Essays on Dynamic Demand Estimation

by

Yucai Emily Wang

Department of Economics
Duke University

Date: ______________________

Approved:

____
Andrew Sweeting, Supervisor

____
Patrick Bayer

____
Arie Beresteanu

____
James Roberts

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2011

Abstract

Essays on Dynamic Demand Estimation

by

Yucai Emily Wang

Department of Economics
Duke University

Date: ______________________

Approved:

____
Andrew Sweeting, Supervisor

____
Patrick Bayer

____
Arie Beresteanu

____
James Roberts

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2011

Abstract

This dissertation consists of three chapters relating to dynamic demand models of storable goods and their application to taxes that are imposed on soft drinks. Broadly speaking, the first chapter builds the estimation strategy for dynamic demand models of storable goods that allows for unobservable heterogeneous preferences in households tastes. The second chapter uses the estimation strategy developed in the first chapter to study the policy implications of taxes that are imposed on sugary soft drinks. The last chapter explores and provides an explanation for the level of pass-through for soda taxes.

To be more specific, the first chapter develops techniques for incorporating systematic brand preferences in dynamic demand models of storable goods. Dynamic demand models are important for correctly measuring price elasticities of products that can be stockpiled. However, most of the literature excludes systematic preferences over consumers brand tastes. This chapter resolves this issue by incorporating random coefficient Logit models into a dynamic demand framework and hence allows for realistic demand substitution patterns. It builds on Hendel and Nevos 2006 Econometrica paper, where the authors introduce a model of dynamic demand that flexibly incorporates observable heterogeneity and estimates it via a three-step procedure that separates brand and volume choices. While a powerful tool, this method is tricky to execute. Therefore, this chapter also discusses the difficulties that may face implementers.

The second chapter predicts the effects of taxes on sugar sweetened soft drinks (sugar taxes) on both total consumption and the welfare of different types of consumers. It specifies and estimates a structural dynamic demand model of storable goods with rational and forward-looking households. It flexibly incorporates persistent heterogeneous consumer preferences and develops a computationally attractive method for estimating its parameters. Sugar taxes have been proposed at both the national and state-level, and passed in three states, as a means of slowing or reversing the growth in obesity and diabetes. To accurately analyze the effects of these policies, this chapter takes two specific aspects of soft drinks into account: storability and differentiation. It compares the results from this model to two benchmark studies: a static model with consumer heterogeneity and a dynamic model without households persistent heterogeneous tastes. It finds that failing to account for dynamics (i.e. storability) results in overestimated reduction in consumption and failing to account for persistent heterogeneous preferences (i.e. differentiation) results in overestimated reduction in consumption and underestimated welfare loss. The model and method developed here are readily applicable to many studies involving storable goods, such as firms optimal pricing behavior and anti-trust policies analyses.

The third and last chapter focuses on the incidence of soda taxes by studying the pass-through level of these taxes. It lays out a framework for thinking about the determinants of the pass-through level. More specifically, it builds theoretical models that examine the pass-through under more complex supply structures with multiple manufactures and retailers. In addition to providing some intuition behind theoretical predictions of the models, this chapter also presents empirical results found in the data along with their implications.

To my beloved mother, Ying Liu.

Contents

Abstract **iv**

List of Tables **x**

List of Figures **xii**

Acknowledgements **xiii**

1 Dynamic Demand Models of Storable Goods with Systemic Brand Preferences **1**

1.1 Introduction 1

1.2 Section I 4

1.2.1 Motivation 6

1.2.2 Model 8

1.2.3 Implications of Model 11

1.3 Section II 12

1.3.1 Model & Computational Structure of HN 12

1.3.2 Numerical Analysis of HN's Estimation 18

1.3.3 Improved Computation Method 25

1.3.4 Application to IRI Data 41

1.4 Conclusion 49

2 Estimating the Distributional Impact of Taxes on Storable Goods: A Dynamic Demand Model with Random Coefficients **50**

2.1 Introduction 50

2.2 Literature 55
2.3 Data 57
2.3.1 General Description 58
2.3.2 Stockpiling Behavior 64
2.3.3 Persistent Heterogeneous Preferences 66
2.4 Model 67
2.4.1 Overview 67
2.4.2 Model Setup 69
2.5 Estimation 75
2.5.1 Overview 76
2.5.2 Estimation Procedure 81
2.6 Results 86
2.6.1 Parameter Estimates 87
2.6.2 Model Fit 91
2.7 Policy Study 92
2.7.1 Price Elasticities 92
2.7.2 Reduction in Consumption 94
2.7.3 Welfare Loss 98
2.7.4 Post Tax Market Shares 101
2.7.5 Benchmark Studies 102
2.8 Concluding Remarks 108

3 Incidence of Taxes on Retail Soft Drinks: Theory & Evidence from Scanner Data 110

3.1 Introduction 110
3.2 Data 111
3.2.1 IRI Data 111

3.2.2 ImpacTeen Data 112

3.2.3 Census Data 113

3.3 Empirical Analysis 113

3.3.1 Large vs. Small Stores 114

3.3.2 Rural vs. Urban Locations 119

3.4 Conclusion 123

Bibliography **124**

Biography **126**

List of Tables

1.1 Dynamic parameter estimates obtained using codes and data posted on Econometrica Supplement Material to HN 2006. The parameters are: γ is the marginal utility of consumption, β_1 is the linear cost of inventory, and β_2 is the quadratic cost of inventory. 19

1.2 Final parameter estimates using initial guess 25% above and below the estimates given in HN 2006. No other changes were made in the code posted on Econometrica. 20

1.3 Decreasing the starting guesses by 10%. The above lists final parameter estimates of household type three after tightening the convergence criteria to 1e-5. 21

1.4 Decreasing the starting guesses by 10%. The above lists final parameter estimates of household type three after increasing the consumption shock draws to 10,000. 21

2.1 Household Demographics 59

2.2 Leading Soft Drinks Market Share 61

2.3 Maximum Household Vol Purchase 62

2.4 Price Variation 64

2.5 Regression of Quantity Purchased 66

2.6 Conditional Probability of Purchase 67

2.7 Specification of Product Utilities for Different Household Types . . . 72

2.8 Mean and Standard Deviations of the Distribution of Parameters . . 87

2.9 Shares of Each Household Type 90

2.10 Observed and Predicted Market Shares 91

2.11 Long Run Price Elasticities 93

2.12 Post-Tax Reduction in Consumption of Both Taxes for All Pass-Through Levels and Income Brackets 96

2.13 Compensating Variation Both Tax Proposals for All Pass-Through Levels and Income Brackets 100

2.14 Tax as a Percentage of Total Spending 101

2.15 Market Shares by Brand (10% Sales Tax) 102

2.16 Predicted Reduction in Consumption 105

2.17 Predicted Reduction in Consumption 107

2.18 Average Compensating Variation 107

3.1 Significant Soda Tax by State 113

3.2 Small Store Diff-n-Diff Regression Results: 118

3.3 Large Store Diff-n-Diff Regression Results: 118

3.4 Simulation Parameters 121

3.5 Simulation Results for Different Cases 121

3.6 Diff-n-Diff Regression Results for Different States: 123

List of Figures

1.1 Likelihood Plot of α 22

1.2 Likelihood Plot of β_2 23

1.3 Likelihood Plot w.r.t. α-changes 24

1.4 Likelihood Plot w.r.t. β_2-changes 25

1.5 Total Utility w.r.t. to Consumption 28

1.6 Properties of HN Utility Function 29

1.7 Properties of Corrected Utility Function 32

1.8 Sensitivity Plot 36

1.9 3D Likelihood Plot - β_1 Constant 39

1.10 Likelihood Plot of HN w.r.t. Cost of Inventory 40

1.11 Likelihood Plot of Alternative w.r.t. Cost of Inventory 41

2.1 Volume of cola drink sales by week 63

2.2 Sales Volume of Regular Coke with and without Price Reductions . . 65

Acknowledgements

I deeply appreciate the guidance and advice provided to me by Andrew Sweeting as well as Arie Beresteanu, James Roberts, and Chris Timmins. I sincerely thank Pat Bayer, Federico Bugni, Paul Ellickson, Gautam Gowrisankaran, Carl Mela, Ken Wilbur as well as all seminar attendants for helpful suggestions and fruitful discussions. I am grateful for the opportunity to work with the IRI data, provided by Paul Ellickson, Carl Mela, and Andrew Sweeting. Furthermore, I would like to thank Christoph Bauner for the support he provided throughout my Ph.D. studies.

1

Dynamic Demand Models of Storable Goods with Systemic Brand Preferences

1.1 Introduction

A key limitation of static demand models is their inability to correctly measure the effects of intertemporal substitution. When the product under study can be stored, static models are misspecified and hence produce inaccurate long run price elasticities. Since long run price elasticities are used in many applications such as antitrust policy analyses, consumer welfare studies, and effects of promotional activities, misestimated long run price elasticities can have significant consequences. Therefore, it is essential to be able to implement dynamic models of demand for goods that are storable. Please note that in this paper, I make a distinction between storable and durable goods. Durable goods are ones like cars, which don't vanish through use. Storable goods, on the other hand, are more like laundry detergent or soft drinks, where a portion of it can be consumed every period. This paper concentrates primarily on storable goods.

There are a few difficulties facing researchers interested in the implementing dynamic demand models of storable goods. First of all, these models are computationally burdensome to estimate. The products under study usually have many characteristics, making computing transitions of states spaces very time consuming. Secondly, many necessary variables are unobservable to the econometrician. For example, consumption decisions of laundry detergent or inventory level of soft drinks are usually known to the consumer but not the econometrician. Recently, dynamic demand models have experienced a boom. Many papers have incorporated dynamics into demand estimations. Most of them implement demand models of durable goods. I list only a few here. Goettler and Gordon (2008) studies competition and innovation of microprocessors. They model dynamic behavior of consumers. Gowrisankaran and Rysman (2007) estimates a dynamic model of consumer preferences for new durable goods. More recently, Conlon (2010) estimates a dynamic demand model in order to study market power. Few, however, have modeled and estimated demand for storable goods.

The most relevant papers in this area, listed in chronological order, are Erdem, Imai & Keane (2003), Hendel and Nevo (2006), and Hartmann and Nair (2007). The last paper estimates a dynamic demand model that incorporates brand switching of tied goods. Hartmann and Nair analyze whether manufacturers can achieve their desired pricing strategy of cheaper primary goods and more expensive aftermarket goods when downstream players have market power and misaligned incentives. Since the product under study are razor blades, the demand model incorporates both durability and storability aspects.

The first two papers both estimate dynamic demand models of storable goods. However they take two different approaches. Erdem, Imai, & Keane implement a complex model and simplify the computational burden by assuming that consumption of each brand follows the proportion of the share of brands in storage. However, the model is restrictive in incorporating heterogeneity and even with simplifying assumption is still computational burdensome.

Hendel and Nevo propose a demand model that can flexibly incorporate observable heterogeneity. The structure of their model allows for the separation of brand and quantity choices. This enables them to consistently estimate many parameters outside the consumer's dynamic problem. Therefore, this model is computationally simpler than the one mentioned previously. Henceforth, I use HN to indicate the authors Hendel and Nevo and HN (2006) to indicate their Econometrica paper.

HN's seminal paper is an important step forward. It paves way for future works on more elaborate models of dynamic demand. The model they consider is based on a conditional discrete choice framework where the idiosyncratic brand and size shock is assumed to be distributed according to type I extreme value distribution. As copious papers have pointed out previously, this Logit error has the independence of irrelevant alternatives property. It produces a priori unreasonable demand substitution patterns. The first goal of this paper is to incorporate systematic brand tastes into dynamic demand models. I allow households to differ in their preferences over characteristics of the products. More specifically, I incorporate a random coefficient on brand choices. I adopt HN's approach of separating brand and volume decisions but implement an estimator that allows for a random coefficient model in the first stage.

Along the way, I will also demonstrate in this paper that there are some drawbacks in HN's estimation procedure: The parameter estimates are highly sensitive to starting values of the optimization routine; the implied consumption decisions are strongly dependent on inventory levels; and the estimation is very time consuming. The second goal of this paper is to identify and explain these weaknesses and to improve HN's dynamic estimation procedure. Along the way, I also implement a more efficient structure of the code that is fast enough for simulated annealing and similar types of optimization methods. For a fair comparison, I use the same optimization routine as HN.

Since there are two separate but related tasks at hand, the body of this paper is divided into two sections. In the first section, the goal is to describe and discuss my random coefficient dynamic demand model of inventory. To that effect, this section is further divided into three subsections. In the first subsection, I motivate the model with past literature and examples seen in the data. In the second subsection, I formally introduce the model. And in the last subsection, I discuss the implications of the model and provide some thoughts on estimation. In the second section, I analyze HN's three stage estimator. I discuss its drawbacks and necessary adjustments for an improved estimation procedure. This section is also divided into subsections. In subsection one, I describe and discuss HN's model and their estimator used. In subsection two, numerical analyses of the estimation procedure are performed to show its weaknesses. Following that, in subsection three, I introduce and discuss in detail the improved computation method and show Monte Carlo results. In subsection four, I apply the improved estimator to household laundry detergent data provided by IRI. At the closing of this paper, in section three, I make some concluding thoughts.

