Section 2 describes the data used for the analyses and Section 3 introduces a series of structural demand models for packaged goods. Section 4 describes the estimation procedure, and section 5 shows the estimation results. Section 6 applies the estimation results to calculate the welfare loss of consumer inattention, and Section 7 concludes the paper.

2.2 Data

I use the weekly panel scanner data collected from Eau Claire, Wisconsin, market, from January 2008 to December 2010. The data is collected by Information Resources Inc. (IRI) and I focus on the peanut butter category. Three household level data sets - peanut butter purchases, trips to grocery stores, and demographics - are merged with product attributes data and store level sales data. The observations that are not matched in the process are dropped. For further details, please see Section 1.2 and Appendix A.1.

I use the same definitions of a product and quantity surcharges as in Chapter 1. 17 products of four brands are identified as shown in Table 1.4. 5 products offer multiple package sizes (small and large), and quantity surcharges are identified for those weeks when a large size item is more expensive per oz than the small size counterpart. Quantity surcharges exist for around 62% of the weeks on average, and the average magnitude of quantity surcharges is bigger than that of quantity discounts. Please see Table 1.5 for details.

The merged data set includes 2,367 households with 23,287 purchase observations of 17 products identified. During quantity surcharge weeks, households show frequent multiple small size jar purchases. On the other hand, roughly 20% of households make at least one miss purchase. Therefore, it is important for a demand model to capture these heterogeneous purchasing behaviors.

2.3 Model

In this section, I present a series of three discrete choice models for packaged goods. I start with a basic model, and then extend the model in two different ways to capture a key purchasing pattern found from the data.

2.3.1 Basic Model

The basic model is adopted from Allenby, Shively, Yang, and Garratt (2004). A consumer has the Cobb-Douglas (C-D) utility function

$$\ln u(x,z) = \alpha_0 + \alpha_x \ln u(x) + \alpha_z \ln(z), \tag{2.3.1}$$

such that she enjoys both the inside good of our interests and an outside good. $x = (x_1, \ldots, x_K)$ is a vector of the amount of each inside good product consumed, where K is the number of products available. Products are differentiated in characteristics, and some of them have multiple package sizes. z represents the amount of the outside good consumption, and $u(x)$ indicates a subutility function.

The subutility function has a linear structure:

$$u(x) = \psi\prime x, \tag{2.3.2}$$

where ψ_k denotes the marginal utility of product k. Let $\ln(\psi_k) = \nu_k + \epsilon_k$, where ϵ_k is a stochastic element. The non-stochastic factor ν_k is determined as

$$\nu_k = \beta_{0b}\text{brand}_k + \beta_c\text{char}'_k, \tag{2.3.3}$$

where brand_k represents the brand which product k belongs to, and char_k is a vector of characteristics of product k.

The consumer determines her consumption on each of K products and the outside good to maximizes her utility function in (2.3.1) subject to the budget constraint

$$\Sigma_{k=1}^{K} p_k(x_k) + z = T, \tag{2.3.4}$$

where $p_k(x_k)$ is the price of x_k units of product k, and T represents the consumer's budgetary allotment. Price of the outside good is one. Price of the inside good is a function of quantity which incorporates any kind of pricing schemes, including linear pricing, quantity discounts, and quantity surcharges.

The quantity choice of product k, x_k, is discrete, as products are offered in certain package sizes. Suppose product k is only available in 16 oz jars. Then x_k should be one of the multiples of 16, such as 16, 32, 48, ... oz. If the product is available in two package sizes, 16 and 28 oz, then x_k should be a combination of the two package sizes.

I assume consumers choose only one product at a time (i.e. only one element of x is nonzero), and evaluate the utility function at all possible combinations of package sizes for each product. Restricting the utility maximization solutions to corner solutions is a key assumption to make the evaluation feasible.

Allenby, Shively, Yang, and Garratt (2004) proves that the C-D utility maximization solution subject to a convex budget constraint is actually at a corner. A budget constraint is convex when quantity discounts exit. On the other hand, it is concave when quantity surcharges exit. Thus, the proof fails when both quantity discounts and quantity surcharges exit. However, the corner solution assumption is still supported by the data, as multiple product purchases are rarely observed.

The solution strategy involves two steps when only one element of x is nonzero. In the first step, the consumer determines the optimal quantity for each product separately. The optimal quantity is chosen from all possible combinations of package sizes available. I substitute $z = T - \Sigma_k p_k(x_k)$ for the outside good, using the budget constraint in equation (2.3.4), and the first stage optimization problem becomes

$$\begin{aligned}
&\max_{a\in A}\{\alpha_0+\alpha_x\ln u(x_{ka})+\alpha_z\ln(T-p_k(x_{ka}))\}\\
=&\max_{a\in A}\{\alpha_x\ln x_{ka}+\alpha_z\ln(T-p_k(x_{ka}))\},
\end{aligned}\tag{2.3.5}$$

where x_{ka} and $p_k(x_{ka})$ represent the quantity and the price of package bundle a of product k. A is the set of all possible combinations of package sizes available for product k. A depends on product k but here I drop the subscript k for convenience. Note that the stochastic factor of the log marginal utility, ϵ_k, cancels in the expression, as it is the same for any possible package bundles within a product. Hence, the optimal quantity of each product in the first stage is deterministic.

In the second step, the consumer decides which product to purchase. In the previous step, the consumer searches for the optimal quantity, given that product k is chosen. In the second step, the consumer lines up the optimal quantities of each product, and compares the utilities. The product choice problem can be written as:

$$\begin{aligned}
&\max_{ka}\{\ln u(x_{ka},T-p_k(x_{ka}))\}\\
=&\max_{k}[\max_{a|k}\{\ln u(x_{ka},T-p_k(x_{ka}))\}]\\
=&\max_{k}[\alpha_0+\alpha_x\ln u(x_k^*)+\alpha_z\ln(T-p_k(x_k^*))],
\end{aligned}\tag{2.3.6}$$

where x_k^* is the optimal quantity for product k in the first step.

Substituting the subutility expression in equation (2.3.2) yields

$$=\max_{k}[\alpha_0+\alpha_x(\nu_k+\epsilon_k)+\alpha_x\ln(x_k^*)+\alpha_z\ln(T-p_k(x_k^*))].\tag{2.3.7}$$

Assume ϵ_k follows the Type I extreme value distribution, EV(0,1). The choice probability can be written as

$$\Pr(x_i)=\frac{\exp[\nu_i+\ln(x_i)+(\alpha_z/\alpha_x)\ln(T-p_i(x_i))]}{\Sigma_{k=1}^{K}\exp[\nu_k+\ln(x_k^*)+(\alpha_z/\alpha_x)\ln(T-p_k(x_k^*))]},\tag{2.3.8}$$

where x_i is the observed demand. Note that x_i replaces x_k^* for the selected product in the estimation procedure.

The model ideally captures the nature of discrete choices consumers make for packaged goods, by searching through all feasible package bundles for each product. However, one critical limitation is that it cannot explain the miss purchases observed in data. Unless the price difference between the small and the large package sizes is negligible, the first stage quantity choice does not yield one or multiple large package size items as the optimal quantity.

I extend the Basic model in two different ways to incorporate the miss purchases. Extension 1 allows consumers to have preference on package sizes, in order to test whether strong preference on large package sizes drives miss purchases. Extension 2 incorporates the concept of consumer inattention and impose some restrictions on the set of package size bundles to consider in the first stage optimization problem.

2.3.2 Extension 1

I assume that consumers have certain preference on package sizes and that affects utilities coming from consuming a product—that is, the marginal utility of a product depends on the package size a consumer chooses. I also assume that the preference on package sizes only affects the deterministic part of the log marginal utility, and not the stochastic part. This implies that there is no uncertainty in consumers' preference on package sizes. The deterministic part is defined as

$$\nu_{ka} = \beta_{0b}\text{brand}_k + \beta_c\text{char}'_k + \beta_s\text{small}_{ka}, \tag{2.3.9}$$

where small_{ka} is a dummy variable indicating whether the package bundle consists of small package sizes or not. Note that ν_{ka} now includes a subscript a to indicate that it depends on the package bundle a. Let ψ_{ka} denote the marginal utility of package bundle a of product k, where $\ln(\psi_{ka}) = \nu_{ka} + \epsilon_k$.

As the package size affects the marginal utility, it also affects the choice of which package bundle to purchase. The first stage optimal quantity choice problem can be written as

$$\begin{aligned}
&\max_{a\in A}\{\alpha_0+\alpha_x \ln u(x_{ka})+\alpha_z \ln(T-p_k(x_{ka}))\} \\
=&\max_{a\in A}\{\alpha_0+\alpha_x(\nu_{ka}+\epsilon_k+\ln x_{ka})+\alpha_z \ln(T-p_k(x_{ka}))\} \\
=&\max_{a\in A}\{\alpha_x(\beta_s \text{small}_{ka}+\ln x_{ka})+\alpha_z \ln(T-p_k(x_{ka}))\}.
\end{aligned} \tag{2.3.10}$$

The stochastic factor of the log marginal utility, ϵ_k cancels here as in equation (2.3.5), and thus the solution is still deterministic. However, there is an additional term, $\beta_s \text{small}_{ka}$, that hasn't been cancelled out as the package size matters.

Let x^*_{ka} be the quantity of the optimal package bundle a of product k identified in the first stage. The second stage problem is the same as in Basic model, since the comparison is across products. The choice probability can be written as

$$\Pr(x_{ia})=\frac{\exp[\nu_{ia}+\ln(x_{ia})+(\alpha_z/\alpha_x)\ln(T-p_i(x_{ia}))]}{\Sigma_{k=1}^{K}\exp[\nu_{ka}+\ln(x^*_{ka})+(\alpha_z/\alpha_x)\ln(T-p_k(x^*_{ka}))]}. \tag{2.3.11}$$

2.3.3 Extension 2

The second extension of Basic model focuses on the consideration set that a consumer searches through in the first stage in order to find her optimal quantities for each product. In Basic model, the consideration set is A, all possible combinations of package sizes available. Let A' be a subset of A that excludes all multiple small package size bundles. I assume that the consumer's consideration set is restricted to A', instead of A, when she makes a miss purchase. In case of a miss, the first stage optimal quantity problem becomes

$$\begin{aligned}
&\max_{a\in A'}\{\alpha_0+\alpha_x \ln u(x_{ka})+\alpha_z \ln(T-p_k(x_{ka}))\} \\
=&\max_{a\in A'}\{\alpha_x \ln x_{ka}+\alpha_z \ln(T-p_k(x_{ka}))\}.
\end{aligned} \tag{2.3.12}$$

Let $x_{ka}^{*\prime}$ be the quantity of the optimal package bundle a of product k identified in the first stage when the consideration set is A'. The second stage problem is the same as in Basic model. The choice probability can be written as

$$\Pr(x_{ia}) = \frac{\exp[v_i + \ln(x_{ia}) + (\alpha_z/\alpha_x)\ln(T - p_i(x_{ia}))]}{\Sigma_{k=1}^{K}\exp[v_{ka} + \ln(x_{ka}^{*\prime}) + (\alpha_z/\alpha_x)\ln(T - p_k(x_{ka}^{*\prime}))]} \tag{2.3.13}$$

when the purchase observation x_{ia} is a miss. Note that the consumer restricts her consideration to A' for not only the chosen product i but also all the other products that are available. For the regular (non-miss) purchase occasions, the first stage problem and the choice probability are the same as in equation (2.3.5) and (2.3.8), respectively.

2.4 Estimation

The C-D utility function in equation (2.3.1) under three different model specifications is estimated as hierarchical Bayes models (Gelfand and Smith (1990)). Since not every parameter can be identified in the utility function, I set $\alpha_0 = 0$ and $\alpha_x = 1$. After the normalization, α_z represents a household's relative preference of outside good to inside good. It should be positive in economic theory, so I set $\alpha_z = \exp(\alpha^*)$ and estimate α^* unrestricted. In addition, I set $\beta_{03} = 0$, which is the preference parameter of Brand 3 (Private Label), in equation (2.3.3) and (2.3.9).

Hierarchical Bayes models allow household heterogeneity for all parameters $(\theta, T) = (\alpha, \beta', T)$. According to the Bayes theorem, the posterior is proportional to the product of the likelihood and the prior:

$$\pi(\theta_h, T_h) \propto \Pi_j \Pr(x_{ij}|\theta_h', T_h) \times \pi(\theta_h|\bar{\theta}, V_\theta) \times \pi(T_h|a, b), \tag{2.4.1}$$

where j denotes a purchase occasion of an household h, and i denotes the product selected. Hierarchical Bayes models impose a hierarchical structure to beliefs on parameters, and as a result, involve two stages of priors:

$$\begin{aligned} \text{first-stage:} \quad & \pi(\theta_h|\bar{\theta}, V_\theta) \times \pi(T_h|a, b) \\ \text{second stage:} \quad & \pi(\bar{\theta}, V_\theta|\tau). \end{aligned} \tag{2.4.2}$$

I assume the normal prior model specified as

$$\begin{aligned} \theta_h &\sim N(\bar{\theta}, V_\theta), \quad T_h \sim N(a, b) \\ \bar{\theta} &\sim N(\bar{\bar{\theta}}, A^{-1}) \\ V_\theta &\sim IW(\nu, V), \end{aligned} \tag{2.4.3}$$

where V_θ follows the Inverted Wishart distribution with degrees of freedom ν and a scale parameter V.

The posterior distribution is simulated by generating sequential draws using Markov Chain Monte Carlo (MCMC) methods with Metropolis-Hastings Algorithm (Chernozhukov and Hong (2003)). I executed 50,000 iterations of the Markov chain and convergence was checked. Please see Appendix A.3 for details.

2.5 Results

The later half of the chain is used to estimate the model parameters. The estimation results of different model specifications are reported in Table 2.1. Column (1)-(3) show the estimation results of Basic Model, Extension 1, and Extension 2, respectively.

Coefficient estimates are similar across the model specifications and the sign of the coefficients are also consistent. The estimate of the preference parameter to small package size (Small Size) in Extension 2 is positive and statistically significant. Thus, households prefer small jars to large jars in overall.

Here I interpret the estimation results of Extension 2. The aggregate estimate of α^* is equal to 0.8078, and taking an exponential gives us the aggregate estimate of α_z equal to

Table 2.1: Aggregate Coefficient Estimates

Model	(1) Basic		(2) Extension 1		(3) Extension 2	
$\alpha^* = \ln(\alpha_z/\alpha_x)$	0.8268	(0.0642)	0.8570	(0.0716)	0.8078	(0.0648)
Skippy	1.8147	(0.1360)	1.7242	(0.1347)	1.8081	(0.1312)
Jif	1.8494	(0.1395)	1.7519	(0.1351)	1.8526	(0.1351)
Peter Pan	0.5660	(0.1599)	0.5118	(0.1418)	0.5645	(0.1605)
Creamy	1.6891	(0.1091)	1.6903	(0.1000)	1.6661	(0.1035)
Flavor	-2.7281	(0.2528)	-2.5760	(0.1888)	-2.5934	(0.2193)
Salt	-0.4499	(0.1728)	-0.4059	(0.1532)	-0.5016	(0.1632)
Sugar	3.7987	(0.3070)	3.6431	(0.2425)	3.8854	(0.3469)
Natural	-1.7324	(0.1797)	-1.6881	(0.1576)	-1.6868	(0.1739)
Reduced Fat	-2.0798	(0.1241)	-2.0199	(0.1270)	-2.0375	(0.1286)
Small Size	-	-	0.2103	(0.0796)	-	-
T	10.952	(0.2049)	10.945	(0.2064)	10.934	(0.2075)
Log Likelihood	-2688.70		-2671.20		-2683.50	

Note: Estimation was conducted with a subsample of 200 households. Standard deviations are reported in parentheses. Extension 1 allows preference on package size and Extension 2 accommodates consumer inattention. The log likelihood is evaluated at the household level parameter estimates $(\hat{\theta}_h, \hat{T}_h)$.

2.2430. As α_x is normalized to 1, this means households enjoy the outside good 2.2430 times more than peanut butter products. All three brand preference parameters are positive, which means households prefer all three national brands to Private Label. Among the three, Jif is most preferred.

Among the product characteristics preference parameters, Creamy and Sugar have positive coefficients and Flavor, Salt, Natural, and Reduced Fat have negative coefficients. Households prefer creamy texture over chunky texture, and regular sugar level over no sugar added. On the other hand, they dislike peanut butter products that are flavored, salted, and naturally processed. Also, they don't like the products with reduced fat contents either.

Estimated posterior of each parameter in Extension 2 are displayed in Figure 2.1. Each panel in the figure represents a histogram of the household means across 25,000 draws. Households show heterogeneous preference on Skippy and Jif brands, but they all prefer those two national brands to Private Label. Most of households also prefer Peter Pan to Private Label, but few of them like Private Label more. Households also have heterogeneous taste for texture, and most of them prefer creamy texture over chunky texture, except for a few.

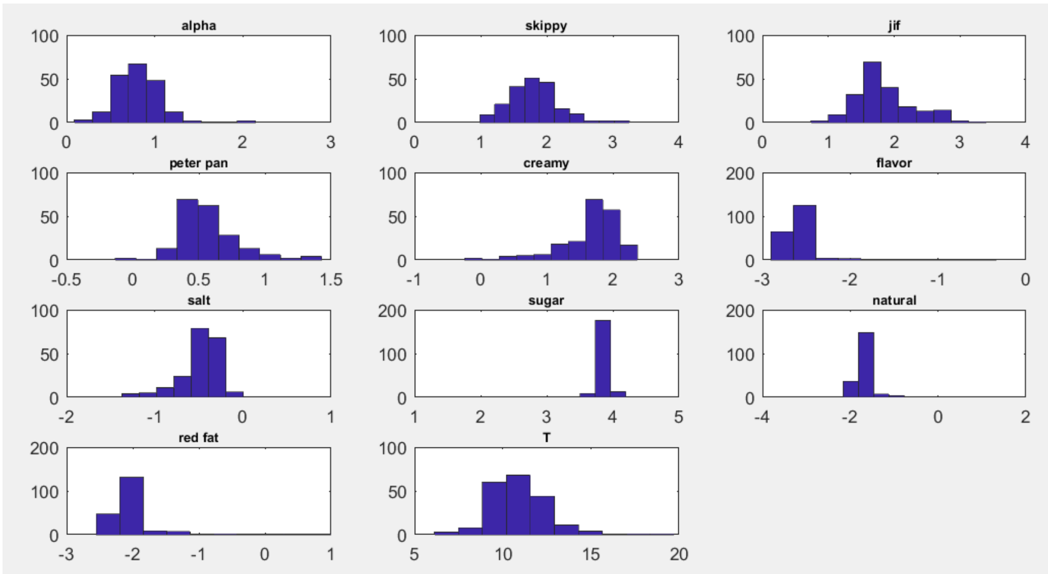


Figure 2.1: Posterior Estimates of Extension 2

Note: Estimation was conducted with a subsample of 200 households. Each panel shows the histogram of the household means for a parameter.

2.5.1 Robustness

I check the robustness of the estimation results by using alternative assumptions on the prior. In this section, I present the evidence of robustness of Basic Model estimation results. The normal prior assumption specified in Equation (2.4.3) is used for the original estimation and the results are presented in Column (1) of Table 2.1. The same results are presented again in Column (1) of Table 2.2 for comparison.

I consider two alternative priors to check robustness, focusing on the distribution of θ. The first alternative is that θ follows the normal distribution but the mean parameter is the half of what is originally assumed. Another alternative is the normal distribution with the same mean but the variance parameter is as twice as big. The estimation results under the alternative priors are presented in Column (2) and (3) of Table 2.2.

The estimation results are consistent across Column (1)-(3). The magnitude of ag-

Table 2.2: Aggregate Coefficient Estimates of Basic Model with Alternative Priors

Prior	(1) $N(\bar{\theta}, V_{\theta})$		(2) $N(\bar{\theta}/2, V_{\theta})$		(3) $N(\bar{\theta}, 2V_{\theta})$	
$\alpha^* = \ln(\alpha_z/\alpha_x)$	0.8268	(0.0642)	0.8246	(0.0676)	0.8294	(0.0623)
Skippy	1.8147	(0.1360)	1.8094	(0.1327)	1.7884	(0.1393)
Jif	1.8494	(0.1395)	1.8553	(0.1338)	1.8254	(0.1454)
Peter Pan	0.5660	(0.1599)	0.5523	(0.1508)	0.5146	(0.1565)
Creamy	1.6891	(0.1091)	1.6848	(0.1014)	1.6972	(0.1050)
Flavor	-2.7281	(0.2528)	-2.6059	(0.2532)	-2.7328	(0.2395)
Salt	-0.4499	(0.1728)	-0.4907	(0.1543)	-0.4499	(0.1651)
Sugar	3.7987	(0.3070)	3.8347	(0.3146)	3.8914	(0.2692)
Natural	-1.7324	(0.1797)	-1.6589	(0.1614)	-1.7077	(0.1335)
Reduced Fat	-2.0798	(0.1241)	-2.0824	(0.1200)	-2.0608	(0.1155)
T	10.952	(0.2049)	10.9541	(0.2038)	10.9577	(0.1973)
Log Likelihood	-2688.70		-2686.80		-2686.80	

Note: Estimation was conducted with a subsample of 200 households. Standard deviations are reported in parentheses. Column (1) shows the estimation results with the original prior as in Column (1) in Table 2.1. Column (2) assumes the mean of θ to be the half of the original, and column (3) assumes the variance of θ to be a double of the original. The log likelihood is evaluated at the household level parameter estimates $(\hat{\theta}_h, \hat{T}_h)$.

gregate coefficient estimates and standard deviations are very similar, although different prior distributions on θ are assumed. Therefore, the estimation results of Basic Model are robust.

2.5.2 Expected Demand

I calculate the expected demand each model predicts, and compare it to the observed demand from the data. The comparison shows that Extension 2 fits the data best, as shown in Table 2.3. The observed demand is calculated as a sum of the total number of jars each household purchased multiplied by the size of the jars in ounces. The expected demand is calculated for each purchase observation, using household level parameter estimates $(\hat{\theta}_h, \hat{T}_h)$. First I solve the first stage problem and find the optimal package bundle x^*_{ka} for each product. In the second stage, I calculate the probability for each product to be chosen, and then multiply it by the quantity of the optimal package bundle.

Extension 2 has the smallest root mean square error (RMSE), which means that it

Table 2.3: Expected Demand and Model Fit

Item ID	Oz	Observed Demand	Expected Demand: Basic	Extension 1	Extension 2
1	16.3	7,665	9,489	9,889	8,911
2	28.0	2,632	2,676	1,697	2,916
3	16.3	2,191	2,034	2,160	1,973
4	28.0	420	483	249	533
5	16.3	521	672	760	730
6	16.3	326	122	137	137
7	16.3	1,958	1,778	1,829	1,783
8	16.3	281	286	293	291
9	15.0	1,558	1,883	1,907	1,892
10	15.0	648	354	363	363
11	18.0	9,756	10,793	12,194	10,077
12	28.0	5,376	5,375	3,194	5,973
13	18.0	1,458	2,816	3,099	2,746
14	28.0	364	901	456	1,005
15	18.0	1,800	2,048	2,076	2,041
16	17.3	502	411	437	387
17	18.0	3,258	4,670	4,707	4,755
18	18.0	414	1,469	1,524	1,391
19	28.0	1,400	1,402	1,304	1,486
20	18.0	666	321	334	321
21	16.3	3,204	3,183	3,202	3,132
22	16.3	793	574	577	581
Root Mean Square Error (RMSE)			31.74	47.82	28.06

Note: Expected demand is calculated using the estimation results obtained from each model specification. Extension 1 allows preference on package size and Extension 2 accommodates consumer inattention.

predicts the demand which is closest to the actual demand observed, compared to the other model specifications. Extension 1 performs poorly; it has RMSE of 47.82%, which is even larger than the Basic model's. These results suggest that consumer inattention better explains households' purchasing behavior than preference on package sizes.

Table 2.3 also shows potential problems in prediction when wrong-specified models are used for the analysis. Structural demand models are widely used to study policy implications and evaluate potential mergers. Thus, it is critical whether the model can predict the demand accurately. In order to achieve that, consumer inattention should be considered in the model properly when packaged goods are of concern. Not considering it (Basic model) or wrong interpretation of consumer behavior (Extension 1) would result poor predictions.

2.6 Implications

I analyze the welfare effects of consumer inattention based on the estimation results from the previous section. I have shown that some households miss substitution opportunities in quantity surcharge periods due to inattention in Section 2.5. Among the 200 households in the subsample, 43 households make at least one miss, with the total number of misses equal to 81.

Choosing an inferior package size reduces the households' utilities. I calculate the loss in welfare for each miss purchase. First I set up a hypothetical scenario where all households are fully attentive. This scenario can be considered as an approximation of a consumer awareness program such as a clear per unit price display requirement. In this scenario, no household makes a miss purchase—that is, the choice set includes all the possible combinations of the package sizes that are available for every purchase occasions, as in Basic model. I set Extension 2 (where some households are inattentive) as a benchmark for comparison.

Then I simulate purchase outcomes under the full attention scenario and Extension 2, using the household level parameter estimates. For miss purchase occasions, utilities are higher under the full attention scenario than the benchmark. I calculate the difference for those cases and take the average across 1,000 simulations.

Lastly, I calculate the additional budgetary allotment needed to compensate the utility difference in order to convert the difference in utilities into dollars. The calculation is based on this approximation:

$$\frac{\partial \ln u(x_k, z)}{\partial \ln T} \times \Delta \ln T = \Delta \ln u(x_k, z), \qquad (2.6.1)$$

which assumes that the impact of a change in T on the optimal quantity x_k is negligible.

Table 2.4 shows the simulated welfare loss.

Table 2.4: Welfare Loss per Miss($)

Min	Q1	Q2	Mean	Q3	Max	Total
0.0088	0.2847	0.6332	0.6858	1.5134	3.1494	50.5926

Note: Welfare loss of consumer inattention is calculated for each miss purchase that the subset of 200 households made. There are 81 misses in total.

The total welfare loss from 81 miss purchases are $50.59. The smallest loss is less than 1 cent, but the biggest loss reaches to $3.15. The median loss is $0.63 and the average is $0.69. Considering the fact that a 16 oz jar of peanut butter is roughly $2, the welfare loss caused by miss purchases is substantial for inattentive households.

2.7 Concluding Comments

I study demand models for packaged goods when quantity surcharges exist in this chapter. Quantity surcharges are frequently observed at grocery stores. Chapter 1 shows heterogeneity in purchasing behavior when quantity surcharges exist, especially the existence of consumers who make miss purchases. Heterogeneity in consumer inattention explains the various purchasing behavior well. Hence it is important to consider consumer inattention when one designs demand models for packaged goods.

This chapter has two main contributions. The first contribution is to develop a demand model that can capture the heterogeneity in consumer attention. The model allows consumers to choose a product and its optimal quantity for packaged goods. The quantity is chosen as a combination of possible package sizes available, and this captures attentive consumers purchasing multiple small size items in quantity surcharge periods. In addition to that, the model accommodates consumer inattention by restricting the choice set in case of miss purchases. The model can be solved as a two stage optimization problem.

I estimate the model using the MCMC methods. The estimation results suggest that consumer inattention explains the household purchasing behavior well. The objective function to maximize is complicated as the model solution requires two stage optimizations, and the number of package bundles to consider is large. The Bayesian approach has advantages in this setting.

I compare the estimation results of the consumer inattention model to the two alternative model specifications: first, a basic model that does not consider quantity surcharges, and then an extension of the basic model that explains miss purchases as a result of strong preference on a large package size. The consumer inattention model shows the minimal prediction errors among the three.

The second contribution is to calculate the costs of consumer inattention. I calculate the welfare loss caused from missing a substitution opportunity, by simulating purchase outcomes using the estimation results from the inattention model. The welfare loss a household suffers is $0.69 on average, which is substantial considering that the average price of a small size jar is roughly $2.

Chapter 3

A Dynamic Demand Model for Packaged Goods

3.1 Introduction

I study the dynamic consumption and purchase decisions consumers make for storable packaged goods in this chapter. I use the same scanner data in the peanut butter category for the analyses. Chapter 1 documents the high frequency of quantity surcharges in the peanut butter category using the detailed scanner data. Chapter 2 develops a demand model for packaged goods that captures the heterogeneous consumer behavior when quantity surcharges exist. However, the model is static and hence fails to consider the consumer's dynamic decision to store for later consumption.

There are three household purchasing patterns to consider. First, households store peanut butter for future consumption. Peanut butter products are storable, as they have long storage life, and I find evidence of storing behavior from my data. Second, multiple jar purchases are highly frequent, especially small size jars in quantity surcharge weeks. Lastly, some households purchase large size jars in quantity surcharge weeks ("miss" purchase). Chapter 1 demonstrates the second and third purchasing behavior in detail.

The last two purchasing patterns combined together suggest that there is heterogeneity in attention among households. Attentive households take advantage of quantity surcharges and purchase two jars of small size items. On the other hand, inattentive households are not aware of the existence of quantity surcharges and make inferior package size choices. Chapter 2 shows that consumer inattention well explains the heterogeneous jar size choices.

I develop a structural demand model for packaged goods that considers the all three purchasing patterns found. First, I allow four types of households: attentive storer, inattentive storer, attentive non-storer, and inattentive non-storer. Different types of households have different preferences on the product. In each period a household chooses consumption quantity, purchase quantity, and the number of jars of each size to purchase. By separating purchase quantity decision and jar size choice, inattentive type households are allowed to make inferior jar size choices for the given purchase quantity.

I model a multiple-discrete choice by considering consumption quantity as a countable variable. More specifically, I assume that consumption quantity follows a negative binomial distribution. This distributional assumption also drives the likelihood function for estimation. I simplify a household's dynamic purchase decision with inventory holding using the concept of effective price by Hendel and Nevo (2013). Under certain assumptions, the storer type household's dynamic problem becomes static where consumption quantity for each period is determined by the effective price.

I estimate the model using the method of maximum likelihood. The rich household level purchase data allow me to estimate the demand parameters. Even though the model is limited to one product, the estimation results still give us helpful insight. The estimation results suggest that attentive storer type households are most price sensitive and inattentive non-storer type households are least price sensitive.

For the literature review on quantity surcharges, please see Section 1.1, and for the literature review on multiple discrete choice model, please see Section 2.1. In addition, the chapter is related to literature on stockpiling behavior. Boizot et al. (2001) and Pesendorfer (2002) documents the evidence that consumer inventory holding behavior using data. Hendel and Nevo (2003) suggest methods to test whether households engage in stockpiling at both household and store levels[1] and Hendel and Nevo (2006b) test consumer inventory holding model using scanner data.

Erdem et al. (2003) introduces a dynamic demand model for storable goods. Hendel and Nevo (2006a) also develops a similar dynamic model with inventory holding. The model separates a household's decisions on how much to consume and how much to purchase. The authors show that static demand analysis mismeasures demand elasticities: it overestimates own-price elasticities and underestimates cross-price elasticities to other products. However, the models in the papers mentioned above do not consider the substitution among different package sizes of the same product. That is, the package size a household chooses is automatically equal to the purchase quantity she chooses. Hence it is not eligible to analyze the effect of quantity surcharges where substitution opportunities exist.