Through out this paper, I focus on two industries: the soft drinks market and the laundry detergent market. I use the soft drinks market to motivate my model and estimation procedure but use laundry detergent to explain HN's model and estimation method. This is because HN use laundry detergent as their application and since fortunately I have a similar data set available I use it for a fair and informative comparison. As such, section one mostly uses examples and facts from Coca-Cola and Pepsi whereas section two uses data from the laundry detergent exclusively.

1.2 Section I

Prior to motivating the more elaborate model, it is important to understand Hendel and Nevo's seminal work. I briefly describe their model and the underlying idea behind their estimation procedure.

Consider HN's model: The market under study is laundry detergent, where there are over fifteen different brands of products provided in each market. Each of those brands sell three to six differently sized containers of detergent. That implies each household faces at least 45 brand-size combination choices in each period. Imagine keeping track of the prices and promotional activities for each household in each period. Furthermore, since prices and advertising activities transition over time, the states space is too large to compute.

Of course, in practice, depending how the researchers treat the choice set, the combination of brands and sizes may decrease significantly. In HN, the available brands and sizes are ones available in the store that the household is observed to have visited. Moreover, the brand choice consists of brands that each household is observed to have ever purchased in the panel. But even so, the states space is still too massive to carry around for the duration of the panel.

To overcome this hurdle, HN introduce a novel approach. They split brand and volume decisions into two parts. That is, households' brand utility is separate from their utility of consumption. Households experience the entire joy and happiness of the brand purchased at the point of purchase. Therefore, households' decision processes can be modeled sequentially. First, households choose their consumption and size purchase. Then conditioned on the already decided size, households choose their brands.

In the brand choice decision phase (stage one estimation), the estimation procedure is a static conditional multinomial discrete choice problem. The parameters associated with brand choice can be obtained through standard statistical packages like STATA, thus reducing the number of parameters to be estimated in the dynamic problem.

1.2.1 *Motivation*

Since the first stage estimation is a multinomial discrete choice model where the error term is distributed type I extreme value, past literature tells us the implied demand exhibits independence of irrelevant alternatives property. The error is an idiosyncratic shock. Households' brand choices are not dependent on the characteristics of brand. Given the prices and the advertising activities of the week, their decisions are driven purely by the random shock.

Perhaps one can argue that one bottle of laundry detergent is just like the next. After all, a household's got to do what a household's got to do: keep its clothing clean at the lowest costs. However, some moms would disagree: There have been arguments made for the enticing scent of a detergent that keeps its users happy. I suppose if on average a typical household does six hundred loads of laundry per year, then that fresh mountain breeze from Purex Ultra would make a difference. Others have insisted that Tide "simply cleans better". It is not surprising then companies spend a fortune on advertisements such as one that depicts women falling for guys with clean and shiny shirts. Scents and cheesy advertising campaigns aside, the data does provide evidence of brand preferences. In the IRI sample, over 46 percent of liquid detergent users purchased only one brand of laundry detergent over a five year period. Sixty percent of those single-brand users are, quoting Dr. Andrew Sweeting, Tide lovers.

There are even more compelling evidences that consumers have strong taste preferences for the market under study: the cola market. In this paper, I focus primarily on Coca–Cola and Pepsi-Cola. These two brands have been rivals ever since the nineteen hundreds when Pepsi first established its trademark. In the present day, Coca-Cola dominates Pepsi. According to Beverage Digest's 2008 report on Carbonated Soft Drinks, Pepsi's U.S. market share is 30.8 percent, lower than Coca-Cola

Company's 42.7 percent.

To clarify, while both companies provide several product lines, for example, Sprite vs. 7Up, I only consider the cola type drinks. They are closer substitutes of each other. In the IRI weekly scanner data for 2001, I find that over 56% of the households purchase one brand exclusively over that year. Below is a table presenting this fact in more detail:

	Frequency	Percentage
One Brand Only	2,561	56.60
Both Brands	1,964	43.40
Total	4,525	100.00

For the 2,561 households who are observed to have purchased only one brand, 90.41% of them purchase Coca-Cola while the rest purchased Pepsi. Another way of thinking about this brand taste persistence is that conditional on a previous purchase of Coca-Cola, there is an 89.84% chance that the next purchase is also Coca-Cola. Similarly, conditional on a previous purchase of Pepsi, the probability that the next purchase is also Pepsi is 61.48 percent. Below, I provide a table that presents this fact.

	Coca-Cola	Pepsi	Total
Remained Loyal	33,140	5,965	39,105
Switched Brands	3,747	3,737	7,484
Total	36,887	9,702	46,589

These are compelling evidence that consumers have systematic brand tastes. It is hard to claim, in the light of the above statistics, that consumers' brand purchase decisions are driven by prices, advertising, and random shocks alone. A Coca-Cola loyal household would not be likely to make a purchase of Pepsi even if it is on sale. Therefore, if individual differences in preferences for brands are not included, the demand substitution pattern would be significantly affected. In the next subsection, I formally state the model.

1.2.2 Model

The objective of each household is to maximize its expected infinite horizon sum of utilities. It achieves this goal by making three choices in each period. One, given the observed consumption shock, it chooses the optimal level of consumption. Two, conditional on the prices and promotional activities, it chooses the desired volume. A purchase of zero units is equivalent to making to purchases. Three, conditioned on that a purchase is made, it chooses the brand. Since soft drink is a storable good, any leftover is stored for potential next period consumption. Carbonated beverages are sold in standardized and pre-defined containers; therefore the volume purchased is a discrete choice. Let h denote the household, t denote the period, s denote the container size, and b denote the brand, households' problem can be broadly stated as

$$V\left(\phi_1\right) = \max_{\{c_h(\phi_t), b_h(\phi_t), s_h(\phi_t)\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[u_{htsb}|\phi_1\right]$$

where u_{hts} is each household's period t utility function that will be discussed in detail below. The discount factor is denoted by the standard Greek letter δ. The state at period t denoted by ϕ_t, includes beginning of the period inventory, current prices and advertising, shock to utility of consumption, and a vector of brand-size shocks. Variables $c_h\left(\phi_t\right)$, $b_h\left(\phi_t\right)$, and $s_h\left(\phi_t\right)$ stand respectively for consumption, brand choice and volume choice.

Each household's per-period utility can be represented as

$$u_{htsb} = \underbrace{U\left(c_{ht}, \upsilon_{ht}|\theta_h\right)}_{(1)} - \underbrace{C\left(i_{ht+1}|\theta_h\right)}_{(2)} + \underbrace{\sum_b \mathbb{1}_{hts} \cdot U_{htb}\left(p_{tbs}, a_{tbs}, x_{tbs}, \epsilon_{htbs}|\theta_h\right)}_{(3)}$$

$$: \quad \text{s.t. } i_{ht+1} = i_{ht} + s_{ht} - c_{ht} \tag{1.1}$$

$$: \quad \text{where } i_{ht} \geqslant 0,\ c_{ht} \geqslant 0,\ s_{ht} \geqslant 0$$

As seen above, the per-period utility can be decomposed into three components. The first component is the utility of consumption. Consumption decisions do not depend on current prices and advertising schemes. Furthermore, it is not influenced by the available brands in inventory. However, it is dependent on the consumption shock drawn. Following the literature, I assume the shock is additive in consumption. Hence, $U\left(c_{ht}, \upsilon_{ht}|\theta_h\right) = U\left(c_{ht} + \upsilon_{ht}|\theta_h\right)$. The second component is the cost of inventory. The inventory at the end of each period is equal to the previous stockpile plus the volume purchased this period minus the amount consumed. At the beginning of the next perid, this inventory becomes the beginning-of-period stockpile. The last component is the utility associated of making a brand-size choice. This utility depends on current prices, p_{tbs}, promotional activities, a_{tbs}, and observed product characterists, x_{tbs}. Furthermore, it includes observed consumer characteristics and unobserved consumer taste parameters. Since no purchase is denoted by a purchase of size zero, one and only one combination of brand and size is always chosen in each period. I use the indicator function $\mathbb{1}_{hts}$ to denote the brand-size choice household h in period t makes. It is equal to one if brand-size combination b and s is chosen and it is equal to zero otherwise. Therefore, in each period the sum of all the indicator functions of a household sum to one. The indirect latent utility function that follows the indicator function is specified below:

$$U_{htbs}\left(p_{tbs}, a_{tbs}, x_{tbs}, \epsilon_{htbs}|\theta_h\right) = \alpha p_{tbs} + \beta a_{tbs} + \gamma_h x_{tbs} + \epsilon_{htbs} \tag{1.2}$$

where α is the marginal utility of income; βa_{tbs} captures the effect of promotional activities on households' choices; and the random coefficients γ_h are unknown consumer taste parameters for the different product characteristics. In the context of carbonated beverages, product characteristics include only the brand information. The brand taste parameter varies across consumers according to

$$\gamma_h = \gamma + \Gamma D_h + \Upsilon \varsigma_h \tag{1.3}$$

where parameter γ denotes the mean of the random coefficients. For each household, observed demographical information is included in D_h and the unobserved characteristics are contained in ς_h. The vector of non-linear demand parameters Γ captures the observed heterogeneity, while Υ captures the unobservable heterogeneity due to random shocks ς_h.

In each period, the timing of events adheres to the following order: At the beginning of each period, consumption and brand-size shocks are realized. Each household then receives pricing and advertising schedules for every brand and each size of cola-related soft drinks. Based on the current period prices, the household forms an opinion of the expected future prices. Then the household makes a brand choice. Since inventory level is perfectly observed by the household, it chooses the level of consumption and the size of purchase simultaneously to maximize the expected sum of their infinite horizon utilities. No purchase is counted as a purchase of zero-liter bottle. Once a brand is purchase, the utility associated with it is exhausted. No brand utility is experienced at the point of consumption. At the end of the period, households consume the chosen consumption level and pay the cost for the leftover inventory.

Following HN, I make some standard assumptions on the distribution of shocks and on the path of prices.

Assumption 1. *Consumption shocks, υ_{ht}, is independently distributed across house-*

holds and over time.

Assumption 2. *Idiosyncratic brand-size shocks, ϵ_{htbs}, follow type I extreme value distribution. It is independently and identically distributed across households, period, and each brand-size combination.*

Assumption 3. *Unobserved random consumer characteristics, ς_h, are normally distributed.*

Assumption 4. *For simplification, prices and promotional activities follow an exogenous first-order Markov process.*

In the next subsection, I discuss the implications of this richer model along with additional computation complexity posed by it.

1.2.3 Implications of Model

Adding random coefficients allow households to have differences in their preferences for product characteristics. In line with all BLP type models, demand substitution patterns generated are more reasonable and realistic. However, incorporating the random coefficients on brand choices poses some computational complexities. This is due to the separation of the brand and volume choices. To allow the separation, the first stage estimation has to condition on the size choice. If households are allowed to have systematic brand tastes, then to compute the choice probabilities, one would have to integrate over the distribution of the different types. However, to compute the brand choice probabilities conditional on the volume purchase, one would have to integrate out the distribution of types conditional on the size purchased. To solve this problem, I estimate the model incorporating the three stage estimator with Arcidiacono and Miller (2007). More thoughts on this will be followed up.

1.3 Section II

In this section, I discuss in detail HN's model and difficulties that implementer of HN's estimation procedure may encounter. To be more specific, implementers are likely to experience the following: The likelihood function is non-smooth resulting in parameter estimates which are largely determined by the starting guesses. The parametric value function often suffers convergence problems. And the implied consumption is overly dependent on available inventory. To avoid these drawbacks, I provide an improved computational formulation. I explain the effects of each modification and present the results of the new procedure using both Monte Carlo simulations and an application to a set of data similar to those of HN's.

1.3.1 Model & Computational Structure of HN

Broadly speaking, there are two difficulties associated with estimating a dynamic demand model of storable goods. The first difficulty resides in our current computational limitation. Since most products have multiple characteristics and many of those characteristics change over time, the states space of the dynamic programming problem is too large to compute or carry around. The second difficulty comes from limited information on households' behavior. Not only are consumption decisions unobservable to most econometricians, data sets also don't include observations of consumers' inventory of the stored goods very often. Hendel and Nevo (2006) builds a model that overcomes both problems and applies the model to the household laundry detergent market using weekly scanner data.

To ease the computation burden, HN estimates many of the parameters outside of the dynamic programming problem. With their chosen stochastic structure of the model, brand choice decisions are seprable from size choice decisions. All brand choice parameters of the model are estimated ahead of the time and used to compute inclusive values. These inclusive values includes prices, promotional activities, and household specific characteristics. To solve the data limitation, HN relies on consumer utility maximization behavior to calculate the sequence of end-of-period inventory levels implied by the model for a given initial inventory. In the following subsections, I'll reiterate their model and estimation strategy prior to discussing the afore stated shortcomings of the estimation strategy[1].

Demand Model

In HN's model, households' utility can be broken down into three components: utility of consumption, utility of purchase, and cost of storage. The utility of purchase can be farther decomposed into the utility of brands purchased and the utility of sizes chosen. In each period, each and every household faces two sources of uncertainty. One is a consumption shock and the other is the set of future prices for each brand size combination.