Hendel and Nevo (2013) show how to simplify a complicated dynamic demand model with inventory holdings using certain assumptions. The model first assumes two types of households, storers and non-storers, and allows the two types to have different preferences. Adding further assumptions on storing technology and households' foresight on future demand and future prices, prices from only a short period of adjacent time window become effective for the current period decisions, instead of the full history of past and future prices. However, again, the authors fail to consider package size choice. Hence I adopt the assumptions from Hendel and Nevo (2013) to simplify the inventory holding part, but include additional components to analyze various jar size choices.

The chapter proceeds as follows. Section 2 describes the data sets used for the analyses. Section 3 shows the evidence of household stockpiling behavior and Section 4 introduces a structural dynamic demand model. Section 5 describes the estimation procedure, and Section 6 shows the estimation results. Section 7 concludes the paper.

[1] I follow the methods the authors suggest to find evidence of consumer inventory holding in Section 3.3.

3.2 Data

I use weekly panel scanner data collected by Information Resources Inc. (IRI) for the analyses. I focus on the peanut butter category in the Eau Claire, Wisconsin, market, from January 2008 to December 2010. I merged the three household level data sets - peanut butter purchase, trips to grocery stores, and demographics - and product attributes data. The observations that are not matched to the store level sales data are dropped. For further details, please see Section 1.2 and Appendix A.1.

Products and quantity surcharges are defined in the same ways as in Chapter 1. Table 1.4 shows 17 products of four brands identified. Quantity surcharges are identified for the five products with multiple package sizes (small and large) available. Quantity surcharges are more frequent than quantity discounts for all the products, and the average price difference between the two sizes is bigger in quantity surcharge weeks than in quantity discount weeks for the majority of the products. Please see Table 1.5 for details.

The merged data set includes 2,367 households with 23,287 purchase observations. Focusing on the 17 products identified, there are 2,288 households with 19,142 purchase observations. At store level, there are 29,340 sales observations across 6 stores in the market.

3.3 Evidence of Stockpiling

This section shows the evidence that households stockpile for future consumption. Stockpiling can evolve as an optimal strategy for households, as peanut butter is storable and stores often offer temporary price reductions on peanut butters. When an item is relatively cheaper than usual, households can purchase a large quantity and store some of it as inventory for future consumption. They also tend to delay their next purchase to consume the inventory. I test for household stockpiling behavior in both store and household levels.

3.3.1 Store Level Evidence

One way to test the household stockpiling behavior is to find evidence of post-promotion dip. If households store a product for later use, the quantity of that product sold during the week following a price promotion is expected to be lower than the quantity sold during the week not following a price promotion at store level. This phenomenon is called "post-promotion dip", as suggested by Hendel and Nevo (2003).

Table 3.1 is the contingency table of the average number of jars sold in sale/non-sale weeks followed by sale/non-sale weeks.[2] A sale is defined as in Section 1.3.3, and the analysis is based on 29,340 sales observations at stores of 22 items of 17 products identified in Table 1.4.

The most important comparison is the two numbers in non-sale weeks. During non-sale weeks following non-sale weeks, more peanut butter jars were sold on average than during non-sale weeks following sale weeks (144.99 versus 34.86). This trend suggests a post-promotion dip, which indicates inventory holding. We can also see that more peanut butter jars were sold on average when there was no sale the previous week than when there was a sale (141.32 versus 123.20). Similarly, more peanut butter jars were sold on average during sale weeks than during non-sale weeks (161.33 versus 106.38).

Table 3.1: Evidence of Stockpiling: Average Number of Jars Sold

	$\text{Sale}_{t-1} = 0$	$\text{Sale}_{t-1} = 1$	
$\text{Sale}_t = 0$	144.99	34.86	106.38
$\text{Sale}_t = 1$	134.63	178.22	161.33
	141.32	123.20	

Note: Sale_t in an indicator of sale, which is equal to 1 if an item was on sale during the week t at a store. The table presents the average number peanut butter jars sold across 156 weeks and six stores during each week.

Next, I run several linear regressions of the quantity sold at stores. I follow Hendel and Nevo (2003) and estimate the following econometric model

[2] I call “sale weeks” the weeks in which the peanut butter products were on sale and “non-sale weeks” the weeks in which the peanut butter products were not on sale.

$$\log(q_{jst}) = \delta_1 \log(p_{jst}) + \delta_2 dur_{jst} + dumvars + \epsilon_{jst}, \qquad (3.3.1)$$

where q_{jst} is the quantity of item j sold at store s in week t, measured as volume normalized by 16 oz; p_{jst} is price per 16 oz of item j at store s in week t; *dumvars* includes feature and display dummy variables, and store, brand, and product specific intercept terms. Lastly, dur_{jst} is the duration from the previous sale, measured as the number of weeks divided by 100 from the previous sale for an item j in store s for any size in week t. I consider the three categories of the duration variable: duration from the previous sale for the brand to which the item belongs, for the product to which the item belongs, and for the item itself.

The key parameter is δ_2, measuring the effect of the duration since the previous sale on quantity purchased. The consumer inventory holding hypothesis predicts that consumers holding inventory can delay purchases waiting for the next price discount, but as the number of weeks since the last sale increases, the inventory level decreases and the consumers become more likely to purchase the product again. A positive δ_2 supports the hypothesis.

Table 3.2 shows the ordinary least squares (OLS) estimation results of the model. The first three columns contain the results of the regression models with each of the duration variables, and the last column includes the all three duration variables. All the coefficients are significant and have the expected signs: less quantity is sold if the price increases, but feature or display increases the quantity sold.

Duration from the last sale shows significantly negative effects, when the three categories are considered separately. However, when all the three are considered together, as shown in Column (4), duration from the last sale for the product has a significantly positive coefficient estimate. This supports the consumer inventory holding hypothesis.

Table 3.2: Evidence of Stockpiling: Store Level

Variable	(1)	(2)	(3)	(4)
Log Price	-2.327	-2.365	-2.328	-2.296
	(0.043)	(0.043)	(0.043)	(0.043)
Duration Since Last Sale				
Brand	-0.547			-0.506
	(0.074)			(0.081)
Product		-0.095		0.173
		(0.028)		(0.041)
Item			-0.151	-0.199
			(0.020)	(0.029)
Feature	0.222	0.212	0.222	0.230
	(0.027)	(0.027)	(0.027)	(0.027)
Display	1.118	1.121	1.114	1.110
	(0.021)	(0.021)	(0.021)	(0.021)

Note: Standard errors in parentheses. The table reports the estimation results of (3.3.1). Each observation used in the analyses is a weekly sales event of 22 items of 17 products a store. A sale for an item is defined as any price lower than or equal to the price level 5% below its modal price. A positive coefficient of duration from the previous sale (δ_2) indicates that consumers hold inventory.

3.3.2 Household Level Evidence

I also estimate an econometric model designed to find evidence of consumer inventory holding at household level. The household level data allow us to measure the impact of a sale on the number of jars purchased or the timing of purchase, in addition to the total quantity purchased. The model is

$$y_{it} = \alpha + \beta S_{it} + \gamma_i + \epsilon_{it}, \tag{3.3.2}$$

where y_{it} represents five different measures of quantity purchased by household i at purchase instance t; S_{it} is an indicator whether the item was on sale; and γ_i is a household-specific effect. The model is adopted from Hendel and Nevo (2003).

Among the five different dependent variables, the first three are quantity variables: quantity normalized to 16 oz, number of jars, and size of each jar purchased. A positive coefficient on the sale dummy is expected for normalized quantity and number of jars to support household stockpiling behavior. As for size of jar, I expect it to have a negative coefficient, as households in the panel data frequently purchased multiple jars of relatively small size (16–18 oz) rather than buying one large size jar, as shown in Table 1.11.

The last two dependent variables are related to the timing of purchases: days from previous purchase and days to next purchase. A negative coefficient for the days from previous purchase and a positive coefficient for the days to next purchase will support the stockpiling hypothesis.

The within estimator allows individual household intercepts, and estimates the effects of household stockpiling. Table 3.3 displays the obtained estimates of β for the five different dependent variables. All the models have significant estimates and the signs are as expected. Column (1) indicates that households purchase a larger volume of an item when it is on sale. Columns (2) and (3) indicate that households purchase a higher number of jars and smaller size jars during sale weeks. Columns (4) and (5) suggest that households advance a purchase and wait longer for the next purchase when there is a sale.

Table 3.3: Evidence of Stockpiling: Household Level

Dependent Variable	(1) Quantity (16 oz)	(2) Num of Jars	(3) Jar Size (16 oz)	(4) Days from Prev Pur	(5) Days to Next Pur
S_{it} (β)	0.207	0.228	-0.0352	-7.221	7.094
	(0.015)	(0.012)	(0.003)	(1.734)	(1.729)

Note: Standard errors in parentheses. The five columns in this table report the estimation results of the model (3.3.2) with five different dependent variables. Each observation used in this analysis is a purchase observation that a household made on a single trip to a store. A sale for an item is defined as any price lower than or equal to the price level 5% below its modal price; and S_{it} is an indicator variable of sale (obtained estimates of the variables other than S_{it} are omitted). Quantity is the amount of an item purchased, normalized to 16 oz, and num of jars is the number of jars of an item purchased. Jar size is the volume of each jar purchased normalized to 16 oz. Days from prev is the number of days from the last peanut butter product purchase, and days to next is the number of days to the next purchase.

I found the evidence of household stockpiling behavior in this section, in both store and household levels. Thus, it is important to consider the dynamic decisions between purchase and consumption quantities consumers make in demand models.

3.4 Model

I develop a demand model that is simple but still captures the key consumer behavior that I find from the data, such as inventory holding and multiple jar purchases.

I propose a demand model when a single good is offered at two package sizes and prices between the two package sizes are not necessarily linear. In each period, households choose not only the quantity to purchase but also the bundle of different package sizes conditional on the given purchase quantity.

3.4.1 The Setup

First, I assume that there is heterogeneity in households' willingness or ability to store.

Assumption 1. *A proportion of households do not store.*

This could arise endogenously if a fraction of the households have storage costs that make it unprofitable to store. Therefore, Assumption 1 can be interpreted as an assumption on the distribution of storage costs.

Assumption 2. *A proportion of households are inattentive.*

Household inattention could arise for various reasons: some households may be unable to calculate the per unit price for an item and compare it to the other package size items. Other households could have high search costs or high opportunity costs to the time spent at a grocery store. Clerides and Courty (2017) call this case rational inattention. Combining Assumption 1 and 2 yields four types of households: attentive non-storer (A-NS), inattentive non-storer (IA-NS), attentive storer (A-S), and inattentive storer (IA-S).

I suppose there are a finite number of periods, R. There is one product and it is offered in two different package sizes, small (S) and large (L). A large size item is twice as big as a small size item. I normalize quantity to small size so that quantity is equal to 1 for the small size item and 2 for the large size item. I define $p_t = (p_{tS}, p_{tL})$ as a vector of the normalized prices (per quantity) for the small and large size items in period t. I allow the four types of households to have different preferences to the product, but none of them has preference to package size. A type h household has indirect utility as follows:

$$U_t^h = \beta_0^h + \beta_p^h \tilde{p}_t + \epsilon_t^h, \tag{3.4.1}$$

where $\tilde{p}_t$ is the representative price of the product in period t, which I will discuss later; ϵ_t^h represents a stochastic preference shock in period t; $\beta^h = (\beta_0^h, \beta_p^h)$ is a vector of preference parameters of type h households; β_0^h represents type h households' baseline preference to the product; and β_p^h captures type h households' sensitivity to the price of the product.

In each period, households decide how much to consume, c_t, and how much to purchase, q_t, in terms of quantity. They also choose package sizes according to the purchase quantity. That is, both consumption quantity and purchase quantity are discrete variables: $c_t, q_t \in \{0, 1, 2, \ldots\}$. If households do not store, purchase quantity and consumption quantity are the same, but if households store, those two quantities are not necessarily the same.

I structure the problem a household faces in a two-stage setting. In stage 1, a household makes quantity decisions: how much to consume and how much to purchase. In stage 2, for the given quantity to purchase, the household decides the number of small size items, y_{tS}, and the number of large size items, y_{tL}, to purchase such that $y_{tS} + 2y_{tL} = q_t$. Let $y_t = (y_{tS}, y_{tL})$.

3.4.2 Negative Binomial Distribution

Note that all the three decision variables households choose are discrete, more precisely, nonnegative integers. I introduce the negative binomial distribution to accommodate the discreteness.[3] Poisson distribution could be an alternative, but negative binomial distribution allows more flexibility, as it has an additional parameter.

First, suppose consumption quantity c follows the Poisson distribution conditional on the parameter λ with density function

$$f(c|\lambda) = \frac{\lambda^c}{c!}e^{-\lambda}, \quad \text{where} \quad c = 0, 1, 2, \ldots \tag{3.4.2}$$

Now allow the parameter λ to be random. Let $\lambda = \mu\nu$, where μ is a deterministic function and $\nu > 0$ is iid with gamma density $g(\nu)$,

$$g(\nu) = \frac{\delta^\delta}{\Gamma(\delta)}\nu^{\delta-1}e^{-\nu\delta}. \tag{3.4.3}$$

Then consumption quantity c follows the negative binomial distribution with a mixture density, as follows:

$$\begin{aligned} h(c|\mu,\alpha) &= \int_0^\infty \frac{e^{-\mu\nu(\mu\nu)^c}}{c!}\frac{\nu^{\delta-1}e^{-\nu\delta}\delta^\delta}{\Gamma\delta}d\nu \\ &= \frac{\mu^c\delta^\delta\Gamma(c+\delta)}{\Gamma(\delta)c!(\mu+\delta)^{c+\delta}} \\ &= \frac{\Gamma(\alpha^{-1}+c)}{\Gamma(\alpha^{-1})\Gamma(c+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^c, \end{aligned} \tag{3.4.4}$$

where $\alpha = 1/\delta$. Using the standard notation of the negative binomial distribution, δ represents the number of failures until the experiment stopped and $\mu/(\mu+\delta)$ is the success probability in each experiment. Lastly, let $\mu = \exp(\beta_0 + \beta_p\tilde{p})$, exponential of the mean utility, and $\nu = \exp(\epsilon)$, exponential of the stochastic preference shock. Note that the mean and the variance of ν are 1 and α, respectively. Then consumption quantity c is determined by the binomial distribution with density $h(c|\mu,\alpha)$.

[3] Saeedi (2014) also uses the negative binomial distribution to model the sellers' discrete choice of quantity to sell.

The two moments of consumption quantity are

$$\begin{aligned} E[c|\mu,\alpha] &= \mu \\ V[c|\mu,\alpha] &= \mu(1+\mu\alpha) \end{aligned} \tag{3.4.5}$$

As we can see, α determines the variance of ν, which is the exponential of the stochastic preference shock, and hence, it also affects the variance of consumption quantity. Thus α can be interpreted as a dispersion parameter.

In order to test whether the consumption quantity can be well represented by the negative binomial distribution, I plot the quantity purchased on a given trip for each household. Figure 3.1 shows the quantity purchased, instead of the quantity consumed, since the quantity consumed is unobserved. We can see that the negative binomial distribution is a good approximation of the quantity purchased. Conditioning that the quantity consumed is not significantly different from the quantity purchased, the negative binomial distribution is also a good approximation of the quantity consumed.

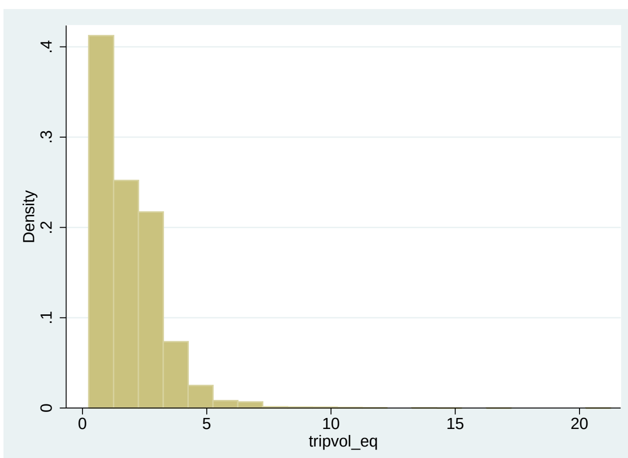


Figure 3.1: Histogram of Quantity Purchased at Household-Trip Level

Note: The figure shows the histogram of quantity (normalized to 16 oz) each household purchased on a trip, conditional on purchase. The data used are the household level weekly purchase data from 2008 to 2010 in the Eau Claire market.