The timing of events goes as follows. At the beginning of each period, households receive pricing and advertising schedules for every brand and each size of the product in interest. After drawing their idiosyncratic shocks, they make a brand choice. Choices are determined by prices, advertising activities, and idiosyncratic shocks. Since households have perfect information on the current level of inventory, they maximize the sum of their infinite horizon utilities by choosing the level of consumption and the size of purchase simultaneously. No purchase is counted as a purchase of size 0. Once a brand is purchase, the utility associated with that brand exhausted. No brand utility is experienced at the point of consumption. At the end of the period, households consume the utility maximizing consumption level and pay the price for the leftover inventory.

[1] For a more detailed discussion of HN's model and estimation procedure, please refer to their 2006 Econometric paper.

To avoid confusion, I adhere to HN's notations closely. Let h denote the household, t denote the period, j denote the brand, and x denote the size of purchase. Household h's per period utility is given as

$$\begin{aligned} & u_{hjxt} \\ = \; & \underbrace{\gamma \log\left(c_{ht} + \upsilon_{ht}\right)}_{\text{consumption utility}} - \underbrace{\left(\beta_1 i_{ht+1} + \beta_2 i_{ht+1}^2\right)}_{\text{cost of inventory}} + \underbrace{\sum_j d_{hjxt}\left(\alpha p_{jxt} + \beta a_{jxt} + \xi_{hjx} + \epsilon_{hjxt}\right)}_{\text{utility of purchase}} \\ : \; & \text{s.t. } i_{ht+1} = i_{ht} + x_{ht} - c_{ht} \qquad (1.4) \\ : \; & \text{where } i_{ht} \geqslant 0,\, c_{ht} \geqslant 0,\, x_{ht} \geqslant 0,\, \sum_j d_{hjxt} = 1 \end{aligned}$$

Here $d_{hjxt}\left(\alpha p_{jxt} + \beta a_{jxt} + \xi_{hjx} + \epsilon_{hjt}\right)$ is an indicator of the chosen brand size combination. It is equalled to one if brand j and size x is chosen and it is zero otherwise. Hence, the sum across all brands is one. Households' objective is to maximize the sum of all expected future utility with respect to consumption and size choice. That is,

$$V\left(s_1\right) = \max_{\left\{c_h(s_t), d_{hjx}(s_t)\right\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[u_{hjxt} | s_1\right] \qquad (1.5)$$

The discount factor is denoted by the standard Greek letter δ. At time t, the state s_t, consists of beginning of the period inventory i_{ht}, current prices p_{jxt}, current advertising a_{jxt}, shock to utility from consumption υ_{ht}, and the vector of brand-size taste shocks ϵ_{hjxt}. Applying standard empirical assumptions: 1) consumption shock is independently distributed over individuals and over time; 2) brand size shock is independently and identically distributed type I extreme value, this model can be estimated in three stages.

Estimation Strategy

The structural parameters of the model are estimated by maximizing the likelihood of the observed sequences of purchases. Similar to most dynamic estimation, the objective function depends on the solutions of the dynamic programming problem. Building on Rust (1987), HN nests households' dynamic programming problem within the parameter search of the estimation. Therefore a numerical solution of the dynamic problem is computed for each likelihood evaluation. Since neither consumption nor inventory is observed, the problem is more complicated than the standard dynamic demand estimation. For readers who are interested in HN's dynamic code structure, a guideline is attached in the appendix.

Assume for the moment that an initial inventory measure is available to the econometrician for each household. By consumers' utility maximization behavior, the optimal consumption and optimal purchase of each household in each period can be calculated for the given model. Therefore, over the duration of the panel, the econometrician can recover the sequences of households' per-period inventory implied by the model. To estimate the initial inventory, HN chooses an arbitrary set of inventory levels for each household and uses the first eleven periods of observation to generate the distribution of inventory.

Recall, the other hurdle of estimating this model is the large dimension of the states space: There are over fifteen brands of laundry detergents in the market. Each brand usually sells three to six different sizes making carrying the states space computational infeasible. HN solve this problem by estimating parameters associated with pricing, advertising, and household characteristics ahead of the dynamic programming problem. By the chosen stochastic nature of the model, these parameters are part of the static brand choice that are used in the dynamic part of the model.

The separation between brand and size is possible because the brand-size error term, ϵ_{hjxt}, is independently distributed over brands and sizes. That is, each brand-size combination has its own shock. Therefore, for each household in each period, conditional on the size purchased, the brand decision is a static multinomial discrete choice with a iid type-I extreme value error. Stage one of the estimation recovers the parameters of marginal utility of income, effects of advertising, and household-brand fixed effects. This is achieved through maximizing the likelihood of the observed brand purchases conditional on observed sizes purchased. Formally, the probability of choosing a brand conditional on observed size chose is:

$$\Pr\left(d_{jt} = 1 | x_{ht}, p_t\right) = \frac{\exp\left(\alpha p_{jxt} + \beta a_{jxt} + \xi_{hjx}\right)}{\sum_k \exp\left(\alpha p_{kxt} + \beta a_{kxt} + \xi_{hkx}\right)} \tag{1.6}$$

The likelihood function is the product of $\Pr\left(d_{jt} = 1 | x_{ht}, p_t\right)$ aross every household in every period conditional on their per-period purchased sizes. Maximizing the likelihood produces consistent static parameters. Theoretically, the choice set of each household in each period should include all brands available for the size chosen. In practice, however, HN's choice set is limited to the set of brands ever observed to have been purchased by the household in the duration of the panel. This reduces computation time as the average set of brands ever purchased is around four.

Once the parameters of the static brand choice model are computed, the inclusive value is simply the utility asociated with purchasing a given size. That is,

$$\omega_{htx} = \log\left\{\sum_k \exp\left(\alpha p_{jxt} + \beta a_{jxt} + \xi_{hjx}\right)\right\} \tag{1.7}$$

A set of ω_{htx}'s are calculated for each household in each period. One omega for each size available to the household in that period, regardless of whether any purchase is made. Furthermore, since the price process is assumed to follow a first order Markov process, a set of AR1 parameters need to be computed. This step follows standard procedures.

The last stage estimates the rest of the parameters (dynamic parameters). It is performed for each of the six pre-specified household types. A household's type is determined by the its location of residence: urban vs. suburban and its size: 1 to 2, 3 to 4, or more than 4. Substituting the set of inclusive values, ω_{ht}, in the households' per-period utility, the objective function can be re-written as follows. Here I omit the household subscript, h.

$$\max_{\{c_t,x_t\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[u_{jxt}|\omega_t, \upsilon_t, \epsilon_{xt}, i_t\right] \tag{1.8}$$

$$= \max_{\{c_t,x_t\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[\gamma \log\left(c_t + \upsilon_t\right) - \left(\beta_1 i_{t+1} + \beta_2 i_{t+1}^2\right) + \omega_{tx} + \epsilon_{xt}|\omega_t, \upsilon_t, \epsilon_{xt}, i_{ht}\right]$$

The Bellman equation associated with this simplified problem is

$$\begin{aligned} & V\left(\omega_t, \upsilon_t, \epsilon_{xt}, i_t\right) \\ = & \max_{\{c,x\}} \Big\{\gamma \log\left(c_t + \upsilon_t\right) - \left(\beta_1 i_{t+1} + \beta_2 i_{t+1}^2\right) + \omega_{tx} + \epsilon_{xt} \\ & + \delta \underbrace{E\left[V\left(\omega_{t+1}, \upsilon_{t+1}, \epsilon_{xt+1}, i_{t+1}|\omega_t, \upsilon_t, \epsilon_{xt}, i_t, c, x\right)\right]}_{EV}\Big\} \end{aligned} \tag{1.9}$$

Even with the reduction of states spaces, the value function is still computationally burdensome. Therefore, EV is estimated by value function approximation with policy iteration detailed in Benitez-Silva et al (2000). HN's parametric value function can be decomposed into the sum of two polynomials. The first polynomial is a function of end-of-the-period inventory and the second is a function of expected next period omega's and lags of omega's. Let $G(\cdot)$ and $H(\cdot)$ denote different polynomials, then $EV = G\left(i_{ht+1}\right) + H\left(E\left[\omega_{xt}\right]\right)$. To be more specific, the actual functional form chosen by HN is

$$EV = \widehat{\theta}_1 \ln(i_{t+1}) + \widehat{\theta}_2 \ln(i_{t+1})^2 + \widehat{\theta}_3 \ln(i_{t+1})^3 + \sum_{x+4}\left(\widehat{\theta}_{x+4} E\left[\omega_{xt}\right]\right)$$

Two things should be noted of the above chosen functional forms. One, consumption level directly depends on inventory. Two, due to additive separability, the term $\sum_{x+4}\left(\widehat{\theta}_{x+4}E\left[\omega_{xt}\right]\right)$ remains constant across all sizes and all consumption levels. Therefore, the expected future price does not influence consumption decisions. Since households' current period consumption does not depend on current inclusive values, consumption does not depend on prices directly at all. It only depends on prices indirectly through purchases and hence inventory levels. In the next sections, I present the weaknesses of this estimation procedure.

1.3.2 Numerical Analysis of HN's Estimation

While replicating HN's 2006 paper with IRI data, I experienced difficulties obtaining dynamic parameters. I substituted IRI data into the set of codes posted on Econometrica's supplemental material section. Minimal changes were made to the code. The final parameters were almost identical to the initial guesses HN used. While HN and I both use the laundry detergent market data, there are fairly significant differences between them. Firstly, the panelists of the two sets of data are located in geographically dissimilar regions: HN use a data set where households are located in a major metropolitan area whereas IRI collects its panel data from two suburban cities. Secondly, the IRI data is gathered at least eight years post HN's data. During those years several upstream competitors entered the market and some producers increased the number of available sizes; as a result, there are more sizes and brand choices available in the IRI data. Hence, there is little reason to believe the utility of consumption and the cost to inventory of the two should be near identical. I perform a series of tests on HN's code, using the data they provide on Econometrica, to study the relationship between initial guesses and final estimates in stage three.

Stages one and two follow standard practices closely. By maximum likelihood properties, parameters to static brand choice are consistent, so I provide no addi-

Table 1.1: Dynamic parameter estimates obtained using codes and data posted on Econometrica Supplement Material to HN 2006. The parameters are: γ is the marginal utility of consumption, β_1 is the linear cost of inventory, and β_2 is the quadratic cost of inventory.

HH Type		γ	β_1	β_2	LLF Value	Run Time
1	Initial Guess	1.3046	9.2332	-3.8442	365.6057	40.05
	Result	1.3056	9.2410	-3.8193		
2	Initial Guess	0.7470	6.4839	1.7746	926.7561	19.45
	Result	0.7473	6.4848	1.8014		
3	Initial Guess	0.5073	21.9555	-35.8595	1.53E+03	81.72
	Result	0.5073	21.9560	-35.8608		
4	Initial Guess	0.0807	4.2231	-8.1845	1.04E+03	38.50
	Result	0.0807	4.2235	-8.1854		
5	Initial Guess	0.9162	4.1324	-6.1341	543.6425	35.78
	Result	0.9154	4.1259	-6.1349		
6	Initial Guess	3.6462	11.5057	-1.1789	1.09E+03	197.50
	Result	3.8003	11.7521	-0.7274		

tional discussion here. Stage three, however, produces dynamic parameter estimates that overly depend on the initial parameter guesses. Please note that the parameter estimates reported in HN (2006) achieve the maximum log likelihood for the given convergence criteria and the number of consumption shock draws. However, to be able to pin down those parameters and their starting guesses, a grid search method would need to be employed. Performing a grid search on a multi-dimensional parameter space with nested dynamic programming problem is very time consuming if not infeasible. Using the set of codes and data posted on Econometrica's supplemental material section, I report HN's dynamic parameter estimates in table 1 below for completeness.

Since econometricians don't usually have foreknowledge of the true parameters of the model, starting values of the likelihood evaluations are usually far off the final estimates. To test the dependency of the dynamic parameters on the starting values, I perform three set of tests: In the first set, I make no changes to the code except altering the starting guesses; in the second set, I modify the convergence criteria;

and in last set, I increase the consumption shock draws.

First, I experiment with alternative starting guesses. I test several starting values, ranging from 5% up and down the reported value to 50%. The conclusion is the same, the resulting parameters of the experiments are far from the ones reported in HN (2006). As an example, Table 2 below shows the results of the dynamic parameter estimates if I increase and decrease the starting values by 25%.

Table 1.2: Final parameter estimates using initial guess 25% above and below the estimates given in HN 2006. No other changes were made in the code posted on Econometrica.