3.4.3 The Non-storing consumer's problem

In this section, I look at the non-storing household's static problem, while in the next section, I look at the storing household's dynamic problem. First, however, there is one problem to discuss: what price to use for the stage 1 quantity decision? In the previous section, I showed that the consumption quantity is drawn from the negative binomial distribution conditional on two parameters, the exponential of mean utility μ^h and the dispersion parameter α^h. The mean utility of a type h household is determined by the household's preference to the product and its price. Here we face the "one product-two prices" problem. That is, there is one product but two prices exist for the product, one for the small size and another for the large size.

In order to determine the mean utility a type h household enjoys from consuming the product, we need to pick a representative price $\tilde{p}_t$ for the product, as shown in equation (3.4.1). There are several candidates for this single price of the product. We can use either the small size price or the large size price. The lower price between the two sizes or the most frequently charged price across the sizes (regular price) can also be an option. Intuitively, attentive households and inattentive households could use different prices to determine their consumption quantities. However, for simplicity, I assume that both types of households use the minimum price between the two sizes in each period. Let $\underline{p_t} = \min(p_{tS}, p_{tL})$.

In the absence of storing, the purchasing quantity and consumption quantity are the same. How much to consume in each period is determined by the binomial distribution with density $h(\cdot)$ in equation (3.4.4). Let $\mu^h(\underline{p_t}) = \exp(\beta_0^h + \beta_p^h \underline{p_t})$, exponential of the mean utility of a type h household when the price she faces for the quantity decision is $\underline{p_t}$. Also, for simplicity, I limit the consumption quantity to 0, 1, or 2 in this analysis. Then the probability of a non-storer household to choose consumption quantity c is

$$\text{Prob}(c_t^h = c) = h(c|\mu^h(\underline{p_t}), \alpha^h), \quad \text{where } c = 0,1,2, \text{ and } h = A-NS, IA-NS. \tag{3.4.6}$$

For the given quantity decision c_t^h, the two types of non-storing household choose a bundle of two package sizes, small and large, y_t^h. When $c_t^h = 0$ for both types of households, $y_t^h = (0,0)$ automatically. When $c_t^h = 1$, there is only one option available: purchasing one small size jar. Hence, $y_t^h = (1,0)$ for both types of households. However, when $c_t^h = 2$, there are two alternatives available: purchasing one large size jar or two small size jars. Here I take a behavioral approach and assume attentive type households minimize their expenditure while inattentive type households minimize the total number of jars to purchase.

Define quantity discounts in period t as $p_{tS} > p_{tL}$ and quantity surcharges as $p_{tS} < p_{tL}$, assuming no tie. Then when $c_t^h = 2$, the attentive non-storer household purchases one large size jar during quantity discount periods and two small size jars during quantity surcharge periods. However, the inattentive non-storer household purchases one large size jar regardless of the period. The two types of non-storing households' jar size choice for a given purchase quantity is presented in Table 3.4.

3.4.4 The Storing Consumer's Problem

If households store, the quantity purchased and the quantity consumed are not necessarily the same. In order to predict storers' purchases, I make the following assumptions:

Assumption 3. *Storage is free, but inventory lasts for only T periods (fully depreciates afterwards).*

One way to interpret the assumption is the product's perishability. For example, $T = 1$ means the product goes bad in two weeks. An alternative interpretation of the assumption is households' capacity to consider their future consumption at purchase. Even though the product may last longer, households only consider purchasing for T periods ahead. The role of this assumption is to dramatically simplify the state space of the model to consider. That is, I do not need to keep track of how much is left in storage in different states, which is what a typical dynamic problem with inventory holdings would do.

Assumption 4. *(perfect foresight): Households have perfect foresight regarding prices T periods ahead.*

The perfect foresight assumption further simplifies the storing household's problem. One may worry that perfect foresight is restrictive, and thus, invalidate demand estimates. However, Hendel and Nevo (2013) show that this perfect foresight assumption fits the data as well as an alternative assumption with households having a rational expectation of future prices.

I assume $T = 1$ for simplicity in the analysis. The storing household's problem is as follows. The type h storing household decides how much to purchase in each period, $\{q_t^h\}_{t=1}^R$, and how much to consume, $\{c_t^h\}_{t=1}^R$, maximizing the sum of utilities subject to the household's budget constraint and storage technology. Here I ignore discounting for utilities from future consumption. As the non-storing type household's problem, both variables are non-negative integers.

Define the effective price in period t as $p_t^{ef} = \min\{\underline{p_{t-1}}, \underline{p_t}\}$. Hendel and Nevo (2013) show that, under Assumptions 3-4, it is equivalent to solve for a series of static optimal consumption quantity with respect to p_t^{ef}. Let $c_t^h(p_t^{ef})$ be the optimal consumption quantity of the type h household. The variable $c_t^h(p_t^{ef})$ is drawn from the negative binomial distribution with density $h(\cdot|\mu^h(p_t^{ef}), \alpha^h)$, as shown in the previous section.

Now I consider the purchase quantity. Since inventory lasts for one period, we only need to consider the two adjacent periods for the purchase quantity decision in period t. The storing household's purchase quantity in period t, q_t^h, is a sum over the consumption in period t if $\underline{p_t}$ is the effective price in period t, and the consumption in period $t+1$ if $\underline{p_t}$ is the effective price in period $t+1$. By definition, $\underline{p_t}$ is the effective price in period t if it is lower than $\underline{p_{t-1}}$, and the effective price in period $t+1$ if it is lower than $\underline{p_{t+1}}$.

Define a sale period S as $\underline{p_t} < \underline{p_{t+1}}$ and a non-sale period N otherwise. Then there are four possible events in period t: a sale in period t followed by a sale in period $t-1$ (SS), a sale in period t followed by a non-sale in period $t-1$ (NS), a non-sale in period t

followed by a sale in period $t-1$ (SN), and lastly, a non-sale in the two periods in a row (NN). The purchase quantity in period t of the type h storing household in each event can be written as

$$q_t^h\left(\underline{p_{t-1}}, \underline{p_t}, \underline{p_{t+1}}\right) = \begin{cases} c_t^h\left(\underline{p_t}\right) & + \quad 0 & \text{in NN;} \\ 0 & + \quad 0 & \text{in SN;} \\ c_t^h\left(\underline{p_t}\right) & + \quad c_{t+1}^h\left(\underline{p_t}\right) & \text{in NS;} \\ 0 & + \quad c_{t+1}^h\left(\underline{p_t}\right) & \text{in SS} \end{cases}, \text{ where } h = A-S, IA-S. \tag{3.4.7}$$

Hence, the purchase quantity in period t for the two storing type households is determined by the minimum price between the two sizes in period t, $\underline{p_t}$, and the sale states in the previous period and the current period. Again, I restrict the consumption quantity to no more than 2 for simplicity. Then the purchase quantity can have an integer value from 0 to 4, depending on the sale event. As described earlier, both $c_t^h(\underline{p_t})$ and $c_{t+1}^h(\underline{p_t})$ follow the negative binomial distribution with density $h(\cdot|\mu^h(\underline{p_t}), \alpha^h)$. Thus, the probability of a storing type household to choose the purchase quantity q in the sale event NN or SS is

$$q_t^h(\underline{p_t}|NN) = q_t^h(\underline{p_t}|SS) = q \quad \text{with prob. } h(q|\mu^h(\underline{p_t}), \alpha^h),$$
$$\text{where } q = 0, 1, 2 \text{ and } h = A-S, IA-S. \tag{3.4.8}$$

$q_t^h(\underline{p_t}|SN)$ is simply zero with probability 1.

It is little bit more complicated in the event of NS since the storing household purchases the consumption quantity for both period t and period $t+1$. For example, there are two possible scenarios why the household purchases quantity 1 in period t in the event of NS: 1) the household's consumption quantity is 0 in period t and 1 in period $t+1$; 2) the household's consumption quantity is 1 in period t and 0 in period $t+1$. Assuming independence between the consumption quantity in period t and period $t+1$, the probability of having each purchase quantity is as follows:

$$q_t^h(\underline{p_t}|NS) = \begin{cases} 0 & \text{with prob. } h(0|\mu^h(\underline{p_t}),\alpha^h)h(0|\mu^h(\underline{p_t}),\alpha^h) \\ 1 & \text{with prob. } h(0|\mu^h(\underline{p_t}),\alpha^h)h(1|\mu^h(\underline{p_t}),\alpha^h) \\ & \quad +h(1|\mu^h(\underline{p_t}),\alpha^h)h(0|\mu^h(\underline{p_t}),\alpha^h) \\ 2 & \text{with prob. } h(0|\mu^h(\underline{p_t}),\alpha^h)h(2|\mu^h(\underline{p_t}),\alpha^h) \\ & \quad +h(1|\mu^h(\underline{p_t}),\alpha^h)h(1|\mu^h(\underline{p_t}),\alpha^h)+h(2|\mu^h(\underline{p_t}),\alpha^h)h(0|\mu^h(\underline{p_t}),\alpha^h) \\ 3 & \text{with prob. } h(1|\mu^h(\underline{p_t}),\alpha^h)h(2|\mu^h(\underline{p_t}),\alpha^h) \\ & \quad +h(2|\mu^h(\underline{p_t}),\alpha^h)h(1|\mu^h(\underline{p_t}),\alpha^h) \\ 4 & \text{with prob. } h(2|\mu^h(\underline{p_t}),\alpha^h)h(2|\mu^h(\underline{p_t}),\alpha^h) \end{cases} \tag{3.4.9}$$

Conditional on the purchase quantity determined, the storing household decides the number of jars to purchase for each size as the non-storing household does. The attentive storer households minimize their expenditure and the inattentive storer households minimize the total number of jars to purchase. In this section, I discuss the cases where the purchase quantity is larger than 2. When the purchase quantity in period t is 3, the attentive storer chooses one small size jar and one large size jar if quantity discount exists in period t, and three small size jars in the case of quantity surcharge. The inattentive storer, on the other hand, purchases one small size jar and one large size jar all the time, for the given purchase quantity 3.

When the purchase quantity is equal to 4, the attentive storer purchases two large size jars if quantity discount exists, and four small size jars if quantity surcharge exists. Again, the inattentive storer's jar size choice does not depend on quantity discountd or quantity surcharged, and the household purchases two large size jars in both cases. Table 3.4 describes each type of households' jar size choice for a given purchase quantity.

Table 3.4: Jar Size Choices of Four Household Types, for Given Purchase Quantity

HH type	A-NS		IA-NS	A-S		IA-S
q	QD	QS		QD	QS	
0	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
1	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
2	(0,1)	(2,0)	(0,1)	(0,1)	(2,0)	(0,1)
3				(1,1)	(3,0)	(1,1)
4				(0,2)	(4,0)	(0,2)

Note: There are four types of households: attentive non-storer (A-NS), inattentive non-storer (IA-NS), attentive storer (A-S), and inattentive storer (IA-S). Variable q represents the given purchase quantity. For a given quantity, the attentive type households choose the number of jars to purchase for each size, (y_S, y_L), depending on quantity discount or quantity surcharge state. However, non-attentive households' jar size choice does not depend on quantity discount or quantity surcharge.

3.5 Estimation

3.5.1 Empirical Model

In this section, I modify the simple demand model from Section 3.4 for estimation using household level purchase data in the peanut butter category. There is one peanut butter product with two package sizes, small and large. There are S stores carrying the product and N households in the market. Assume there are four types of households: attentive non-storer (A-NS), inattentive non-storer (IA-NS), attentive storer (A-S), and inattentive storer (IA-S). Household types are not observable to researchers, but can be identified based on the households' purchasing patterns. Let N_1 and N_2 be the number of A-NS and IA-NS type households, and N_3 and N_4 the number of A-S and IA-S type households, respectively.

There is a finite number of weeks, R. I consider a household's visit to a store s in week t as given. This assumes that households do not plan to visit a store just to purchase the peanut butter product. Upon a visit, a type h household i observes the volume prices of the small and large size items, p_{tSs} and p_{tLs}, respectively, and decides how much to consume, c_{it}^h, how much to purchase, q_{it}^h, and how many jars of each size to purchase, $y_{it}^h = (y_{itS}^h, y_{itL}^h)$, such that $y_{itS}^h + 2y_{itL}^h = q_{it}^h$. From the data we observe the following: week t and store s each household visited, $y_{it}^h = (y_{itS}^h, y_{itL}^h)$, and $p_{ts} = (p_{tSs}, p_{tLs})$.

The probability to observe a purchase incident of $y_{it}^h = (y_{itS}^h, y_{itL}^h)$ is the probability for the type h household i of having the purchase quantity $q_{it}^h = y_{itS}^h + 2y_{itL}^h$ multiplied by the probability of making such a jar choice (y_{itS}^h, y_{itL}^h) for the given purchase quantity. This can be written as

$$\text{Prob}(y_{it}^h|\alpha^h, \beta^h) = \text{Prob}(y_{it}^h|q_{it}^h) \times \text{Prob}(q_{it}^h|\mu_{it}^h, \alpha^h), \tag{3.5.1}$$

where α^h is the dispersion parameter and $\beta^h = (\beta_0^h, \beta_p^h)$ is a vector of the preference parameters. The first term on the right hand side, $\text{Prob}(y_{it}^h|q_{it}^h)$, represents the probability of making such a jar size choice y_{it}^h for the given purchase quantity q_{it}^h. The second term on the right hand side, $\text{Prob}(q_{it}^h|\mu_{it}^h, \alpha^h)$, represents the probability of purchasing quantity choice q_{it}^h, where $\mu_{it}^h = \exp(\beta_0^h + \beta_p^h \underline{p_{ts}})$ when s is the store household i visited in week t.

The quantity that a non-storer type household $(h = A - NS, IA - NS)$ purchases is equal to the consumption quantity drawn from the negative binomial distribution. Hence the household's purchase quantity follows the same negative binomial distribution as the household's consumption quantity. Therefore, $\text{Prob}(q_{it}^h|\mu_{it}^h, \alpha^h)$ is equal to $h(q_{it}^h|\mu_{it}^h, \alpha^h)$, where $h(\cdot)$ is the density function of the negative binomial distribution, as shown in equation (3.4.4).

For the two storer types $(h = A-S, IA-S)$, the purchase quantity depends on the consumption quantity and the sale event, as shown in equation (3.4.7). When the sale event is NS or SS, storer type households purchase the future consumption quantity (for week $t+1$) in week t. I would like to emphasize that storer type households also draw their consumption quantity in week $t+1$ from the same negative binomial distribution with density $h(\cdot|\mu_{it}^h, \alpha^h)$ as the households draw their consumption quantity in week t, since these households face the same price $\underline{p_{ts}}$ for both. Thus the probability of a purchase quantity choice of a storer type household is

$$\text{Prob}(q_{it}^h|\mu_{it}^h,\alpha^h) = h(q_{it}^h|\mu_{it}^h,\alpha^h)^{1-\mathbf{1}(SN)-\mathbf{1}(NS)} \times \mathbf{1}(q_{it}^h=0)^{\mathbf{1}(SN)} \times \left(\sum_{l=0}^{q_{it}^h} h(l|\mu_{it}^h,\alpha^h)h(q_{it}^h-l|\mu_{it}^h,\alpha^h)\right)^{\mathbf{1}(NS)}, \quad h = A-S, IA-S. \tag{3.5.2}$$

For a given visit to a store s in week t and the purchase quantity q_{it}^h, the type h household i decides how many jars to purchase for each package size, y_{itS} and y_{itL}. Attentive type households ($h = A-NS, A-S$) choose y_{it}^h to minimize their expenditure. Hence the jar size choice depends not only on the purchase quantity but also on the quantity discount or quantity surcharge status. During quantity discount weeks, an attentive household purchases as many large size jars as possible for the given purchase quantity. During quantity surcharge weeks, on the other hand, an attentive household purchases small size jars only. Thus, attentive type households' jar size choice probability for the given purchase quantity, depending on the quantity discount or quantity surcharge status, is

$$\begin{aligned}
\text{Prob}(y_{it}^h|q_{it}^h,\text{QD}_t) &= \begin{cases} 1 & \text{if } y_{it}^h = (q_{it}^h - 2\lfloor\frac{1}{2}q_{it}^h\rfloor, \lfloor\frac{1}{2}q_{it}^h\rfloor) \\ 0 & \text{otherwise} \end{cases} \\
\text{Prob}(y_{it}^h|q_{it}^h,\text{QS}_t) &= \begin{cases} 1 & \text{if } y_{it}^h = (q_{it}^h, 0) \\ 0 & \text{otherwise} \quad \text{if } h = A-NS, A-S. \end{cases}
\end{aligned} \tag{3.5.3}$$

However, inattentive type households ($h = IA-NS, IA-S$) choose y_{it}^h to minimize the total number of jars to purchase. Therefore, inattentive households' jar size choice depends on the purchase quantity only, and they purchase as many large size jars as possible all the time. Inattentive type households' jar choice probability for the given purchase quantity is

$$\text{Prob}(y_{it}^h|q_{it}^h) = \begin{cases} 1 & \text{if } y_{it}^h = (q_{it}^h - 2\lfloor\frac{1}{2}q_{it}^h\rfloor, \lfloor\frac{1}{2}q_{it}^h\rfloor) \\ 0 & \text{otherwise} \quad \text{if } h = IA-NS, IA-S. \end{cases} \tag{3.5.4}$$

Let ϕ^h_{it} be the probability of jar choice of a type h household i in week t for simplification. Then the log likelihood function can be written as

$$\log L(\alpha,\beta) = \sum_h \log L^h(\alpha^h,\beta^h), \quad \text{where} \tag{3.5.5}$$

$$\log L^h(\alpha^h,\beta^h) = \sum_{t=1}^{R}\sum_{i=1}^{N^h}\left\{\log\phi^h_{it} + \log h(q^h_{it}|\mu^h_{it},\alpha^h)\right\} \quad \text{for } h = A-NS, IA-NS,$$

$$\begin{aligned}\log L^h(\alpha^h,\beta^h) = \sum_{t=1}^{R}\sum_{i=1}^{N^h}\Big\{&\log\phi^h_{it} + (1-\mathbf{1}(SN)-\mathbf{1}(NS))\log h(q^h_{it}|\mu^h_{it},\alpha^h) \\ &+\mathbf{1}(SN)\log(\mathbf{1}(q^h_{it}=0)) \\ &+\mathbf{1}(NS)\log\left(\sum_{l=0}^{q^h_{it}} h(l|\mu^h_{it},\alpha^h)h(q^h_{it}-l|\mu^h_{it},\alpha^h)\right)\Big\} \quad \text{for } h = A-S, IA-S.\end{aligned} \tag{3.5.6}$$

The identification of the parameters is discussed in Appendix A.4.

3.5.2 Data Cleaning

I clean the data in the following way to fit them to the empirical model presented in section 3.5.1. The data set I use for the estimation includes household level peanut butter product purchase data combined with trip data from the Eau Claire, Wisconsin, market from 2008 to 2010. I also use the store level peanut butter sales data for price information.

As the model assumes a single product, I combine the products with multiple package sizes available into one. Among 17 products I identified in Chapter 1 (Table 1.4), five of them have multiple package sizes, small and large. Four of them are national brands (Product 1,2,9, and 10), and one is the store brand (Product 14). I exclude the store brand product for the analyses, since it shows somewhat different pricing history compared to the other national brand products [4]. Any purchase of the four national brand products from the data is considered a purchase, and a purchase of any other products or no purchase is considered a no purchase.

To define the package size, I consider the 16.3 oz jar of Product 1 and 2, and the 18

[4] See Table 1.5.

oz jar of Product 9 and 10 a small size item, and I consider the 28 oz jar of any of the four products mentioned a large size item. The large size item is not exactly twice as big as the small size item in the data. In order to handle this issue, I set the quantity to 1 if the jar size is between 16 oz and 18 oz, and to 2 if the jar size is between 28 oz and 36 oz.

To define the price of each size item, I calculate the average price of the four items of each size at week-store level. Let p_{tks} be the average price per 16 oz of the four size k items in week t at a store s. Then I define sale and quantity discounts or quantity surcharges at week-store level using the minimum price between the two sizes, $\underline{p_{ts}}$.

There are cases of multiple trips to the same or different stores in a week, and some households purchased peanut butter items multiple times on those trips. However, the model assumes maximum one visit per week. Hence I drop or merge some of those observations in the following way. First, I identify the trip observations with no purchase. If there are no other trip observations the same week, I keep the observation. If there are other trip observations the same week but there was no purchase, I keep only one representative trip observation.[5] If there is any purchase observation the same week, I drop all the other trip observations with no purchase that week.

Some households made multiple purchases at multiple stores the same week. In that case, I drop all the observations.[6] In the case of multiple purchases from the same stores in a week, I merge those purchase observations of the same size. In addition, I drop a few observations that have no matching price information in the store level data. Lastly, I drop the purchase observations with a total quantity purchased greater than 4. Those observations take only 2.51% of the total purchase observations. As a result, the clean data set includes 1,986 households with 9,810 purchase observations and 228,242 trip observations.

[5] I choose the observation recorded earliest in the week.

[6] An alternative way to handle this case is keeping one purchase observation out of the multiple purchase observations and drop all the rest.

3.5.3 Estimation Procedure

I estimate the parameters of each household type h, (α^h, β^h), using the method of maximum likelihood. The first step is to assign a type to each household. I identify the unobserved household types using the method of *kmeans*. Let h_i be the individual specific moments which are informative about the unobservable individual effects. The *kmeans* method estimates a partition of individual units by finding the best grouped approximation to the moments $\{h_i\}$ based on K groups

$$(\hat{\varphi}, \hat{k_1}, \ldots, \hat{k_N}) = \text{argmin}_(\varphi, k_i, \ldots, k_N) \sum_{i=1}^{N} \| h_i - \varphi(k_i) \|^2, \tag{3.5.7}$$

where $\| \cdot \|$ denotes the Euclidean norm; $\{k_i\} \in \{1, \ldots, K\}^N$ are partitions of $\{1, \ldots, N\}$ into at most K groups; and $\varphi = (\varphi(1)', \ldots, \varphi(K)')'$ are $K \times 1$ vectors.

Since I assume four types of households, I set K to 4. I choose one moment for each dimension of household heterogeneity: storing and attentiveness. The first moment is the percentage of quantity purchased in sale weeks (PSALE) for each household. This moment represents the individual household's storing behavior. The second moment is the percentage of quantity purchased as a miss (PMISS). Here a miss is defined as a large size jar purchase in quantity surcharge weeks. The second moment measures the household's attentiveness. For example, a household with low PSALE and low PMISS is likely to be the attentive non-store type.

Using the two moments, I assign an initial household type according to the following criteria: 1) a household is assigned to the storer type if PSALE is larger than 0.5, and to the non-storer type otherwise; 2) a household is assigned to the attentive type if PMISS is 0, and to the inattentive type otherwise. With this initial type assignment, I perform *kmeans* and obtain the estimation results as presented in Table 3.5. There are significantly more attentive type households than inattentive ones. The attentive storer type has the largest number of households, and the inattentive non-storer type has the smallest number of households.

Table 3.5: Results of Household Type Assignments

Type	Num of HHs	Num of Obs	Ave PMISS (%)	Ave PSALE(%)
A-NS	622	70,651	2.33	25.76
IA-NS	88	9,287	70.54	15.25
A-S	1,075	125,829	1.03	85.13
IN-S	201	22,475	49.56	68.39
Total	1,986	228,242	9.43	61.75

Note: The four household types are attentive non-storer (A-NS), inattentive non-storer (IA-NS), attentive storer (A-S), and inattentive storer (IA-S). I assign a type to each household using the method of kmeans. Two household moments are used to perform kmeans: PSALE and PMISS, where PSALE represents the percentage of quantity purchased on sale weeks and PMISS represents the percentage of quantity purchased as a miss.

In addition, I adjust the probability of observing certain jar size choices. When the purchase quantity is larger than 1, multiple alternatives exist in terms of jar size choice. For the attentive types ($h = A - NS, A - S$), the "right" choice is determined by quantity discounts or quantity surcharges, as shown in equation (3.5.3). For example, when the purchase quantity is 2 and quantity surcharges exist, the right choice is to purchase two small size jars. The model predicts the probability of observing $y_{it}^h = (2,0)$ in that case to be 1, and the probability of observing the "wrong" choice, $y_{it}^h = (0,1)$, to be 0. However, logarithm of 0 goes to negative infinity and it makes the log likelihood function explode to negative infinity. The same problem arises for the inattentive type households in some cases. Hence, for technical reasons, I adjust the probability of making the right choice to 0.9 and the probability of making the wrong choice to 0.1.

In addition, for the storing type households, the model predicts the probability of having the purchase quantity equal to 0 to be 1 in the sale event SN, as shown in equation (3.4.7). Again, the same technical issue arises, so I adjust the probability to have purchase quantity 0 in the SN event to 0.9 instead of 1, and the probability of having a positive purchase quantity to 0.1.

3.6 Results

I first present the estimation results when all the households are assumed to be non-storers (see Table 3.6). Column (1) presents the demand estimates when all the households are assumed to be attentive non-storer types. The obtained estimates of the two preference parameters $\beta^{A-NS} = (\beta_0^{A-NS}, \beta_p^{A-NS})$ are 1.787 and -2.126. The interpretation of the coefficients is as follows. As shown in equation (3.4.5), the mean of consumption quantity c_{it}^h is equal to the exponential of the mean utility μ_{it}^h. Thus, one unit change in $\underline{p_{ts}}$ increases the expectation of c_{it}^h by

$$\frac{\partial E[c_{it}^h | \underline{p_{ts}}]}{\partial \underline{p_{ts}}} = \beta_p^h \exp(\beta_0^h + \beta_p^h \underline{p_{ts}}). \tag{3.6.1}$$

We can calculate the average response by taking the average across individual households' response in each week as follows:

$$\frac{1}{Rn^h} \sum_{t=1}^{R} \sum_{i=i}^{N^h} \frac{\partial E[c_{it}^h | \underline{p_{ts}}]}{\partial \underline{p_{ts}}} = \frac{1}{RN^h} \sum_{t=1}^{R} \sum_{i=i}^{N^h} \beta_p^h \exp(\beta_0^h + \beta_p^h \underline{p_{ts}}). \tag{3.6.2}$$

The formula yields the average response -0.1733. That is, when all the households are assumed to be the attentive non-storer types, the average consumption quantity decreases by 0.1733 units when the minimum price between the two sizes of the product increases by one unit.

Using the equation (3.6.1), we can calculate the price elasticity of consumption for each purchase occasion as

$$\frac{\partial E[c_{it}^h | \underline{p_{ts}}] | \underline{p_{ts}}]}{\partial \underline{p_{ts}}} \frac{\underline{p_{ts}}}{E[c_{it}^h | \underline{p_{ts}}]} = \beta_p^h \underline{p_{ts}}. \tag{3.6.3}$$

Price elasticity of consumption evaluated at the median price of \$2.08 is presented in Table 3.7. Column (1) in the table shows that when all households are assumed to be attentive storer type, 1% increase in price decreases consumption by 4.47%.

Column (2) in Table 3.6 shows the estimation results when the households are di-

Table 3.6: ML Estimation Results: Non-Storer Types Only

Parameter	Description	(1) Est	(1) Std	(2) Est	(2) Std
β_0^{A-NS}	preference to the PB product of A-NS	1.788	0.106	2.579	0.124
β_p^{A-NS}	sensitivity to price of A-NS	-2.126	0.052	-2.646	0.061
α^{A-NS}	dispersion parameter of A-NS	21.925	0.372	20.761	0.458
β_0^{IA-NS}	preference to the PB product of A-NS			0.535	0.202
β_p^{IA-NS}	sensitivity to price of A-NS			-1.234	0.097
α^{IA-NS}	dispersion parameter of A-NS			18.552	0.513
-logL	negative value of the log like function	575857		575506	

Note: Column (1) shows the estimation results when every household is assumed to be attentive non-storer types. Column (2) shows the estimation results when households are divided to attentive non-storer group and inattentive non-storer group. In total, 1,547 households who never purchased a large size jar in quantity surcharge weeks are assigned to the attentive non-storer group. The rest of 439 households are assigned to the inattentive non-storer group. Standard errors are obtained using numerical Hessian.

Table 3.7: Estimated Price Elasticity of Consumption

Type	(1)	(2)	(3)	(4)	(5)
A-NS	-4.4700	-5.5040			-1.1759
IA-NS		-2.5670			-0.0713
A-S			-4.0860	-5.4310	-6.2786
IN-S				-1.6651	-1.5022

Note: Price elasticity of consumption for each type is evaluated the median price of $2.08

vided into two groups: attentive non-storer type and inattentive non-storer type. attentive non-storer and attentive storer types are assumed to be attentive non-storer type, and the two other inattentive types are assumed to be inattentive non-storer type. The obtained estimates of the parameters are $\hat{\beta}^{A-NS} = (2.579, -2.646)$ and $\hat{\beta}^{IA-S} = (0.535, -1.124)$. The two types of households' price elasticity of consumption are -5.5040 and -2.5670, respectively, as shown in Table 3.7.

Next I estimate the model assuming all households are storer type. Column (1) in Table 3.8 shows the estimation results when all the households are assumed to be attentive storer type. The price elasticity of consumption when all households are assumed to be storer type is -4.0862 as shown in Table 3.7. This number is similar to the price elasticity obtained when all households are assumed to be attentive non-storer type, but slightly smaller.

Table 3.8: ML Estimation Results: Storer Types Only

Parameter	Description	(1) Est	(1) Std	(2) Est	(2) Std
β_0^{A-S}	preference to the PB product of A-S	1.091	0.111	2.109	0.132
β_p^{A-S}	sensitivity to price of A-S	-1.965	0.056	-2.611	0.067
α^{A-S}	dispersion parameter of A-S	28.806	0.550	26.651	0.654
β_0^{IA-S}	preference to the PB product of IA-S			-0.683	0.231
β_p^{IA-S}	sensitivity to price of IA-S			-0.801	0.115
α^{IA-S}	dispersion parameter of IA-S			25.381	0.805
-logL	negative value of the log like function	577761		577535	

Note: Column (1) shows the estimation results when every household is assumed to be attentive storer types. Column (2) shows the estimation results when households are divided to attentive storer group and inattentive storer group. In total, 1,547 households who never purchased a large size jar in quantity surcharge weeks are assigned to the attentive non-storer group. The rest of 439 households are assigned to the inattentive non-storer group. Standard errors are obtained using numerical Hessian.

Column (2) in Table 3.8 shows the estimation results when the households are divided into attentive storer and inattentive storer types. The obtained estimates are $\hat{\beta}^{A-S} = (2.109, -2.611)$ and $\hat{\beta}^{IA-S} = (-0.683, -0.801)$, and the price elasticities are -5.4308 and -1.6651, respectively. That is, when all the households are assumed to be storer types, when the minimum price increases by 1%, attentive and inattentive type households decrease their consumption quantities by -5.4308% and -1.6651%, respectively. Inattentive households are less sensitive to price change than attentive households are when they are assumed to be storer types.