HH	Par	Reported	25% Below Reported Value		25% Above Reported Value	
			Initial Guess	Final Estimate	Initial Guess	Final Estimate
1	γ	1.3056	0.9785	0.7444	1.6308	2.1144
	β_1	9.2410	6.9249	4.7250	11.5415	12.5210
	β_2	-3.8193	-2.8832	-1.6875	-4.8053	-5.1533
2	γ	0.7473	0.5603	0.4367	0.9338	0.5570
	β_1	6.4848	4.8629	1.8877	8.1049	10.6092
	β_2	1.8014	1.3310	1.3595	2.2183	3.0355
3	γ	0.5073	0.3805	0.3311	0.6341	1.2944
	β_1	21.9560	16.4666	15.7983	27.4444	25.3154
	β_2	-35.8608	-26.8946	-26.8834	-44.8244	-39.8502
4	γ	0.0807	0.0605	0.0419	0.1009	0.1338
	β_1	4.2235	3.1673	3.3443	5.2789	6.8162
	β_2	-8.1854	-6.1384	-1.6791	-10.2306	-11.2379
5	γ	0.9154	0.6872	0.5002	1.1453	1.3706
	β_1	4.1259	3.0993	2.7438	5.1655	5.8423
	β_2	-6.1349	-4.6006	-4.2487	-7.6676	-8.4105
6	γ	3.8003	2.7347	3.9116	4.5578	5.3692
	β_1	11.7521	8.6293	15.5033	14.3821	12.9772
	β_2	-0.7274	-0.8842	2.5480	-1.4736	-1.8681

These results indicate that the likelihood function is non-smooth. Simplex search methods, e.g. fminsearch, is not finding the global minima. One remedy to this is simply decreasing the convergence criteria. There are two convergence criteria in stage 3 estimation. One governs the convergence of the dynamic programming problem and the other on likelihood function evaluation. The convergence criteria on the dynamic programming problem is set to vary, but its minimum is achieved at 1e-3. The convergence criteria on fminsearch is also set to 1e-3. In the second experiment, I decrease both to 1e-5 and run the previous experiments with alternative starting values. As table three shows, this does not solve the dependency issue. To reduce redundancy, instead of listing all the results, I list only parameter estimates of a 10% decrease on the starting guesses of household type 3. This household type has the largest number of households present (49 households).

Table 1.3: Decreasing the starting guesses by 10%. The above lists final parameter estimates of household type three after tightening the convergence criteria to 1e-5.

HH	Par	Truth	10% Below Truth: Initial Guess	10% Below Truth: Final Estimate
	γ	0.5073	0.4566	0.4513
3	β_1	21.9560	19.7600	20.8025
	β_2	-35.8608	-32.2736	-33.5005

Although the parameter estimates of the cost of inventory shows some improvement, the marginal utility of consumption remains close to the initial guess. The likelihood function could also be non-smooth because the number of consumption shock draws is small. For each household in each period, HN draws 100 consumption shocks. In experiment three, I increase the number of consumption shock draws to 10,000. I once again use household type 3 as an example to show that this fix does not achieve the desired result.

Table 1.4: Decreasing the starting guesses by 10%. The above lists final parameter estimates of household type three after increasing the consumption shock draws to 10,000.

HH	Par	Reported	10% Below Reported Values: Initial Guess	10% Below Reported Values: Final Estimate
	γ	0.5073	0.4566	0.4354
3	β_1	21.9560	19.7600	19.8022
	β_2	-35.8608	-32.2736	-31.6783

If anything it seems increasing the consumption shock draws actually make the estimates closer to the starting values. To understand this problematic phenomenon,

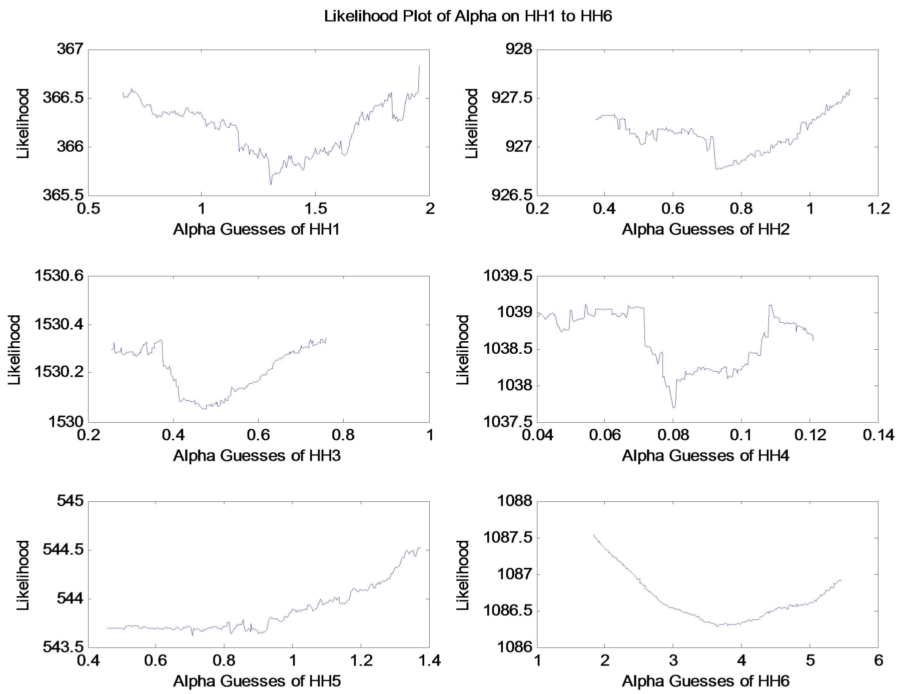

FIGURE 1.1: Likelihood Plot of α

I plot the negative of log-likelihood functions for each parameter while holding all others constant at the reported values. The first set of graphs below plot the likelihood function of the marginal utility of consumption for each household type while holding the cost to inventory terms constant.

As is apparent from the set of plots above, not only are most of the curves highly non-smooth, it is even unclear if household type five's log likelihood has a well defined minimum. Only household type six has a smooth likelihood curve with a clear global minimum. However, as we have seen from table two, even this household type has parameter estimates that depend on the starting guesses. Part of this could be explained by the likelihood function plots of the cost of inventories. The two cost of inventory parameters have very similar likelihood plots, here I present the less smooth of the two.

From the set of graphs above, it is not surprising that estimates may be off the

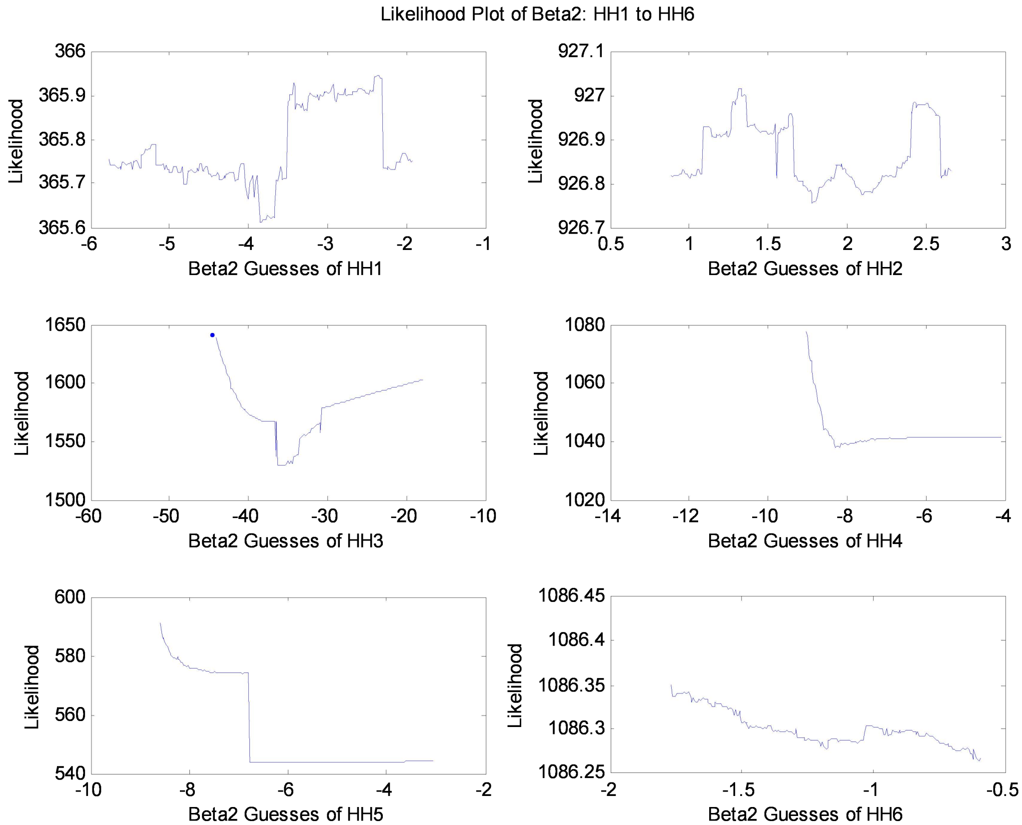

FIGURE 1.2: Likelihood Plot of β_2

mark. While household types one, two, and three have very rugged likelihood functions, the likelihoods feature well defined global minima. The same cannot be said for household types four and five. Types four and five have likelihood functions that are almost flat for a range of parameter values. Household type six's likelihood function is almost monotonically decreasing, so it is unclear whether parameter estimate achieves a global minimum.

If the likelihood is flat, then smoothing the function would not improve the parameter estimates. However, since some of household types have likelihood functions with well defined minima, smoothing them should provide better estimates. Increasing the consumption shock draws of these households, the two plots below shows the

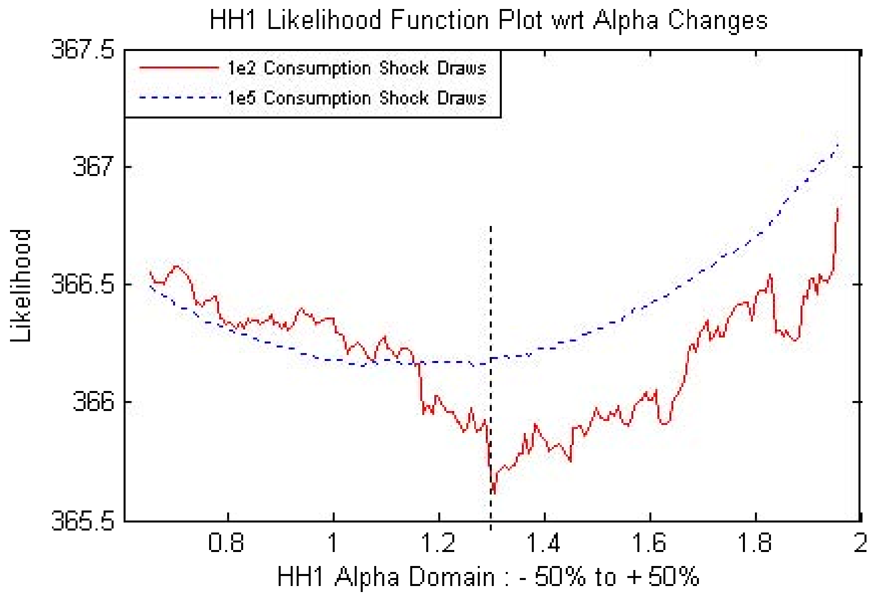

FIGURE 1.3: Likelihood Plot w.r.t. α-changes

likelihood functions before and after increasing consumption shock draws. The first is a graph of type one households' likelihood function with respect to alpha holding betas constant at the reported values in HN 2006. The second one is for beta 2 for the same household. The plots show likelihood functions before and after increasing the consumption shock draws. The solid red line indicate 100 consumption shocks and the dotted blue line 10,000 consumption shocks. Increasing consumption shock draws significantly smoothed the likelihood function. Unfortunately, it does not improve the estimates. For both of the plots above, the likelihoods after increasing consumption shocks no longer achieve a global minima as before. In fact the likelihood function of the quadratic inventory cost term became almost monotonically deceasing. It is not surprising then increasing consumption shock draws does not lessen parameter estimates' dependencies on starting guesses.

I hope that it is self-evident at this point that simple modifications of HN's stage three estimation procedure do not solve the dependency issue. In the next section, I provide some explanation for this weakness and discuss the necessary adjustments

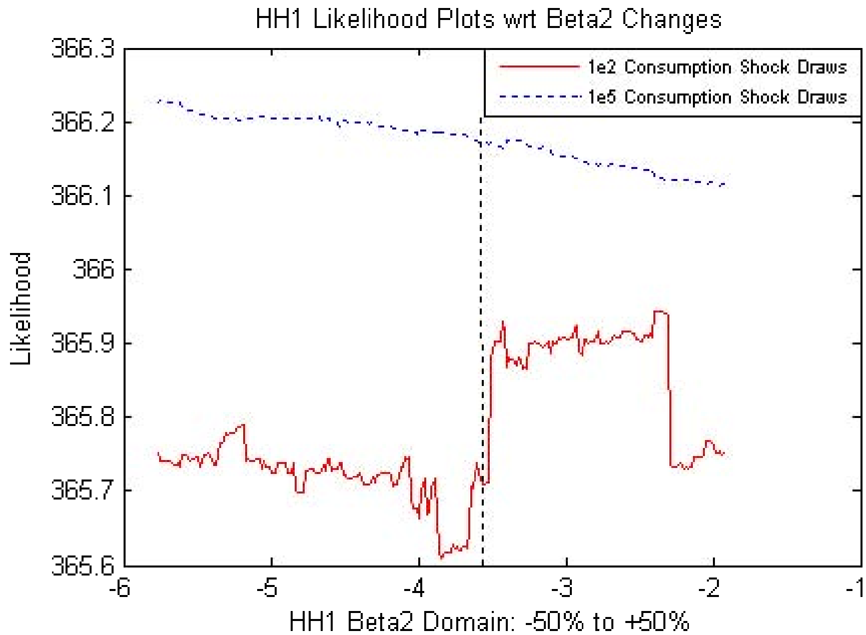

FIGURE 1.4: Likelihood Plot w.r.t. β_2-changes

to obtain consistent estimates.

1.3.3 *Improved Computation Method*

Recall, the third stage of Hendel and Nevo's estimation routine is maximum likelihood with the numerical solution of the dynamic programming problem nested in. The value function is estimated via a parametric approximation approach. Theoretically, once the numerical solution to the parametric value function is calculated, the relative simplicity of the model allows for a straight forward maximization of the likelihood function. However, in practice, this estimation routine is difficult to carry out due to 1) computation burden and 2) unobserved initial inventory. The dependency of the parameter estimates on starting guesses is a result of three factors: inaccurate approximations of the parametric value function, current period utility choices, and the use of random initial inventory levels.

Adjustments are made on all fronts. I improve the approximation to the parametric value function and modify the current period utility function by choosing basis functions that better fit the industry under study. I also interact consumption decision with expected future prices. Moreover, instead of setting arbitrary initial inventories, I estimate the initial inventory levels by combining observed households' purchase behavior with the nature of detergent needs. I present a detailed discussion below.

Function Choice

The quality of value function approximation depends on the dimension of the states spaces and the number of basis functions used. Theoretically, with a large enough number of basis functions, any function can be well approximated. Moreover, with enough terms, the functional forms of the basis functions also don't matter: Polynomials or Fourier transformations both achieve good approximations. The matter becomes more complicated when the dimension of the function is large. As the number of dimensions increases (number of parameters increases), it becomes computational burdensome to provide enough basis functions for a good approximation. When this becomes an issue, as is the case here, the functional forms of the basis functions play an important role for good approximations.

HN bases the parametric value function on twenty-one variables. The basis functions include a constant term, three polynomial terms of the log of current inventory, and expected future period omegas and their lags. To fit a close approximation for this value function, a large number of basis functions has to be used. However, as the number of basis function increases, so does the computation burden. Hendel and Nevo choose a polynomial consisting of twenty-four basis functions for value function approximation.

Given Hendel and Nevo's chosen parametric value function, the Bellman equation can be expressed explicitly as

$$
\begin{aligned}
& V\left(i_t, \omega_t, \nu_t, \varepsilon_t\right) \\
= \quad & \max_c \left\{u\left(c+\upsilon_t\right) - C\left(i_{t+1}\right) + u\left(sz_t\right) + \varepsilon_t \right. \\
& \left. +\delta E\left[V\left(i_{t+1}, \omega_{t+1}, \nu_{t+1}, \varepsilon_{t+1}\right) | i_t, \omega_t, \nu_t, \varepsilon_t, c, sz\right]\right\} \\
= \quad & \max_c \left\{\gamma \log\left(c_t + \upsilon_t\right) - \left(\beta_1 i_{t+1} + \beta_2 i_{t+1}^2\right) + u\left(sz_t\right) + \varepsilon_t + \right. \\
& \left. \delta\left[\widehat{\theta}_1 \ln(i_{t+1}) + \widehat{\theta}_2 \ln(i_{t+1})^2 + \widehat{\theta}_3 \ln(i_{t+1})^3 + \sum_{x=1}^{24}\left(\widehat{\theta}_{x+4} E\left[\omega_{xt}\right]\right)\right]\right\}
\end{aligned}
$$

Since this function is additively separable, given a purchase decision, only consumption and inventory terms are relevent to consumption decisions. The inclusive values and the expected future prices remain constant across all choices of consumption. Dropping inclusive values and future expected prices, the Bellman equation for objective function is:

$$
\begin{aligned}
& V\left(i_t, \omega_t, \nu_t, \varepsilon_t\right) \\
= \quad & \max_c \left\{\gamma \log\left(c_t + \upsilon_t\right) - \left(\beta_1 i_{t+1} + \beta_2 i_{t+1}^2\right) \right. \\
& \left. +\delta\left[\widehat{\theta}_1 \ln(i_{t+1}) + \widehat{\theta}_2 \ln(i_{t+1})^2 + \widehat{\theta}_3 \ln(i_{t+1})^3\right]\right\}
\end{aligned}
$$

The functional choices along with reported parameter estimates imply the following utility curves with repect to consumption:

Several implications of the Bellman equation functional forms are clear from the above utility plot. First of all, optimal consumption decision is highly correlated with the available level of current inventory. The consumption level that maximizes the infinite horizon utility strictly increases in initial inventory level. Secondly, the optimal consumption level is close to the total inventory., suggesting that each household consumes most of their inventory in each period. Moreover, the value of the Bellman equation at optimal consumption is strictly increasing in inventory. Together, these

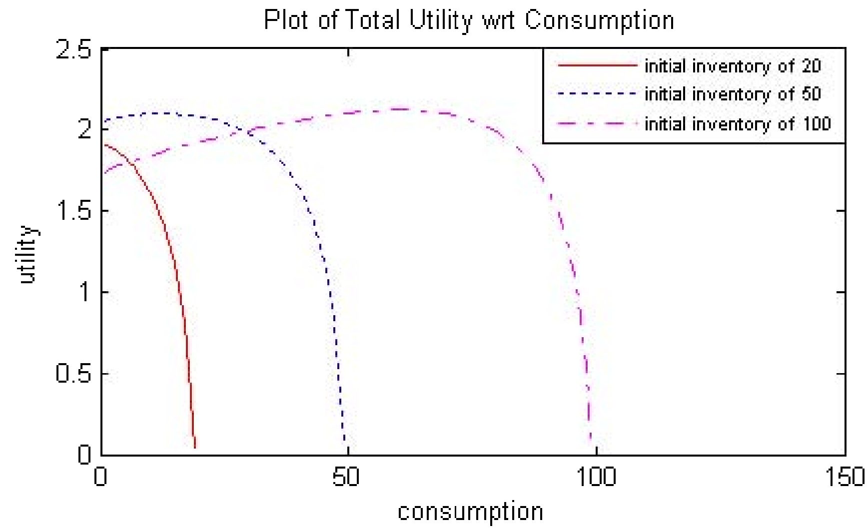

FIGURE 1.5: Total Utility w.r.t. to Consumption

properties imply that households will always chosen to purchase the largest quantity available, consume most of the stock, and leave a small beginning of next period inventory. Without size shocks, this generates purchases of large quantities in each period for each household. Since the period is defined at the week level, this data generating process would not generate data that matches what is observed. Any variation in quantity purchased is driven by the random size shocks.

Furthermore, since consumption is very dependent on inventory levels, correctly estimating the initial inventory is important. An error created here would stretch to inaccurate optimal consumptions and end of period inventories. This eventually implies purchase decisions which are not consistent with actual consumer behavior. Not surprisingly, we see the implied initial inventory levels generated by the model behave somewhat strangely. In the figure below I plot initial inventory and optimal consumption generated by HN's model using data provided on Econometrica's web site. I also plot the convergence of a randomly choosen parameter of the parametric value function.

The parameters of the parametric value function converges well. The convergence

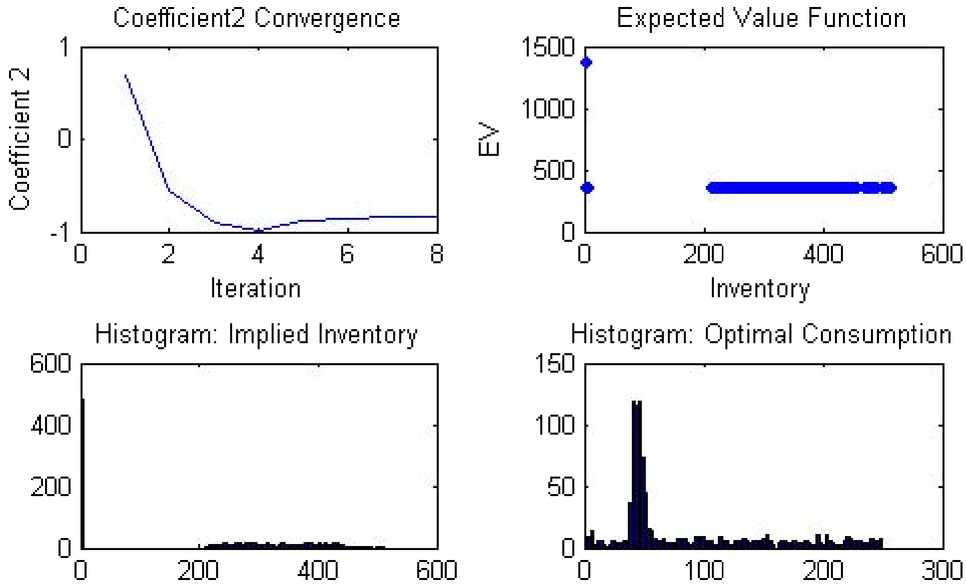

FIGURE 1.6: Properties of HN Utility Function

criterion is 1e-3, the parameters converge in eight iterations. The expected value function generated, however, seems to be independent of the inventory level. This implies that households have little need for detergent inventory. They derive almost the same utility from a wide range of inventory levels. While the distribution of implied optimal consumption has a irregular shape and a large variance, the mean of the distribution is not too far from the average household detergent use. The implied inventory, however, does not behave as well. From the histogram of the implied inventory level, we see that there is large gap between no inventory and two hundred units of inventory. It's difficult to understand its cause. It seems that households with less than two hundred units of inventory will chosen to consume all and leave no inventory for the next period. Another drawback of the Bellman equation is that, depending on the dynamic parameters the optimization routine searches over, the parametric value function does not converge. Parametric value function parameter would osolate between several values. The use of a small update parameter doesn't solve this problem. HN's remedy for this to to allow the convergence criterion to vary depending on the number of iterations.

Take a moment to consider the industry under study. Laundry detergent is one of the consumer product that is needed usually on a weekly basis. However, few households go shopping for detergent every week. While the weekly laundry needs depend on inventory availability, the variation of weekly detergent needs for each household should not be very large. In line with Hendel and Nevo, I also specify current period utility with three components and use parameteric value function approach to approximate the above problem. However, I chose current period utility functions and basis functions that reduce the dependency of consumption level on initial inventory level. Furthermore, I allow consumption to depend on expected future prices (inclusive values).

$$\begin{aligned} & V\left(i_t, \omega_t, \nu_t, \varepsilon_t\right) \\ = \quad & \max_c \left\{ \gamma \cdot \sqrt[5]{(c_t + \upsilon_t)} - \left(\beta_1 i_{t+1} + \beta_2 i_{t+1}^{1.5}\right) + \omega_t + \right. \\ & \left. \delta \left[\widehat{\theta}_1 \sqrt[2]{i_{t+1}} + \widehat{\theta} \sqrt[3]{2 i_{t+1}} + \widehat{\theta}_3 \sqrt[5]{i_{t+1}} + \sum_{x=1}^{4} \left(i_{t+1} - 1\right)_{t+1}^{-1} \left(\widehat{\theta}_{x+4} \cdot E\left[\omega_{xt}\right] \right) \right] \right\} \end{aligned}$$

Instead of using log of consumption, the utility of consumption follows a root-n form. This function better fits the utility of detergent consumption. While both functions are concave in consumption, the root-n form avoids negative or undefined utility values associated with log functions when consumption levels are smaller than one. Furthermore, the marginal utility of consumption decreases faster in the root-n form than that of log, so households' consumptions become less dependent on inventory. The basis functions of the inventory terms in the adjusted parametric value function follow a similar story. I exchange the log of the inventory with root-n forms. Furthermore, to allow households' consumption to depend on expected future prices, I interact end-of-period inventory with expected future prices.

As discussed in HN(2006), it's important to allow consumption to vary in response to price changes because this is the main alternative explanation to why consumers buy more during sales. However, since the expected future prices enter into the Bellman equation as additive separable terms, they do not affect consumption. I add interactions between inventories and prices. I argue that laundry detergent usage is largely driven by need. That is, households may not consume more when expected future prices are low. However, they would stockpile more laundry detergent if the expected next period prices are high. In another word, inventory depends not only on current period prices but also expected future prices. And since consumption is directly related to inventory, interactions between consumption and expected future prices come through interactions between inventory and expected future inclusive values. To better understand the implication of the functional choice, I plot below a similar graph of the Bellman equation:

As shown in the graph above, the inifinite horizon utility is still strickly increasing in initial inventory level. However, consumption is no longer overly dependent on available inventory. In fact, consumption levels remain close to the reasonable average household detergent use. This implies that while housedholds value inventory, they do not over-consume simply because they have large stocks of laundry detergent at home. This functional structure also improves the convergence of the parametric value function. The figure below shows the convergence of a randomly chosen coefficient of the parametric value function and the implied consumption and inventory levels.