Lastly, I estimate the model assuming the four groups of households as assigned in Table 3.5. The estimation results are presented in Table 3.9, and the estimated coefficients vary across household types. Price elasticities calculated using the obtained estimates are presented in Column (5) in Table(3.7). Price elasticity for each type is -1.1759, -0.0713, -6.2786, and -1.5022 for attentive non-storer type, inattentive non-storer type, attentive storer type, and inattentive store type, respectively.

Storer types are more sensitive to price change than non-storer types. For example, Attentive storer type households decrease consumption quantities by 6.2786% when

Table 3.9: ML Estimation Results: All Four Types

Parameter	Description	Est	Std
β_0^{A-NS}	preference to the PB product of A-NS	-1.412	0.210
β_p^{A-NS}	sensitivity to price of A-NS	-0.565	0.101
α^{A-NS}	dispersion parameter of A-NS	25.176	0.769
β_0^{IA-NS}	preference to the PB product of IA-NS	-2.442	0.681
β_p^{IA-NS}	sensitivity to price of IA-NS	-0.034	0.328
α^{IA-NS}	dispersion parameter of IA-NS	30.640	2.484
β_0^{A-S}	preference to the PB product of A-S	2.990	0.142
β_p^{A-S}	sensitivity to price of A-S	-2.965	0.073
α^{A-S}	dispersion parameter of A-S	20.959	0.557
β_0^{IA-S}	preference to the PB product of IA-S	-0.879	0.374
β_p^{IA-S}	sensitivity to price of IA-S	-0.712	0.185
α^{IA-S}	dispersion parameter of IA-S	30.520	1.467
-logL	negative value of the log likelihood function	577716	

Note: Standard errors are obtained using numerical Hessian.

price increases by 1%, while attentive non-storer type households decrease consumption quantities by only 1.1759%. When the price in the current period changes, non-storer type households adjust their consumption quantities for the current period. As for the storer type households, change in the current period price affects the consumption quantity in the current period if the price in the current period is the effective price in the current period. It also affects the consumption quantity in the next period if the price in the current period is the effective price in the next period.

Between the two non-storer types, attentive non-storer type households are more price sensitive than inattentive non-storer type households. Attentive non-storer type households decrease consumption quantities by 1.1759% when price increases by 1%, while inattentive non-storer decreases consumption quantities by 0.0713%. The same pattern exists between the two storer types. These results implies that inattentive households are insensitive to not only price different between package sizes but also the price level itself in each period.

3.7 Concluding Comments

I study the dynamic optimization problem consumers face for packaged goods in this chapter. I focus on the heterogeneity of consumers in storing and attentiveness based on the evidence I find from data. Quantity surcharges exist for packaged goods in addition to quantity discounts. Chapter 1 documents the high frequency of quantity surcharges in the peanut butter category. It also finds the strong evidence of heterogeneous consumer behavior when quantity surcharges exist: some consumers purchase multiple small size jars to avoid the surcharge, but some others still purchase large size jars and pay the surcharge.

The main contribution of this chapter is to develop a simple dynamic demand model that can capture the heterogeneity in consumer attention. There are four components of the model. First, the model is dynamic, as peanut butter is storable. Second, the model considers a multiple discrete choice in terms of quantity. Third, the model separates a purchase quantity decision and a package size choice. The fourth component is two-dimensional heterogeneity in consumers: attention and storability.

In each period, consumers decide consumption quantity, purchase quantity, and the number of jars of each size to purchase. The second and the third components together generate multiple jar purchases that are frequently observed in the data. Also, the third and fourth components leave room for inattentive type consumers to make inferior package size choices.

I impose each component to the model in the following way. First, I assume four types of consumers: attentive non-storer, inattentive non-storer, attentive storer, and inattentive storer. In order to handle multiple discreteness, I consider consumption quantity as a countable variable that follows a negative binomial distribution. Consumption quantity and purchase quantity are the same for non-storing type consumers, but not necessarily the same for storing type consumers.

As for a storing type consumers' dynamic problem, I introduce the concept of effective price and few additional assumptions to simplify the state space. For the given purchase quantity, consumers decide how many jars of each size to purchase. I take a behavioral approach and assume that an attentive type consumer chooses a package bundle that minimizes expenditure, and an inattentive type consumer chooses the one that minimizes the total number of jars to purchase.

I derive a log likelihood function using the distributional assumption on consumption quantity, and then estimate the parameters using the method of maximum likelihood. As household types are unobservable from the data, I estimate the partition of households minimizing the difference within a group in terms of the two characteristics of purchasing patterns. The estimation results suggest that storer type households are more sensitive to price than non-storer type households. Also, attentive type households are more price sensitive than inattentive type households.

Chapter 2 develops a demand model for packaged goods that captures the heterogeneity in attention. The model allows households to choose the product, quantity, and package bundle to purchase. However, the model omits the dynamics of household purchase and consumption decisions. In this chapter, I find the evidence that households stockpile peanut butter when the price is lower than usual and delay the next purchase. Hence, I introduce the model that simplifies the product choice problem but separates the purchase and consumption decisions in this chapter.

Chapter 2 shows that consumer inattention causes welfare loss to households. An interesting question is whether the magnitude of welfare loss varies by household storing behavior. In order to answer that question, one needs to extend the demand model to include multiple products. With the extended dynamic demand model, one can estimate the demand more accurately, and further analyze the welfare loss by storing behavior.

Bibliography

Jagdish Agrawal, Pamela E Grimm, and Narasimhan Srinivasan. Quantity surcharges on groceries. *Journal of consumer affairs*, 27(2):335–356, 1993.

Greg M Allenby, Thomas S Shively, Sha Yang, and Mark J Garratt. A choice model for packaged goods: Dealing with discrete quantities and quantity discounts. *Marketing Science*, 23(1):95–108, 2004.

Rick L Andrews and Ajay K Manrai. Mds maps for product attributes and market response: an application to scanner panel data. *Marketing Science*, 18(4):584–604, 1999.

Christine Boizot, Jean-Marc Robin, and Michael Visser. The demand for food products: an analysis of interpurchase times and purchased quantities. *The Economic Journal*, 111(470):391–419, 2001.

Bart J Bronnenberg, Michael W Kruger, and Carl F Mela. Database paperâĂŤthe iri marketing data set. *Marketing science*, 27(4):745–748, 2008.

A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.

Kwangpil Chang, Sivaramakrishnan Siddarth, and Charles B Weinberg. The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects. *Marketing Science*, 18(2):178–192, 1999.

Victor Chernozhukov and Han Hong. An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346, 2003.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–77, 2009.

Siddhartha Chib, PB Seetharaman, and Andrei Strijnev. Analysis of multi-category purchase incidence decisions using iri market basket data. In *Advances in Econometrics*, pages 57–92. Emerald Group Publishing Limited, 2002.

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

Sofronis Clerides and Pascal Courty. Sales, quantity surcharge, and consumer inattention. *Review of Economics and Statistics*, 2017.

Christopher T Conlon and Julie Holland Mortimer. An experimental approach to merger evaluation. Technical report, National Bureau of Economic Research, 2013.

Brenda Cude and Rosemary Walker. Quantity surcharges: are they important in choosing a shopping strategy? *The Journal of Consumer Affairs*, pages 287–295, 1984.

Peter R Dickson and Alan G Sawyer. The price knowledge and search of supermarket shoppers. *The Journal of Marketing*, pages 42–53, 1990.

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

Jean-Pierre Dubé. Product differentiation and mergers in the carbonated soft drink industry. *Journal of Economics & Management Strategy*, 14(4):879–904, 2005.

Tülin Erdem, Susumu Imai, and Michael P Keane. Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, 1(1):5–64, 2003.

Alan E Gelfand and Adrian FM Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

Bari A Harlam and Leonard M Lodish. Modeling consumers' choices of multiple items. *Journal of Marketing Research*, pages 404–418, 1995.

Igal Hendel. Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies*, 66(2):423–446, 1999.

Igal Hendel and Aviv Nevo. The post-promotion dip puzzle: What do the data have to say? *Quantitative Marketing and Economics*, 1(4):409–424, 2003.

Igal Hendel and Aviv Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673, 2006a.

Igal Hendel and Aviv Nevo. Sales and consumer inventory. *The RAND Journal of Economics*, 37(3):543–561, 2006b.

Igal Hendel and Aviv Nevo. Intertemporal price discrimination in storable goods markets. *American Economic Review*, 103(7):2722–51, 2013.

Tanjim Hossain and John Morgan. ... plus shipping and handling: Revenue (non) equivalence in field experiments on ebay. *Advances in Economic Analysis & Policy*, 5(2), 2006.

Jaehwan Kim, Greg M Allenby, and Peter E Rossi. Modeling consumer demand for variety. *Marketing Science*, 21(3):229–250, 2002.

Pankaj Kumar and Suresh Divakar. Size does matter: Analyzing brand-size competition using store level scanner data. *Journal of Retailing*, 75(1):59–76, 1999.

Sanghak Lee, Jaehwan Kim, and Greg M Allenby. A direct utility model for asymmetric complements. *Marketing Science*, 32(3):454–470, 2013.

Bartosz Mackowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803, 2009.

Puneet Manchanda, Asim Ansari, and Sunil Gupta. The âĂIJshopping basketâĂİ: A model for multicategory purchase incidence decisions. *Marketing science*, 18(2):95–114, 1999.

Jordi Mondria et al. Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 145(5):1837, 2010.

Kent B Monroe and Angela Y Lee. Remembering versus knowing: Issues in buyersâĂŹ processing of price information. *Journal of the Academy of Marketing Science*, 27(2): 207–225, 1999.

Robert W Nason and Albert J Della Bitta. The incidence and consumer perceptions of quantity surcharges. *Journal of retailing*, 59(2):40–54, 1983.

Martin Pesendorfer. Retail sales: A study of pricing behavior in supermarkets*. *The Journal of Business*, 75(1):33–66, 2002.

Ricardo Reis. Inattentive consumers. *Journal of monetary Economics*, 53(8):1761–1800, 2006.

Peter E Rossi, Greg M Allenby, and Rob McCulloch. *Bayesian statistics and marketing*. John Wiley & Sons, 2012.

Maryam Saeedi. Reputation and adverse selection, theory and evidence from ebay. *Theory and Evidence from eBay*, 2014.

Jorge M Silva-Risso, Randolph E Bucklin, and Donald G Morrison. A decision support system for planning manufacturers' sales promotion calendars. *Marketing Science*, 18 (3):274–300, 1999.

Christopher A Sims. Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356. Elsevier, 1998.

Christopher A Sims. Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690, 2003.

David E Sprott, Kenneth C Manning, and Anthony D Miyazaki. Grocery price setting and quantity surcharges. *Journal of Marketing*, 67(3):34–46, 2003.

Stanley M Widrick. Measurement of incidents of quantity surcharge among selected grocery products. *Journal of Consumer Affairs*, 13(1):99–107, 1979.

JR Wygant, A Keiling, CA Cattell, l M Johnson, RL Lysak, M Temerin, FS Mozer, CA Kletzing, JD Scudder, W Peterson, et al. Polar spacecraft based comparisons of intense electric fields and poynting flux near and within the plasma sheet-tail lobe boundary to uvi images: An energy source for the aurora. *Journal of Geophysical Research: Space Physics*, 105(A8):18675–18692, 2000.

Appendix A

Details

A.1 Data

A.1.1 Data Set Descriptions

In this thesis, I use five different data sets provided from IRI from 2008 to 2010 for the analyses. The first data set is product attributes data recorded at UPC level. The product attributes data originally contain 1,018 UPCs of peanut butter and peanut butter related products. For each UPC listed, information on its parent company, brand, product type, texture, flavor, and few other product characteristics is provided. As for product type, approximately 85% of UPCs belong to peanut butter and the rest belong to five other peanut butter related product types, such as peanut butter combo and peanut butter spread. There are 17 different textures and, ranging from super chunky and chunky to smooth and creamy. I later group them into two categories, creamy and chunky.

The second data set is store level data. It contains information on sales of peanut butter UPCs at store-week level. The available variables are store ID assigned by IRI, week, UPC, number of jars sold, total sales in dollars, and promotional activities such as feature and display. There are three types of stores in the market: two mass, three drug, and six grocery stores.[1] I focus on grocery stores only for the following reasons: 1)

[1] One of the three drug stores opened in 2009

household level data of grocery trips to mass stores are not available; 2) drug stores have minimal sales volume compared to grocery stores [2]; 3) grocery stores also have somewhat different pricing strategies and promotion activities than grocery stores. Dropping observations from mass stores and drug stores, the data contain six stores, 31 brands, 159 UPCs, and 42,634 sales observations. In the rest of paper, I call grocery stores as stores.

The rest of three data sets are household level: household trip data, household demographic characteristics data, and household peanut butter purchase data. Trip data contain records of trips to stores each household made for the period 2008-2010. For each household I can tell the store visited, the checkout time, and the total dollar amount spent on the trip. This trip data originally contain 5,727 households and 719,711 trip observations to both grocery stores and drug stores. I focus on trip observations to grocery stores only, and there are 5,702 households and 684,927 trip observations left after dropping the drug store observations. Not every household has complete three-year trip records: some have trip observations from one year and no record for the next year. I keep the households with at least one trip observation each year. As a result, 4,076 households and 680,851 trip observations remain.

Next, household demographic characteristics data provide information on household income, family size, age and education level of the household head, and few other characteristics. For the period 2008-2010, the demographic information was collected in Summer 2012, so the demographic characteristics for each household stay the same over the three years I analyze. There are 2,994 households listed on the data with at least one year observation. Using the fact that the demographic characteristics do not vary over the three years, I filled up the missing observations.

The last and the most important household level data are household purchase data. The data include the complete peanut butter product purchase records of household panels during the time period analyzed: UPC and the number of jars purchased, dollar amount paid, and the store, week, and minute where and when the purchase occurred.

[2] The sum of sales in dollar amount of all the drug stores during the three year time period in Eau Claire takes roughly 1% of the total market sales.

That is, each observation is at household-store-week and minute-UPC level. Thus, if a household visited a store at a certain time and purchased two peanut butter jars with different UPCs, this purchase event is recorded as two separate purchase observations. However, if a household purchased two jars of the same peanut butter (same UPC), then this purchase event is recorded as a single purchase observation.

There are initially 2,713 households with 27,838 purchase observations in the data. Those purchases occurred at three different types of stores: mass, drug, and grocery stores. However, the purchase observations from mass stores and drug stores are very minimal.[3] Due to this negligible number of observations, and also in order to maintain the coherence with the other data sets, I keep the purchase observations from grocery stores only. This leaves 2,713 households and 27,661 purchase observations.

A.1.2 Merging Data Sets

I merged the five different data sets into one data set in the following order: 1) household trip data and household demographic characteristics data; 2) household peanut butter purchase data and peanut butter product attributes data; 3) results of merge 1 and merge 2; 4) result of merge 3 and store peanut butter sales data. In the process of merging, I dropped the unmatched observations. Here are some details of each merge.

Merge 1: First, I merged the household demographic characteristics data to the household trip data. I dropped 88 households from the demographic characteristics data who do not have any matching trip observation in the trip data. Also 1,170 households (53,625 corresponding trip observations) from the household trip data have no demographic characteristics information available, so I dropped all of their trip observations. In addition, one household (45 trip observations) has most of demographic characteristics information available but no income information. Household income is one of the most important characteristics, so I dropped all of the her trip observations. As a result 2,905 households and 605,688 trip observations are left.

[3]Only 149 and 28 purchase observations occurred at drug stores and mass stores, respectively.