The convergence of coefficients is smoother in comparison to HN's, but this is the result of using a small update parameter. From the above plots, we see that the expected value is increasing in inventory levels, which conforms to observed households' stockpiling behavior. Moreover, the implied inventory level match the

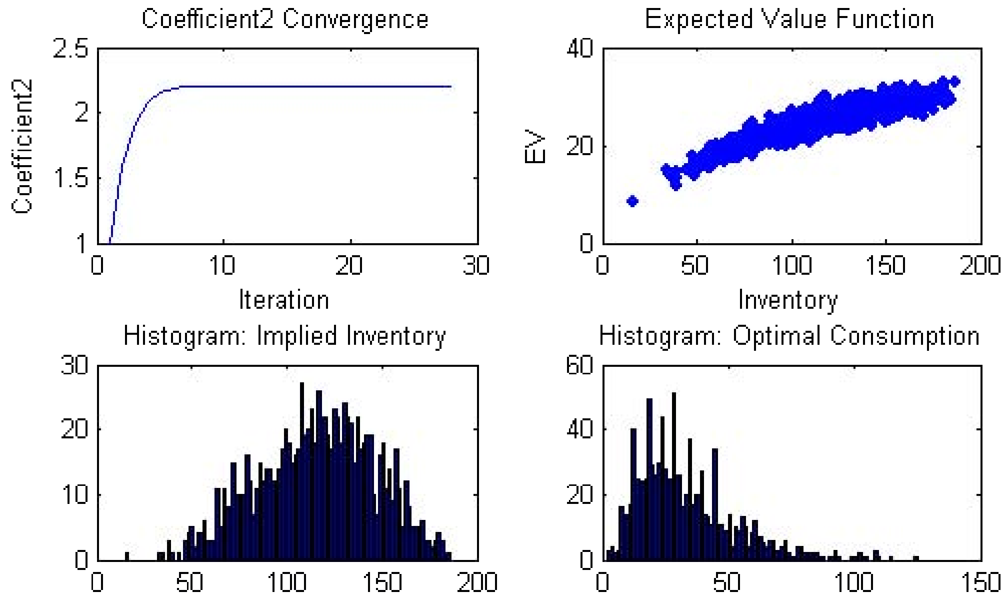

FIGURE 1.7: Properties of Corrected Utility Function

simulated inventory very well. The implied optimal consumtion also seems more reasonable.

Improving the functional structure of the utility does not remove the dependence of purchase decisions on initial inventory levels. If the initial inventory level is significantly different from the actual inventory, then the implied purchase decision would be different from the observed ones. In the next subsection, I discuss the method I use to estimate the initial inventory.

Initial Inventory Estimation

Following standard approach, Hendel and Nevo estimate each household's initial inventory by arbitrarily drawing a set of 100 values and use the first 11 periods of observations to generate the distribution of inventory. Since consumption is very dependent on the amount of stock available, this process doesn't lead to a distribution of the initial inventory that converges to the true level. Hence, the entire sequence of end of the period inventories are mismeasured and the parameter estimates implied are affected. I take a different approach. While I also draw a set of initial inventory levels, their values are not drawn arbitrarily. I combine initial inventory implied by the observed purchasing behavior with randomly drawn stockpiles.

Households' laundry detergent needs have little variance over the duration of a panel study, unless its family structure experiences major changes. Since changes in household structures are observed, this can be easily taken care of. Assume each household have a preferred level of stockpile, over the duration of the panel, the observed amount of detergent used between any two purchases implies an average weekly consumption need. Therefore, the amount of household detergent consumption prior to the first observed purchase can be approximated. I use this consumption level along with a set of randomly drawn stockpiles to generate the distribution of initial inventory. The ramifications of this approach are reported and discussed in the Monte Carlo section.

Monte Carlo Study

As is evident from the previous sections, HN (2006) proposed a dynamic demand model of storable goods that allow brand taste parameters to be estimated prior to the dynamics. Thus the estimation procedure is less computationally intensive. The main drawback of HN's estimator is that the dynamic parameter estimates are overly sensitive to starting guesses of the optimization routine. I propose three changes to their estimation routine to decrease that dependency. I test my proposals on data simulated to resemble the set used in HN's paper.

In order to generate data that is similar to the set used in HN's paper, I omit estimation stages one and two. Instead, I use the AR1 parameters reported in their paper to generate the sequences of inclusive values. I also use distributions given in their code to generate initial inventory levels and per period consumption shocks. Since it is time consuming to generate data according to HN's model. I only simulate ten sets of data, each containing 500 households with 52 periods. Using these sets of data, I perform rigorous tests on the proposed changes. Results are discussed below. After discussing the performance of the proposed adjustments, I show how each adjustment influence the estimation and its results.

Sensitivity of Parameter Estimates to Starting Values Naturally, I start the series of tests by varying the initial parameter guesses. The starting parameters range from 70% of the true parameters to 130%. All parameters are decreased or increased by the same percentage. I do not test how parameter estimates differ if only one parameter is different from its true value. Although, the results here should hold for those cases. The average of the results is presented in the table below:

True Parameter	γ	β_1	β_2	:	2	0.00150	0.00150

	Starting Value			Estimation Result		
	γ	β_1	β_2	$\widehat{\gamma}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
True Parameter	2	0.00150	0.00150	1.9609	0.00154	0.00155
5% Below Truth	1.9	0.00143	0.00143	1.9661	0.00140	0.00154
5% Above Truth	2.1	0.00158	0.00158	1.9674	0.00165	0.00158
10% Below Truth	1.8	0.00135	0.00135	1.9736	0.00134	0.00157
10% Above Truth	2.2	0.00165	0.00165	1.9727	0.00177	0.00157
15% Below Truth	1.7	0.00128	0.00128	1.9658	0.00129	0.00158
15% Above Truth	2.3	0.00173	0.00173	1.9572	0.00166	0.00159
20% Below Truth	1.6	0.00120	0.00120	1.9696	0.00120	0.00157
20% Above Truth	2.4	0.00180	0.00180	1.9632	0.00185	0.00154
25% Below Truth	1.5	0.00113	0.00113	1.9569	0.00117	0.00158
25% Above Truth	2.5	0.00188	0.00188	1.9684	0.00199	0.00147
30% Below Truth	1.4	0.00105	0.00105	1.9630	0.00110	0.00161
30% Above Truth	2.6	0.00195	0.00195	1.9666	0.00233	0.00151

To keep an impartial comparison, I use the same convergence criteria as HN, 1e-3. Based on the results presented above, the marginal utility of consumption, γ, is fairly consistent. The parameter estimates have little dependence on the starting guesses. And estimates deviate less than 2.5% from the truth. The second inventory cost term, β_2, follows a similar story. Its parameters estimates do not depend on the starting guesses of the optimization and maximum deviation from the truth is less than 7.4 percent. Unfortunately, the linear cost of inventory parameter, β_1, is more susceptible to starting values. It seems to be negatively correlated with its initial guess. In theory, this parameter is identified. However, due to its extremely low value, it cannot be very precisely estimated with a low tolerance. Overall, estimates of the dynamic parameters are close to their true values. In addition, the improved estimator takes, on average, less than fifteen seconds per iteration.

Recall, from the discussion of HN's Bellman equation structure, I show evidence that not only do consumption levels depend highly on inventory, the implied per period consumption is also very close to the available current period inventory. This feature implies that dynamic parameters are sensitive to initial inventory estimates. Although my proposed functional structure removes some of the correlation between consumption and inventory, initial inventory levels still influence parameter estimates. In the next subsection, I test the sensitivity of parameters on initial inventory estimates.

Sensitivity of Parameter Estimates to Initial Inventory Estimates To test how sensitive parameter estimates are to estimates of initial inventory, I let the initial inventory vary but hold constant the starting values of the optimization routine. The initial parameter guesses used are the true parameters. I test various initial inventory levels. They range from minus 30% of the true starting inventories to plus 30%. Results of the Monte Carlo study are presented below:

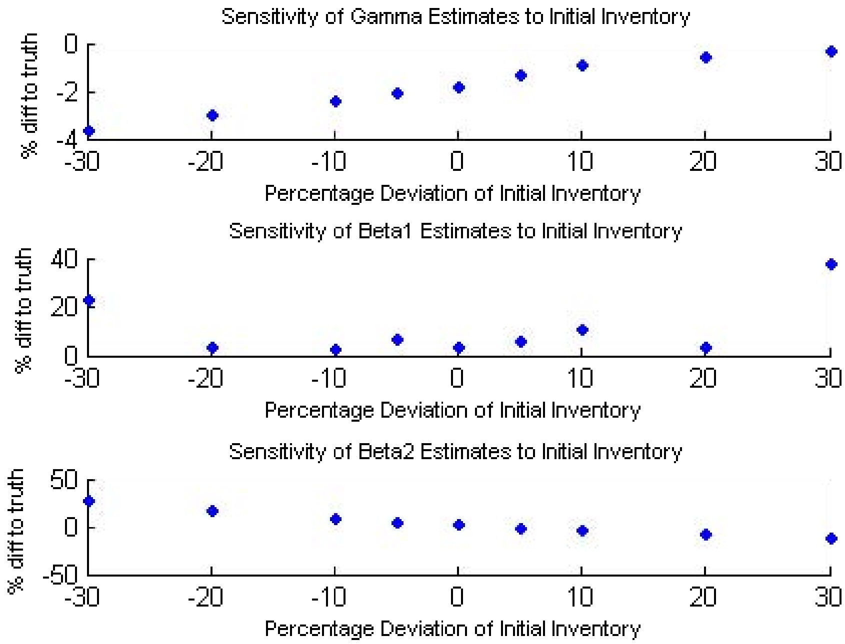

FIGURE 1.8: Sensitivity Plot

Initial Inventory	$\widehat{\gamma}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
True	1.96447	0.00155	0.00155
5% Below	1.96000	0.00160	0.00157
5% Above	1.97460	0.00159	0.00150
10% Below	1.95280	0.00154	0.00164
10% Above	1.98290	0.00166	0.00146
20% Below	1.94067	0.00156	0.00176
20% Above	1.98867	0.00156	0.00138
30% Below	1.92727	0.00185	0.00191
30% Above	1.99400	0.00207	0.00132

To ease discussion, for each parameter, I plot the percentage deviation of its estimates from the truth against the percentage deviation of the initial inventory level.

For the marginal utility of consumption, γ, the estimates are fairly invariant to initial inventory estimates. There is a slight correlation between the initial inventory level and the parameters estimates: Marginal utility of consumption estimate increases as the initial inventory estimate increases. As the graph below shows, if the initial inventory is uniformally overestimated by 30%, then the parameter estimate is 0.3% below the true parameter. On the other hand, if the initial inventory is uniformally underestimated by 30%, then the parameter estimate is 3.6% below the true parameter. The key result here is that the correlation is very small. Even if the initial inventory is underestimated by 30%, the parameter estimates will only be 3.6% away from the true marginal utility of consumption.

The first term of cost of inventory term, β_1, share a different story from marginal utility of consumption. Beta 1 is more sensitive to initial inventory levels. However, as the graph below shows, there is little trend between the level of the initial inventory and the linear cost of inventory. While the initial inventory deviates less than 20% away from the truth, the estimates of the first term of inventory cost stays invariant to the initial inventory guess. As the initial inventory deviates farther away from truth (30%), the parameter estimates worsens to 40 percent over the true parameter.

The parameter most influenced by initial inventory estimates is the second term of inventory cost, β_2. Estimates of the second parameter of inventory is correlated with initial inventory levels: As initial inventory estimates increase, the parameter estimates of β_2 decrease. If I uniformally increase the initial inventory by 30%, on average β_2 is estimated to be 12% below the true cost of inventory parameter. If I uniformally decrease the initial inventory by 30%, on average β_2 is estimated to be over 25% of the truth.

This result does not come as a surprise. It is sensible for this cost parameter to be more sensitive to initial inventory. If the initial inventory is estimated too high, then given the observed household purchasing behavior, the cost of inventory has to be low to justify a larger stock of detergent. Similarly, if the initial inventory is too low, but households are not observed to replenish their inventory, then the cost of inventory must be high.

The analysis suggests that consistent estimates of initial inventory is important. While small deviations from the true initial inventory level does not seem to influence parameter estimates too much, large deviations from the truth could cause cost of inventory parameters to significantly differ from their actualy values. In the following section, I show how the likelihood function reacts if I use only one or two of the three proposed changes.