Merge 2: The second step is merging the household purchase data and the product attributes data. All the purchase observations from the household purchase data are successfully matched to the UPCs listed in the product attributes data. Households purchased four different kinds of product type: peanut butter, peanut butter combo, peanut butter spread, and peanut spread. I kept purchase observations of peanut butter and peanut butter spread types only. The reason is as follows: 1) peanut butter combo is a type of products that has peanut butter and jelly or peanut butter and chocolate spread in one jar; 2) peanut spread is usually flavored with honey or chocolate. However, peanut butter spread has no difference in observable characteristics than peanut butter. Also, the number of peanut butter jars that household purchased during the time period analyzed, peanut butter combo and peanut spread product types together take a small share.[4] After dropping, the same number of households remain, 2,713, but the number of purchase observations are reduced to 27,527.

Merge 3: The third step is to combine the two merged data sets from step 1 and step 2. I dropped 1,229 purchase observations with no matching trip observations. This yields 2,905 households in total with 606,833 trip observations. 2,638 households bought at least one peanut butter jar. The total number of purchase observations is 26,298, and they include 28 brands and 129 UPCs. However, there is an additional grocery store in this merged data set other than the six grocery stores listed in the store level peanut butter sales data. 22 households made 54 trips to the additional store during the time period analyzed, and made 2 peanut butter product purchases. In order to make it compatible to the store level peanut butter sales data, I dropped those 22 households. 2,883 households with 601,114 trip observations remain. 2,616 households purchased at least one peanut butter product, and these households made 26,024 purchase observations in total.

Merge 4: The last step is to merge the store level sales data and the combined household level data obtained by the merge 3. The final data set contains 42,634 sales obser-

[4] Households purchased 42,425 jars in total, and 125 jars are peanut butter combo type and 20 jars belong to peanut spread type.

vations from six stores in the Eau Claire market for the period 2008-2010, including 31 brands and 159 UPCs of peanut butter. There are 2,883 households in total with 601,114 trip observations. 2,616 households purchased peanut butter products at least once, and they made 26,024 purchase observations in total. Those purchase observations include 27 brands and 128 UPCs.

A.1.3 Data Cleaning

I cleaned the merged data to get them ready for the later analyses in the following order. First I dropped household with no peanut butter purchase. Then I dropped households who purchased too many jars on a single trip. I defined purchasing more than 6 jars of peanut butter on a single trip as too many. 50 households were dropped and 2,566 households with 24,829 purchase observations are left. The next step is to drop household who purchased too little or too many during the whole time period. I dropped households who purchased only one jar for the three year time period. I also dropped households who purchased more than 100 jars. That left us 2,369 households with 24,154 purchase observations.

One of the brands, Jif to Go, sells packs of 1.5 oz individual discs that are distinctive to rest of the jar peanut butters. Hence I dropped the purchase observations of Jif to Go, and 2,369 households with 24,083 purchase observations are left.

144 purchase observations have no matching sales information at stores. Some other purchase observations have matching sales information at stores, but the price level recorded at the household side and the one recorded at the stores are not exactly the same. 652 purchase observations have the price differences more than $0.05. I dropped the both cases, and 2,367 households with 23,287 purchase observations are left.

A.2 Summary Statistics of Households

This section provides the summary statistics of the households who purchased at least one item of the five products with multiple package sizes, as defined in Table 1.4.

Table A.1 shows the summary statistics of the household demographic information. The median income of the head of the household (HH) is 7, which represents the range of $35,000 to $44,999. The average family size is 2.4, and less than 25% of households have children. The median age of HH is 5, which falls in the range between 55 and 64. The median education level reached by HH is graduate high school.

Table A.1: Summary Statistics of Household Demographics

Variables	Obs	Mean	SD	Min	P25	P50	P75	Max
HH Income	2013	7.37	3.16	1	5	7	10	12
Family Size	2013	2.40	1.24	1	2	2	3	6
Num of Children	2013	0.27	0.59	0	0	0	0	3
HH Age	2008	4.78	1.14	1	4	5	6	6
HH Educ	1979	4.20	1.33	1	3	4	5	8
HH Race	2008	1.05	0.35	1	1	1	1	4

Note: The summary statistics are calculated from 2,013 households who purchased at least one jar of the five products of the focus. Five households have no information available in HH age and race, and 34 households (including those five households) have no information available in HH educ. HH Income is a category variable representing the combined pre-tax income of the head of household (HH). Income is equal to 1 if the combined pre-tax income of HH is in the range of $00,000 to $9,999 per year, 2 if in the range of $10,000 to $11,999, 3 if in the range of $12,000 to $14,999, 4 if in the range of $15,000 to $19,999, 5 if in the range of $20,000 to $24,999, 6 if in the range of $25,000 to $34,999, 7 if in the range of $35,000 to $44,999, 8 if in the range of $45,000 to $54,999, 9 if in the range of $55,000 to $64,999, 10 if in the range of $65,000 to $74,999, 11 if in the range of $75,000 to $99,999, and 12 if greater than or equal to $100,000. Family size and num of children represent the number of family members and the number of children in the household, respectively. HH Age is a category variable representing the age of HH. HH age is equal to 1 if the HH's age lies in the range of 18 to 24, 2 if 25 to 34, 3 if 35 to 44, 4 if 45 to 54, 5 if 55 to 64, and 6 if greater than equal to 65. HH Educ is a category variable representing the education level reached by HH. HH Educ is equal to 1 if some grade school or less is reached, 2 if grade school is completed, 3 if some high school, 4 if graduated high school. It is equal to 5 if technical school, 6 if some college, 7 if graduated from college, and 8 if post graduate work. Race is a categorical variable representing the HH's ethnicity. HH Race is equal to 1 if the HH is White, 2 if Black-African American, 3 if Hispanic, 4 if Asian, 5 if Other, 6 if American Indian-Alaska Native, and 7 if Native Hawaiian-Pacific Islands. 96.94% of the households in the panel have white HHs.

The household demographics of the panel data are broadly similar to the national population's. The median household income in the United States is 51,371 dollars in 2012 [5]. This number includes the income of the head of household's and the other household members who are 18 years or older, so the sample median of 35,000 to 44,999 dollars seem to be compatible. The summary statistics of the national population demographics are presented in Table A.2.

The family size of the household in the sample represent the national population

[5] Household Income: 2012, American Community Survey Briefs issued by U.S. Census Bureau

very well. The mean and the standard deviation are very similar to the national level, and also the median and the third quantile are the same. The number of children are somewhat less than the national level. This can be explained by the fact that the average age of the head of household in the data is higher than the national level. The most distinctive difference is the race the head of household's ethnicity. The predominant ethnicity of the head of household is White, which does not represent the diversity in the national population.

Table A.2: Summary Statistics of National Population Demographics

Variables	Mean	SD	Min	P25	P50	P75	Max
Family Size	2.48	1.38	1	1	2	3	6
Num of Children	0.60	1.00	0	0	0	1	4
HH Age	3.60	1.31	1	3	4	5	6
Education	5.27	1.60	1	4	6	7	7
HH Race	1.64	1.19	1	1	1	2	7

Note: The summary statistics are calculated based on the America's Families and Living Arrangements: 2012 issued by U.S. Census Bureau. Family Size, Num of Children, HH Age, and HH Race are defined the same as in Table A.1. Education is the education achieved by 25 years or order population. Education is equal to 1 if 0 to 4 years of Elementary school, 2 if 5 to 8 years of Elementary school, 4 if 1 to 3 years of High school, 4 if 4 years of High school, 6 if 1 to 3 years of College, and 7 if 4 years or more of College.

Table A.3 shows the summary statistics of grocery shopping trips households make for the three year time period. Households make 227 trips to grocery stores on average, which means 1.45 trips per week on average. The average number of weeks between trips being less than 1 also indicates that households go on more than one grocery shopping in a week on average.

There are six grocery stores in the Eau Clare market. The median number of stores a household visits is four, and 20.67% households visit all six stores. Households spend $42.01 per grocery shopping and $53.58 per week on average.

The summary statistics of peanut butter purchases are presented in Table A.4. Households frequently purchase multiple peanut butter jars on a trip, as the average number of jars purchased on a trip is 1.6. The average quantity of peanut butter purchased on a trip is 18.72 oz. This implies that households purchase small size jars more

Table A.3: Summary Statistics of Grocery Shopping Trips

Variables	Mean	SD	Min	P25	P50	P75	Max
Num of Trips	226.93	130.95	18.00	140.00	197.00	276.00	1317.00
Num of Stores	4.24	1.38	1.00	3.00	4.00	5.00	6.00
Ave Num of Weeks Btw Trips	0.86	0.44	0.12	0.55	0.77	1.08	5.35
Ave Exp per Trip	42.01	22.66	5.34	25.62	37.46	53.72	233.19
Ave Exp per Week	53.58	28.59	5.87	32.93	48.28	68.00	252.51

Note: The summary statistics are based on total grocery shopping trips 2,013 households made for the three years.

often than large size jars.

Households purchase a subsequent amount of peanut butters, 1 jar minimum to 95 jars for the three year time period. Households purchase 13.44 jars and 154.76 oz of peanut butter on average. Some households stick with only one UPC or one brand (13.61% and 23.6%, respectively), but the rest of households enjoy some variety. The median household purchases three different UPCs and two different brands.

Table A.4: Summary Statistics of Peanut Butter Purchases

Variables	Mean	SD	Min	P25	P50	P75	Max
A. On a Trip							
Ave Num of Jars	1.60	0.45	1.00	1.31	1.57	1.88	6.00
Ave Oz	18.72	2.80	15.28	17.03	17.66	19.32	28.00
B. For Three Years							
Total Num of Jars	13.33	12.62	1.00	5.00	9.00	18.00	95.00
Total Oz	154.76	141.38	16.30	56.00	115.80	204.60	1552.76
Total Num of UPCs	3.84	2.31	1.00	2.00	3.00	5.00	14.00
Total Num of Brands	2.53	1.18	1.00	2.00	2.00	3.00	6.00

Note: The summary statistics are based on the peanut butter purchases 2,013 households made for the three years. The purchase observations include 17 products defined, and not limited to 5 products with multiple package sizes.

A.3 Estimation Details

A.3.1 Estimation Procedure

I estimate the Hierarchical Bayes models in Chapter 2 using Markov Chain Monte Carlo methods. In this section I show the detailed estimation procedure of Basic Model presented in Section 2.3.1. I simulate draws from the posterior distribution as follow:

1. Generate $\{\theta_h', T_h\} = \{\alpha_h, \beta_h', T_h\}$ for $h = 1, \ldots, H$ households, using the Metropolis-Hastings (M-P) algorithmn with a random walk chain. The posterior is proportional to the product of the likelihood and the prior

   $$\pi(\theta_h, T_h) \propto \Pi_j \mathrm{Pr}(x_{ij}|\theta_h', T_h) \times \pi(\theta_h|\bar{\theta}, V_\theta) \times \pi(T_h|a, b),$$

   where $\Pi_j \mathrm{Pr}(x_{ij}|\theta_h', T_h)$ is the likelihood, with the choice probability $\mathrm{Pr}(x_{ij}|\theta_h', T_h)$ given by Equation (2.3.8). $\pi(\theta_h|\bar{\theta}, V_\theta)$ and $\pi(T_h|a, b)$ are the prior distributions of heterogeneity and budget limit, respectively.

2. Generate $\bar{\theta}$ for the given draws of $\{\theta_h\}$ and V_θ

   $$\pi(\bar{\theta}|\{\theta_h\}, V_\theta) = \mathrm{Normal}(H(V_\theta + A)^{-1}(HV_\theta \sum_h \theta_h / H + A), H(V_\theta + A)^{-1})$$

   where $A = 100I^{-1}$ and H is the number of households.

3. Generate V_θ for the given draws of $\{\theta_h\}$ and $\bar{\theta}$ generated in Step 2

   $$\pi(V_\theta|\{\theta_h\}, \bar{\theta}) = IW(g_0 + H, G_0 + \sum_h (\theta_h - \bar{\theta})(\theta_h - \bar{\theta})'),$$

   where $g_0 = 50$ and $G_0 = 50I$

4. Repeat

For each household h, I start with the initial guess of $\theta_0 = (1, 2, 2, 1, 1, -2, 1, 2, -2, -2)'$

and $T_0 = 20$ and prior distributions of

$$\theta_h \sim N(\theta_0, 25I), T_h \sim N(10, 9).$$

A.3.2 Convergence Check

Monitoring convergence is important to ensure the validity of the posterior obtained by MCMC methods. Plotting the sequence of MCMC output is a good way to start in practice (Rossi, Allenby, and McCulloch (2012)). Figure A.1 shows the time series plots of MCMC draws for Basic Model. The figure provides strong evidence of convergence. The mean parameters begin to show ergodicity soon after the initial draws, revisiting the same region over and over again.

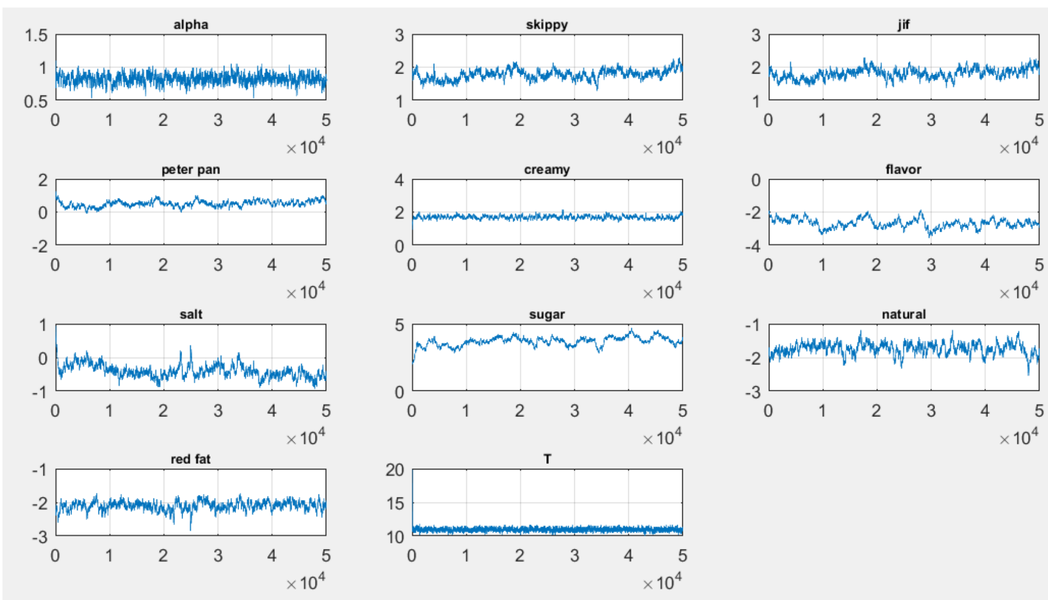


Figure A.1: Time Series Plots of Average Parameters Across Households by MCMC Draws

Note: The figure shows the MCMC draws of each parameter in Basic Model. Since parameters are drawn for each household, the average across households is used for the plots. 50,000 iterations are conducted in total, and the first 25,000 iterations are used as a burn-in period.

Another way to check convergence is starting the chain from different initial values (Gelman and Rubin (1992)). The original estimation results of Basic Model presented in Column (1) of Table 2.1 use $\theta_0 = (1, 2, 2, 1, 1, -2, 1, 2, -2, -2)'$ as a starting value of θ_h for each household. I consider three alternative starting points: $-\theta_0$, θ_0+1, and θ_0-1. The estimation results are displayed in Table A.5.

Table A.5: Aggregate Coefficient Estimates of Basic Model with Various Starting Values

Starting Value of θ_h	(1) $-\theta_0$		(2) θ_0+1		(3) θ_0-1	
$\alpha^* = \ln(\alpha_z/\alpha_x)$	0.8242	(0.0641)	0.8289	(0.0594)	0.8232	(0.0619)
Skippy	1.7376	(0.1163)	1.7908	(0.1317)	1.7854	(0.1236)
Jif	1.7713	(0.1160)	1.8299	(0.1329)	1.8224	(0.1180)
Peter Pan	0.4797	(0.1508)	0.5279	(0.1542)	0.5261	(0.1513)
Creamy	1.6934	(0.1073)	1.6906	(0.1074)	1.6734	(0.0958)
Flavor	-2.6138	(0.2138)	-2.6729	(0.2293)	-2.6581	(0.2178)
Salt	-0.4025	(0.1442)	-0.4398	(0.1480)	-0.4363	(0.1545)
Sugar	3.7743	(0.3228)	3.7931	(0.2772)	3.7855	(0.3110)
Natural	-1.713	(0.1517)	-1.6839	(0.1491)	-1.6644	(0.1459)
Reduced Fat	-2.0558	(0.1185)	-2.0741	(0.1222)	-2.0698	(0.1235)
T	10.9498	(0.2045)	10.959	(0.2010)	10.9503	(0.2112)

Note: Estimation was conducted with a subsample of 200 households. Standard deviations are reported in parentheses. The three columns shows the aggregate coefficient estimates of Basic model with three different starting values of θ_h. The original estimation whose results are presented in Column (1) in Table 2.1 uses $\theta_0 = (1,2,2,1,1,-2,1,2,-2,-2)'$ as a starting value. Column (1)-(3) use $-\theta_0$, θ_0+1, and θ_0-1 as starting values, respectively.

The estimate results suggest successful convergence. The estimation results across Column (1)-(3) are very similar to each other, and also similar to the estimation results obtained with the original initial value θ_0. Figure A.2 shows how MCMC draws for each parameter vary through iterations, with an alternative starting value $-\theta_0$. We can see that the chains explore the parameter space at first but then quickly converge to the ergodic distributions.

A.4 Identification

I first introduce the identification of the standard negative binomial maximum likelihood estimator, and then show how the estimator maximizing the log likelihood function in equation (3.5.5) is related to the standard negative binomial maximum likelihood estimator.

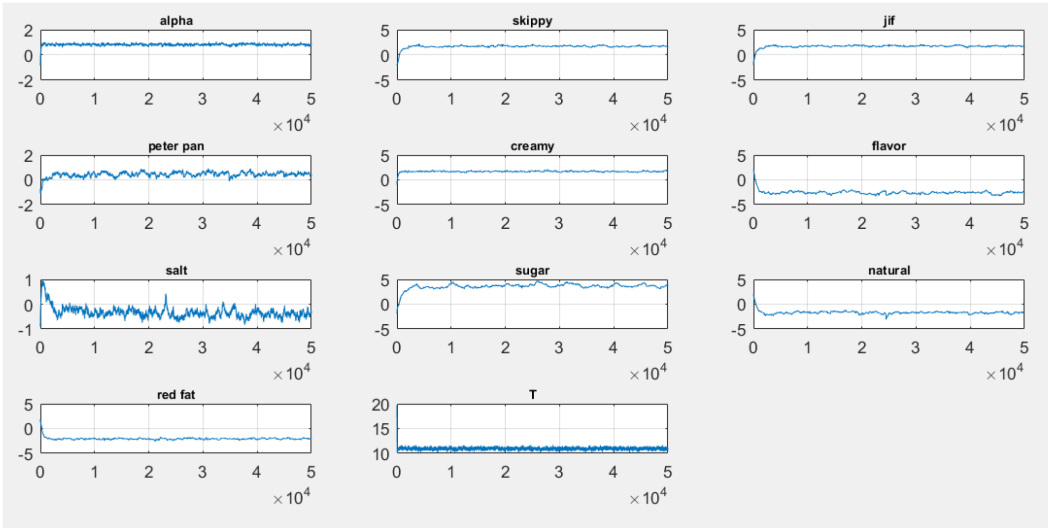


Figure A.2: Time Series Plots of Average Parameters Across Households by MCMC Draws with an Alternative Staring Value

Note: The figure shows the MCMC draws of each parameter in Basic Model, with an alternative starting value of θ_h, $-(1,2,2,1,1,-2,1,2,-2,-2)'$. Since parameters are drawn for each household, the average across households is used for the plots. 50,000 iterations are conducted in total, and the first 25,000 iterations are used as a burn-in period.

Suppose q follows a negative binomial distribution with density

$$h(q|\mu,\alpha) = \frac{\Gamma(q+\alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(q+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{q}, \tag{A.4.1}$$

where $\alpha \geq 0$, $q = 0,1,2,\ldots$, and $\mu = \exp(x'\beta)$. Cameron and Trivedi (2013) show that $\Gamma(q+\alpha^{-1})/\Gamma(\alpha^{-1}) = \prod_{j=0}^{q-1}(j+\alpha^{-1})$ when q is an integer, which means

$$\log\left(\frac{\Gamma(q+\alpha^{-1})}{\Gamma(\alpha^{-1})}\right) = \sum_{j=0}^{q-1}\log(j+\alpha^{-1}). \tag{A.4.2}$$

Then the log likelihood function can be written as

$$\begin{aligned} \log L(\alpha,\beta) &= \sum_{i=1}^{n}\log h(q_i|\mu_i,\alpha) \qquad \text{(A.4.3)} \\ &= \sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{q_i-1}\log(j+\alpha^{-1})\right) - \log q_i! \right. \\ &\left. -(q_i+\alpha^{-1})\log(1+\alpha\exp(x_i'\beta)) + q_i\log(\alpha) + q_i x_i'\beta\right\}. \end{aligned}$$

The negative binomial maximum likelihood estimator $(\hat{\alpha}, \hat{\beta})$ is the solution to the first-order conditions

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{q_i - \mu_i}{1+\alpha\mu_i} x_i = 0, \\
&\sum_{i=1}^{n} \left\{ \frac{1}{\alpha^2} \left(\log(1+\alpha\mu_i) - \sum_{j=0}^{q_i - 1} \frac{1}{(j+\alpha^{-1})} \right) + \frac{q_i - \mu_i}{\alpha(1+\alpha\mu_i)} \right\} = 0.
\end{aligned} \tag{A.4.4}$$

Now I show the identification of the parameters (α^h, β^h) for the two non-storer types $(h = A - NS, IA - NS)$. The log likelihood function is given in equation (3.5.5). Note that the probability of jar size choice, ϕ^h_{it}, does not depend on the parameters (α^h, β^h). Let $x'_{it} = (1, \underline{p_{ts}})'$. The maximum likelihood estimator $(\hat{\alpha}^h, \hat{\beta}^h)$ for the two non-storer types is the solution to the first order conditions

$$\begin{aligned}
&\sum_{t=1}^{R} \sum_{i=1}^{N^h} \frac{q^h_{it} - \mu^h_{it}}{1+\alpha^h \mu^h_{it}} x_{it} = 0, \\
&\sum_{t=1}^{R} \sum_{i=1}^{N^h} \left\{ \frac{1}{(\alpha^h)^2} \left(\log(1+\alpha^h \mu^h_{it}) - \sum_{j=0}^{q^h_{it} - 1} \frac{1}{(j+(\alpha^h)^{-1})} \right) + \frac{q^h_{it} - \mu^h_{it}}{\alpha^h(1+\alpha^h \mu^h_{it})} \right\} = 0,
\end{aligned} \tag{A.4.5}$$

which is the negative binomial maximum likelihood estimator.

Next I show the identification of the parameters (α^h, β^h) for the two storer types $(h = A - S, IA - S)$. The log likelihood function is given in equation (3.5.5). The first order condition with respect to β^h is

$$\begin{aligned}
&\frac{\partial \log L^h(\alpha^h, \beta^h)}{\partial \beta^h} \\
&= \sum_{t=1}^{R} \sum_{i=1}^{N^h} \left\{ (1 - \mathbf{1}(SN) - \mathbf{1}(NS)) \frac{\partial \log h(q^h_{it} | \mu^h_{it}, \alpha^h)}{\partial \beta^h} \right. \\
&\qquad \left. + \mathbf{1}(SN) \frac{\partial \log(\mathbf{1}(q^h_{it} = 0))}{\partial \beta^h} + \mathbf{1}(NS) \frac{\partial \log D}{\partial \beta^h} \right\} = 0, \\
&\text{where} \quad D = \sum_{j=0}^{q^h_{it}} h(j | \mu^h_{it}, \alpha^h) h(q^h_{it} - j | \mu^h_{it}, \alpha^h).
\end{aligned} \tag{A.4.6}$$

The three partial derivatives in the equation (A.4.6) are

$$\begin{aligned}
\frac{\partial \log h(q_{it}^h|\mu_{it}^h,\alpha^h)}{\partial \beta^h} &= \frac{q_{it}^h-\mu_{it}^h}{1+\alpha^h\mu_{it}^h}x_{it}, \\
\frac{\partial \log(\mathbf{1}(q_{it}^h=0))}{\partial \beta^h} &= 0, \\
\frac{\partial \log D}{\partial \beta^h} = \frac{1}{D}\frac{\partial D}{\partial \beta^h} &= \frac{q_{it}^h-2\mu_{it}^h}{1+\alpha^h\mu_{it}^h}x_{it}.
\end{aligned} \tag{A.4.7}$$

Plugging the equation (A.4.7) to equation (A.4.6) yields

$$\begin{aligned}
&\sum_{t=1}^{R}\sum_{i=1}^{N^h}\left\{(1-\mathbf{1}(SN)-\mathbf{1}(NS))\left(\frac{q_{it}^h-\mu_{it}^h}{1+\alpha^h\mu_{it}^h}x_{it}\right)+\mathbf{1}(NS)\left(\frac{q_{it}^h-2\mu_{it}^h}{1+\alpha^h\mu_{it}^h}x_{it}\right)\right\}=0 \\
&\Rightarrow \sum_{t=1}^{R}\sum_{i=1}^{N^h}\left(\frac{(1-\mathbf{1}(SN))q_{it}^h-(1-\mathbf{1}(SN))+\mathbf{1}(NS))\mu_{it}^h}{1+\alpha^h\mu_{it}^h}\right)x_{it}=0.
\end{aligned} \tag{A.4.8}$$

The first order condition with respect to α^h is

$$\begin{aligned}
&\frac{\partial \log L^h(\alpha^h,\beta^h)}{\partial \alpha^h} \\
&=\sum_{t=1}^{R}\sum_{i=1}^{N^h}\Bigg\{(1-\mathbf{1}(SN)-\mathbf{1}(NS))\frac{\partial \log h(q_{it}^h|\mu_{it}^h,\alpha^h)}{\partial \alpha^h} \\
&\qquad +\mathbf{1}(SN)\frac{\partial \log(\mathbf{1}(q_{it}^h=0))}{\partial \alpha^h}+\mathbf{1}(NS)\frac{\partial \log D}{\partial \alpha^h}\Bigg\}=0, \\
&\text{where} \quad D=\sum_{l=0}^{q_{it}^h} h(l|\mu_{it}^h,\alpha^h)h(q_{it}^h-l|\mu_{it}^h,\alpha^h).
\end{aligned} \tag{A.4.9}$$

The three partial derivatives in equation (A.4.9) are

$$\frac{\partial \log h(q_{it}^h|\mu_{it}^h,\alpha^h)}{\partial \alpha^h}=\left\{\frac{1}{(\alpha^h)^2}\left(\log(1+\alpha^h\mu_{it}^h)-\sum_{j=0}^{q_{it}^h-1}\frac{1}{(j+(\alpha^h)^{-1})}\right)+\frac{q_{it}^h-\mu_{it}^h}{\alpha^h(1+\alpha^h\mu_{it}^h)}\right\}, \tag{A.4.10}$$

$$\begin{aligned}
\frac{\partial \log(\mathbf{1}(q_{it}^h=0))}{\partial \alpha^h} &= 0, \\
\frac{\partial \log D}{\partial \alpha^h} &= \frac{1}{D}\frac{\partial D}{\partial \alpha^h},
\end{aligned}$$

where the last term can be written as

$$\frac{\partial D}{\partial \alpha^h} = \sum_{l=0}^{q_{it}^h} \left\{ h(l|\mu_{it}^h, \alpha^h) \frac{\partial h(q_{it}^h - l|\mu_{it}^h, \alpha^h)}{\partial \alpha^h} + \frac{\partial h(l|\mu_{it}^h, \alpha^h)}{\partial \alpha^h} h(q_{it}^h - l|\mu_{it}^h, \alpha^h) \right\} \tag{A.4.11}$$

$$= \left(\sum_{l=0}^{q_{it}^h} h(l|\mu_{it}^h, \alpha^h) h(q_{it}^h - l|\mu_{it}^h, \alpha^h) \right) \times \left(\frac{1}{(\alpha^h)^2} 2\log(1+\alpha^h \mu_{it}^h) + \frac{q_{it}^h - 2\mu_{it}^h}{\alpha^h(1+\alpha^h\mu_{it}^h)} \right)$$

$$- \sum_{l=0}^{q_{it}^h} \left\{ h(l|\mu_{it}^h, \alpha^h) h(q_{it}^h - l|\mu_{it}^h, \alpha^h) \frac{1}{(\alpha^h)^2} \times \left(\sum_{j=0}^{q_{it}^h - l - 1} \frac{1}{(j+(\alpha^h)^{-1})} + \sum_{j=0}^{l-1} \frac{1}{(j+(\alpha^h)^{-1})} \right) \right\}$$

$$\Rightarrow \frac{\partial \log D}{\partial \alpha^h} = \left(\frac{1}{(\alpha^h)^2} 2\log(1+\alpha^h\mu_{it}^h) + \frac{q_{it}^h - 2\mu_{it}^h}{\alpha^h(1+\alpha^h\mu_{it}^h)} \right) - \frac{\sum_{l=0}^{q_{it}^h} \left\{ h(l|\mu_{it}^h, \alpha^h) h(q_{it}^h - l|\mu_{it}^h, \alpha^h) \frac{1}{(\alpha^h)^2} \left(\sum_{j=0}^{q_{it}^h-l-1} \frac{1}{(j+(\alpha^h)^{-1})} + \sum_{j=0}^{l-1} \frac{1}{(j+(\alpha^h)^{-1})} \right) \right\}}{\sum_{l=0}^{q_{it}^h} h(l|\mu_{it}^h, \alpha^h) h(q_{it}^h - l|\mu_{it}^h, \alpha^h)}.$$

Thus

$$\sum_{t=1}^{R} \sum_{i=1}^{N^h} \{ (1 - \mathbf{1}(SN) - \mathbf{1}(NS)) \times \tag{A.4.12}$$

$$\left[\frac{1}{(\alpha^h)^2} \left(\log(1+\alpha^h\mu_{it}^h) - \sum_{j=0}^{q_{it}^h - 1} \frac{1}{(j+(\alpha^h)^{-1})} \right) + \frac{q_{it}^h - \mu_{it}^h}{\alpha^h(1+\alpha^h\mu_{it}^h)} \right]$$

$$+\mathbf{1}(NS) \left[\left(\frac{1}{(\alpha^h)^2} 2\log(1+\alpha^h\mu_{it}^h) + \frac{q_{it}^h - 2\mu_{it}^h}{\alpha^h(1+\alpha^h\mu_{it}^h)} \right) \right.$$

$$\left. \left. - \frac{\sum_{l=0}^{q_{it}^h} \left\{ h(l|\mu_{it}^h, \alpha^h) h(q_{it}^h - l|\mu_{it}^h, \alpha^h) \frac{1}{(\alpha^h)^2} \left(\sum_{j=0}^{q_{it}^h-l-1} \frac{1}{(j+(\alpha^h)^{-1})} + \sum_{j=0}^{l-1} \frac{1}{(j+(\alpha^h)^{-1})} \right) \right\}}{\sum_{l=0}^{q_{it}^h} h(l|\mu_{it}^h, \alpha^h) h(q_{it}^h - l|\mu_{it}^h, \alpha^h)} \right] \right\} = 0.$$

The maximum likelihood estimator $(\hat{\alpha}^h, \hat{\beta}^h)$ for the two storer types is the solution to the first order conditions in equation (A.4.8) and equation (A.4.12).