Reaction of Estimates to Functional Form Changes Recall, three changes were made to HN's estimation routine: 1) I adjust the parametric value function. I adopt a set of basis functions that are based on a root-n form instead of log function. I also interact consumption with expected future prices. 2) I alter the current utility functions. The new utility of consumption is also based on a root-n form instead of the log of the consumption level. This function avoids negative utilities associated with low levels of consumption and allows the rate of decreasing marginal utility of consumption to be higher in comparison to its log counter part. Moreover, the nonlinear term of the cost of inventory is raised to the power of 1.5 instead of the squared term. This normalizes the utility of zero consumption to be zero. 3) The last change implemented uses observed data to help estimate the initial level of inventory. Prior to discussing how each of those changes affect the likelihood, for reference later on, I plot below the likelihood after all adjustments are made using simulated data. Since there are a total of three parameter, the likelihood lives in a four dimensional space. Since we humans have a hard time plotting and comprehending four-dimensional spaces, I keep the linear cost of inventory parameter constant at its truth and plot the likelihood altering the rest of the two parameters:

The surface of the likelihood function is fairly smooth and hence would lead to parameter estimates that are not dependent on initial guess. First, we would want to see the resulf of running HN's set of code on the simulated data. Unfortunately,

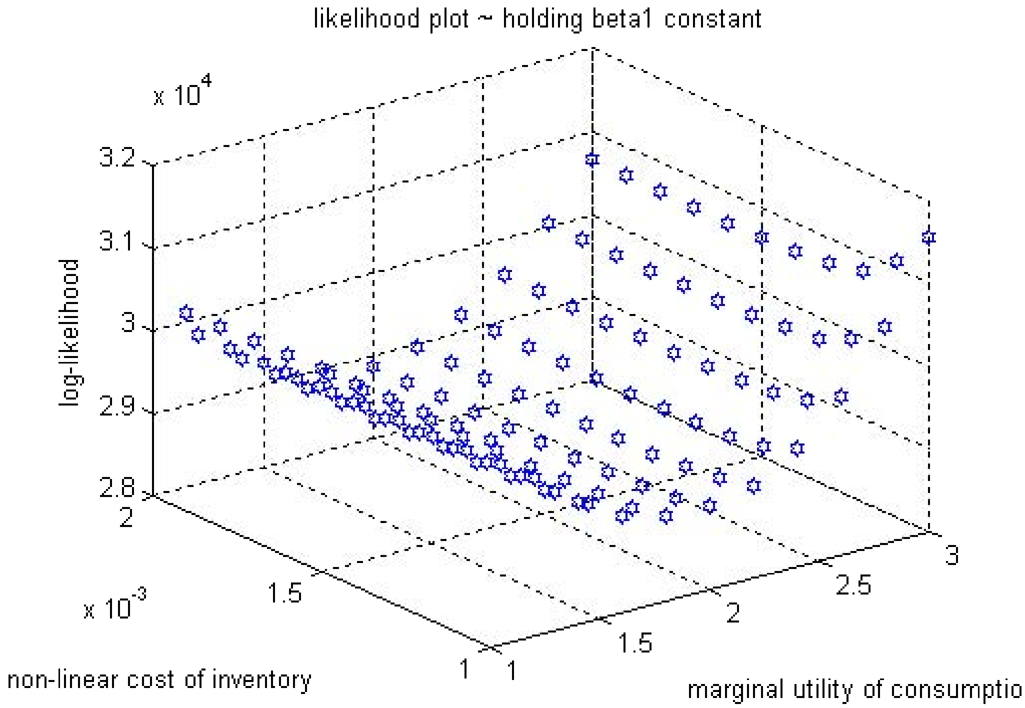

Figure 1.9: 3D Likelihood Plot - β_1 Constant

the parametric value function refuses to converge. The parameter value occelates

The surface of the likelihood function is fairly smooth and hence would lead to parameter estimates that are not dependent on initial guess. First, we would want to see the result of running HN's set of code on the simulated data. Unfortunately, the parametric value function refuses to converge. The parameter osculates between several values even with a very small update parameter (0.01). In fact, to use HN's parametric value function, the parametric value function has to be force to exit after some fixed (large) number of iterations. I tried simulating data using this method; however, households in the resulting data consume nearly everything in their inventory and make large purchases in every period.

Suppose I exchange the parametric value function with the one proposed here while keeping the current period utility and initial inventory estimation as they are in HN(2006), then I arrive at the following likelihood plot.

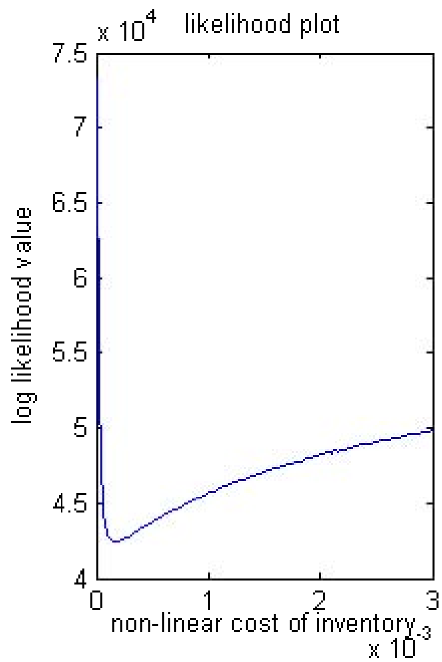

FIGURE 1.10: Likelihood Plot of HN w.r.t. Cost of Inventory

The first plot is the likelihood of the marginal utility of consumption when I keep the cost of inventory terms constant at their true values. The second is a similar plot for the non-linear cost of inventory term. Recall the 3D likelihood plot at the beginning of this subsection, the true values are located 2 and 1.5 respectively. It is understandable that the minimum of the likelihood functions don't match the coefficients. Since the simulation uses the adjusted current utility function and the parametric value function. The point I want to make here is that the parametric value function converges well. Keeping HN's way of estimating initial inventory but adjusting the Bellman equation, I plot a similar set of plots.

We see the likelihood function of the marginal utility of consumption resembles the 3D plot above. Its minimum is close to the true value. However, the likelihood is a little bit more rugged. The likelihood of the cost of inventory does not reach the minimum at the true value 1.5. This makes sense since initial inventory is important for estimating the cost of stocking products. This problem can be remedied by using

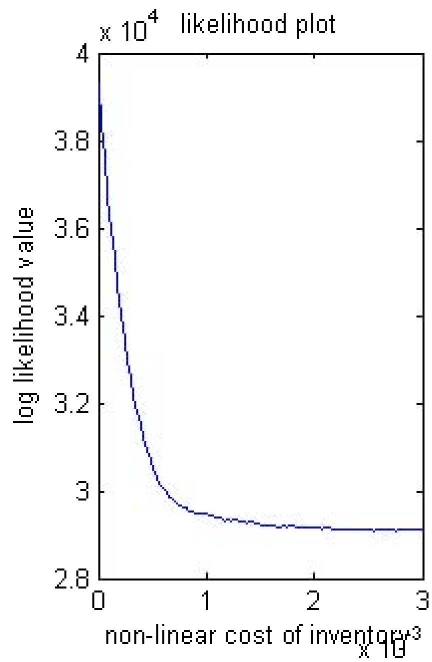

Figure 1.11: Likelihood Plot of Alternative w.r.t. Cost of Inventory

a large number of initial inventory draws, but each draw increases the computation burden.

Knowing that the improved estimation procedure no longer produce dynamic parameter estimates that are sensitive to starting guesses, I apply HN's model and their three stage estimator with my adjustments to laundry detergent panel collected by IRI.

1.3.4 Application to IRI Data

There are 8,939 panelists included in the IRI data set. Participating panelists come from two suburban cities of the United States. One around the Great Lakes area and the other the New England area. Panelists are followed for five years. The format of the data is similar to that used by HN. All shopping trips of all panelists are recorded. For each product purchased, available information include prices, promotional activity, and characteristics of the product. Some product attributes are observable as well. Among those are brand information, volume of purchase, and the type of detergent (liquid vs. powder). Moreover, households' demographical information are also available. Among the observable are race, income, number of children, and size of the family.

Data Description

Although there are many households available, some of them don't satisfy the basic criteria for this project. Not all households reported purchases every year. I drop all households that have any year missing during the five years of the panel. Moreover, if a household does not satisfy reporting criteria set up by IRI, then it is marked as non-static household. These households may have missing reports for some weeks in some year. These households are also eliminated, leaving a total of 1,722 households. Below I present basic statistics of these households.

Being from suburban areas, panelists in the IRI data are more homogenous than the one used by Hendel and Nevo. Over 97.15% of the households are white, leaving only 49 non-Caucasian households. The majority of houesholds are two people households, which does not depart too far for the avarage US household size (2.59). The table of the distribution of household family size is provided below:

Family Size	Frequency	Percentage
1	279	16.20
2	718	41.70
3	286	16.61
4	273	15.85
5	121	7.03
6	45	2.61
Total	1,722	100.00

It's not surprising then to see that 75.15% of households don't have any children. Following Hendel and Nevo, I also consider liquid detergents only. Since the panel is very long, leaving out households who have ever used powdered detergents means eliminating many households. There are 908 households who have consistently used liquid detergent only through the five years.

Prior to applying the data to the improved estimation procedure, a series of further eliminations are performed. These eliminations are due to inconsistencies in the data. I describe each step in detail.

Reduction 1: Based on reported Procter & Gamble estimate, the average American household does 600 loads of laundry per year. Approximately that translates into 1000 fluid ounces of laundry detergent per year per household. In the IRI data set there are many households that grossly violate this estimate. Below a table describes the maximum annual laundry detergent use reported by IRI households:

Household Max Detergent Usage	Frequency	Percentage
1000 oz and less	490	53.98
2000 oz and less	238	80.26
3000 oz and less	108	92.12
4000 oz and less	40	96.50
5000 oz and less	16	98.32
6000 oz and less	7	99.06
Total	899	100.00

Dropping all households who uses more than 3000 fluid ounces of laundry detergent per year, 836 panelists are left.

Reduction 2: Since IRI collect its data separately for stores and panelists, sometimes households' claimed purchases are not observed to have been sold in the weeks reported. When this happens, no promotional activities are observed. Therefore, these households are dropped from the panel.

Reduction 3: Some households make purchases of multiple brands of laundry detergent per week. Most of the time, purchases of the second brand are detergents for delicate fabrics. Since the current model is not adapted to multi-brand purchases, I further reduce the panel to exclude these households. Leaving households who have only made purchases of one brand per week over the duration of the panel, the set of data used in estimation includes 234 households. In the following table below I present summary statistics of this set of data.

Category	Value	Frequency	Percentage
IRI Maket	1	122	52.14
	3	112	47.86
Race	Caucasian	229	97.86
	Non-Caucasian	5	2.14
Children	No Children	191	81.62
	Have Children	43	18.38
Family Size	1	64	27.35
	2	99	42.31
	3	21	8.97
	4	30	12.87
	5	17	7.26
	6	3	1.28

Although only 26% of the data remain, I argue that most households are dropped at random. That is the distribution of the observable household characteristics and purchase behaviors post reduction imitate those prior to reduction. The ratio between Caucasion households and non-Caucasion households remain comparable. The same is true for the distribution of family sizes. The number of households with no children increased by 6.47%, however. This is due to the reduction of households who make multiple brand purchases per week.

Compared to the data set used by Hendel and Nevo, there are many more brands available to households in the 2000's. There are a total of thirty liquid detergent brands reported in IRI. Although, top ten brands still account for roughly 90% of the market. Private label detergents account for slightly under 5% of the market. The shares of each brand is presented in the following table.

Brand Name	Frequency	Share
TIDE	2,902	22.11
ALL	1,852	14.11
ERA	1,650	12.57
PUREX	1,550	11.81
WISK	1,205	9.18
ARM & HAMMER	1,154	8.79
XTRA	583	4.44
AJAX	353	2.69
DYNAMO	316	2.41
TREND	120	0.91
PRIVATE LABEL	598	4.56
Others	749	5.71
Total	13,124	100

The dominant brands are similar between the IRI data and HN's data set. Tide and All are still top share holders of the market. Shares of those brands also remained nearly constant. Tide commanded a market share of 21.4% in HN's analysis and 22.11 reported in IRI. Brand All's market share also stayed around 15% in both sets of data. The only brand that has a drastic change in market share is Era. This brand held a bare 3.7% of the market in HN's data but enjoys the position as the third top brand in IRI, holding a 12.57% market share.

The distribution of sizes are different between the two sets of data. In market HN uses, the top households' pick for bottle sizes are 64 and 128 oz. Together those two shares make over 85% of the market. As the table below shows, that is not the case for IRI. The dominant sizes that households in IRI markets purchase are sizes four and five. That is, over 96% of households buys bottles containing either between 96 and 128 oz or between 128 and 256 oz of detergent.

Size Code	Volume (fl oz)	Frequency	Share
1	[16 32)	2	0.02
2	[32 64)	250	1.90
3	[64 96)	179	1.36
4	[96 128)	8,411	64.09
5	[128 256)	4,258	32.44
6	[256 400)	24	0.18
	Total	13,124	100

Moreover, the options of volume per container also increased in IRI's market. Quoting HN, about 97 percent of the volume of liquid detergent sold was sold in five different sizes, {32, 64, 96, 128, 256}. The same is not true in IRI, aside from the dominance of bottles with 100 oz of laudry detergent (62.71%), other sizes have a more even spread. In estimation, instead of using the same volume defined in HN, I use a weighted average for each size. In the following I apply the improved estimation to the set of data described above.

Results

Hendel and Nevo divide the set of households available to them into six different types and estimate dynamic parameters for each of the six types independently. A household's type depends on where it is located and how many people are in the household. There are two types of locations: urban and suburban and three types of households, households with 2 people or less, 4 people or less, and more than 4 people. Since IRI data are collected exclusively in suburban areas, I only differentiate the types of households by the number of people in the household, definition follows HN.

The first stage features a standard conditional multinomial discrete choice model. Household choose among their prefered brand choices given a predetermined size. Here I compare the first stage estimates of Hendel and Nevo's data set to the IRI data set. Below the first stage estimates are presented:

Stage 1: Brand Choice Conditional on Size

	Hendel & Nevo	IRI
Price	-0.75	-0.73
Feature	1.05	1.61
Display	1.52	1.21
Dummy Variables		
Non-White HH	-0.26	0.51
Large Family	-0.43	-0.12
HH-Brand	$\surd$	$\surd$

The first stage results are comparable to ones obtained in Hendel and Nevo. The only significant difference lies in the coefficient on the dummy for non-Caucasion households. However, due to the small sample size (5 household) of non-white households in the IRI data, I would argue that HN's result for this dummy is more trust worthy. Next, I present the AR1 parameter estimates. Here I only present results for household type I.

Stage 2: Estimates of Price Process

	ω_{1t}	ω_{2t}	ω_{3t}	ω_{4t}
$\omega_{1,t-1}$	0.4858832	0.0052551	0.2928311	0.0911492
$\omega_{2,t-1}$	0.0275118	0.8209847	0.1543715	-0.0277496
$\omega_{3,t-1}$	0.0831225	0.0230693	0.8198198	0.0441283
$\omega_{4,t-1}$	0.0367693	-0.0244625	0.0849779	0.9402269

Unfortunately, since Hendel and Nevo do not present their results for suburban families, there is no reference for comparison. This set of price processes are estimated with inclusive values from the first stage. They are used in estimating households' expected future prices and promotional activities. I use this set of AR1 parameters in the last stage of estimation.

Following the improved estimation method of dynamic parameters, I compare the estimates of the last stage to that of Hendel and Hevo's. I estimate the dynamic parameters for only using households with less than three household members. There are 163 households of this type in the set of IRI data. Each household is followed for 260 periods. (52 weeks for 5 years) This is the counter part to household type four in HN (2006). In Hendel and Nevo's data, household type four has 27 households. Each household is followed for 34 periods. The parameter estimates are presented below

	HN	W
Household Size	1-2	1-2
Utility of Consumption	0.08	0.13
Linear Cost of Inventory	-4.24	2.28
Non-Linear Cost of Inventory	8.20	1.92
Run Time (min)	55.4	20.9

The marginal utility of consumption share the same sign; however, IRI households seem to enjoy consumption more. The cost of inventory is different between the two. The cost of inventory resembles hosuehold type two instead of household type four of HN's data. That is, IRI households have a lower cost of inventory. This is not un-expected, as IRI households live in suburban areas where housing prices are on average lower. Moreover, since inventory interacts with expected future prices, the model implies that households take more advantage of decreased prices. Hence, they are more like to stock on laundry detergent when items are on sale.

A note on the improved estimation procedure: The parameter estimates obtained are not dependent on the starting guesses. In fact, the results presented here are estimated with starting values [2 0.0015 0.0015]. Moreover, even with over nine times as much data and a starting value farther away from the result, the improved estimation is still significant faster than that of Hendel and Nevo.

1.4 Conclusion

Taking inventories and intertemporal substitution into account is imperative for the correct estimation of own and cross price elasticities for storable goods. HN make a big step towards this goal by proposing a model which simplifies the task along several dimensions. This eases the computational burden and therefore allows the estimation of dynamic demand that can incorporate flexible observable heterogeneities. However, the brand choice is modeled as a multinomial discrete choice with a GEV error term. This model implies that households have no systematic brand preferences and any observed differences come from the idiosyncratic shock. The implied demand substitution pattern does not fit the industry under study; hence, the first goal of this paper is to incorporate random coefficients on brand tastes to allow individual differences in brand preferences to generate more reasonable demand substitution patterns.

Along the way, as discussed above, I also discuss several drawbacks from HN's estimation procedure. Moreover, I show how to revise HN's method in order to improve the accuracy of the parameter estimates. I am also able to speed up the estimation significantly compared to HN which is important in order to enrich the model and use more data in future work. I then use the improved estimation routine to identify demand parameters with data in the laundry detergent market.

This paper, in conjunction with HN, is one of the first to help economists understand demand for storable goods. Despite its contributions, dynamic demand estimation largely remains a white spot on the map; much further work will be necessary to remedy this.

2

Estimating the Distributional Impact of Taxes on Storable Goods: A Dynamic Demand Model with Random Coefficients

2.1 Introduction

Obesity is related to increased risks in numerous serious and potentially fatal health problems, including hypertension, type 2 diabetes, heart disease, stroke, etc. With one in four American adults believed to be obese and an estimated medical costs of 147 billion dollars a year (Finkelstein et al. 2009), obesity has become an alarming concern to health professionals and policy makers alike. The American Heart Association recently implicated soda consumption as a major contributing factor to weight gain[1]. In response, the US senate, the District of Columbia, and twelve states have drafted, and in some cases passed, a variety of sugar tax proposals in an attempt to curb households' soda consumption and to generate revenue for health programs. For example, Washington DC passed a legislation that imposes a 6% sales tax on

[1] In an Aug. 2009 American Heart Association statement, soft drinks and sugar sweetened beverages are stated as the number one contributor of added sugars in Americans' diets. The AHA notes that excessive intake of added sugars is implicated in the rise in obesity and cautions that no more than half of a person's daily discretionary calorie allowance should come from added sugars.

sugary soda and New York is debating a one penny per ounce tax. These policies and others similar to them aim to reduce the consumption of sugar-sweetened soft drinks and to generate revenue for an assortment of health initiatives[2].

In this paper, I analyze the welfare impacts of these taxes and predict post-tax consumption changes for different types of consumers. I specify and estimate a structural dynamic demand model of storable goods, where I incorporate households' persistent heterogeneous preferences for product characteristics as household-specific brand and diet random coefficients. This model allows me to accurately estimate consumers' price sensitivity and thus predict the implications of the policies at hand.

To examine the implications of sugar taxes, I predict the total decrease in soft drink consumption, measure the total welfare loss, and determine the post tax market shares. The key to obtaining accurate answers to these questions is to correctly predict consumers' responses to the taxes. Since these policies all primarily impose permanent price hikes, consumers' reactions to policy changes depend on their responsiveness to price variations: Suppose that demand is inelastic. Then the taxes will have little impact on consumption behavior and hence public health, but will generate a substantial revenue. My model accurately predicts consumers' price sensitivity because it captures two particular aspects of soft drinks:

1. Soda is a storable good. Hence, households can take advantage of sales, making extra purchases when prices are low and storing additional products for later consumption. I show in the next section that the data support the hypothesis that households engage in stockpiling. This behavior gives rise to intertemporal substitution. Previous literature on storable goods has shown that static models, which ignore intertemporal substitution, overestimate long run own

[2] There is a large literature debating the effectiveness of sugar taxes in health journals. Most of these works promote the taxes. This line of literature include Jacobson and Brownell (2000), Brownell and Frieden (2009), Brownell et al. (2009), and Kaplan (2010).

price elasticities and underestimate long run cross price elasticities. Therefore, I specify the demand of soda using a dynamic model.

2. Soda is also a differentiated product. There are two dominant brands in this market: Coca-Cola and its rival Pepsi Cola. Moreover, each brand offers diet and regular varieties of drinks. Berry, Levinsohn, and Pakes (1995) and the literature that followed have shown that for differentiated products incorporating consumer heterogeneity in the demand system is important in obtaining realistic predictions. Standard logit models restrict substitution patterns since they exhibit the independence of irrelevant alternatives property. To allow for flexible and realistic substitutions across products, I incorporate household-specific brand-diet random coefficients in the dynamic model of storable goods.

Accounting for both of these aspects, storability and differentiation, is crucial for obtaining accurate predictions for the policies at hand.

I show how each of these aspects influences consumption change predictions and welfare loss estimates by comparing my results with two alternative models: In the first one, I ignore households' stockpiling behavior but allow for consumer heterogeneity in preferences. That is, I specify the demand of soda using a static model with brand and diet random coefficients. I find that the reduction in consumption is exaggerated when households' intertemporal substitution behaviors are not taken into account. In the second model, I ignore persistent consumer heterogeneity but take dynamics into account. I specify the demand of soda using an inventory model that does not allow for persistent random components. In this case, households' reduction in consumption is overestimated and households' welfare loss is underestimated.

Incorporating persistent heterogeneity in consumer preferences is important for accurately predicting the welfare effects of sugar taxes. However, this accuracy does not come without costs. Consider, for a moment, a dynamic demand model of storable goods that does not account for persistent consumer heterogeneity in unobserved product tastes[3]; that is, the model features no random coefficients. This model has a large state space: Not only are all current prices and promotional activities of all products and sizes part of the state, the current period inventories and consumption shocks are also state variables. Thus, even with a relatively coarse grid, the state space easily exceeds 16,000 elements[4]. In the dynamic programming problem[5], the transition matrix, holding $16{,}000^2$ transition probabilities, has to be inverted, which implies that the model and the value function are extremely time consuming to calculate.

Allowing for random coefficients in this dynamic demand model further exacerbates the computation burden: Since households' preferences for product characteristics are unknown, they have to be integrated out. That is, the model, the value function, and the choice probabilities need to be re-calculated for each household, each possible product preference, and parameter try, and then integrated over the distribution of all preferences. Estimating this model, which has 29 parameters, using a fixed point approach would take months, if not years. I adopt a simulated maximum likelihood approach and adapt Ackerberg (2009) to significantly reduce the computation time. My methodology enables me to compute the model, the value functions, and the choice probabilities ahead of the maximization routine. Furthermore, since households' decision processes are independent from each other, I can perform the calculations for each household separately, which allows for straightforward parallel

[3] For example, the dynamic demand model developed in Hendel and Nevo (2006B).

[4] Assuming 6 products with 2 sizes only, we have 12 different prices. Each product can have two types of promotional activities, feature or display, which makes 4 possible advertising schemes. Suppose I limit the inventory to a maximum of 35 servings and allow consumption shocks to have only 10 values, then I arrive at 16,800 (12*4*35*10) possible states of the world.

[5] I use a policy function iteration approach, where all states are discretized. I find that this approach is generally more stable than parametric value function iteration approaches when the state space is very large.

computing. Using the pre-calculated choice probabilities, the maximization over the parameter search takes less than 3 minutes to complete.

I use the estimated distribution of households' preferences to perform policy analyses. Particularly, I study two specific tax proposals: a 10% sales tax and a penny per ounce tax. Using the 2002 to 2004 panels of weekly scanner data provide by IRI, I divide households into three income brackets and for each of them I calculate the welfare effects of the sugar taxes at four pass-through levels: 25%, 50%, 75%, and 100%. More specifically, I simulate the consumption change as a result of these sugar taxes, calculate the compensating variation, and estimate the consumers' welfare loss. Furthermore, I predict the post-tax substitution pattern and market shares. I find that failing to account for dynamics overestimates the long run price elasticity of soft drinks and hence results in overestimating the total reduction in the consumption of sugary soda. Failing to account for persistent consumer heterogeneity in brand and diet tastes, on the other hand, overestimates the total reduction of sugary soda consumption while underestimating the total welfare loss. I show that the reduction in consumption is roughly half of what was previously predicted. Moreover, I find that the policies generate a small deadweight loss but tax poor households more than their rich counterparts.

This paper contributes to the literature on dynamic demand models of storable goods by flexibly incorporating persistent household specific product preferences. Previous literature fails to allow for persistent random components in product tastes. I build on the framework introduced by Hendel and Nevo (2006), but add household specific brand and diet random coefficients. This permits consumers to have intrinsic product tastes. Thus, my model removes the restrictive substitution patterns induced by Hendel and Nevo's conditional logit product choice specification and hence allows for more realistic substitutions across all products. The model and methodology developed here are readily applicable to many studies regarding storable goods, including firms' optimal pricing strategies and antitrust policy analyses.

The rest of the paper will proceed as follows. In the next section, I place this paper in the context of the literature. In section 3, I present industry details and the data. Also, I show evidence for storage behavior. I discuss my model and estimation procedure in detail in sections 4 and 5. In section 6, I show the parameter estimates, examine the model fit, and discuss the implied long run price elasticities. In section 7, I discuss the welfare implications of the sugar taxes and compare these results against the two benchmark studies. I show how welfare implications differ when either dynamics or persistent heterogeneity are ignored. I conclude in section 8 and talk about extensions for future work.

2.2 Literature

Sugar tax policies have been debated in the public health literature extensively. Brownell and Frieden (2009) are in favor of imposing sugar taxes. They argue that sugar sweetened soft drinks are the largest cause for obesity and conclude that sugary sodas should be taxed in order to improve public health. This view is shared by a few other papers, including Mello, Studdert, and Brennan (2006) and Brownell et al (2009). Opposing this point of view, Kaplan (2010) points out that the medical trials used in Brownell et al (2009) do not provide enough evidence for the claim that sugar taxes will decrease obesity in the population. In another line of reasoning, some papers support these policies from a revenue generation point of view. For example, Jacobson and Brownell (2000) acknowledge the fact that sugar taxes may not improve public health but claim that they will generate sizable revenues for health programs. On the opposite side, Gostin (2007), Byrd (2004), and Powell et al. (2007) argue against these taxes on the grounds that they are regressive and target poor and minority groups. These papers all make predictions based on reduced form regressions. In contrast, this paper takes a structural approach to estimating the demand of soft drinks.

There are two strands of literature on structural demand models that this paper fits into: static demand models with consumer heterogeneity and dynamic demand models of storable goods. In terms of static demand models, there has been a plethora of papers that incorporate random coefficients, for example, Berry, Levinsohn, and Pakes (1995) and Nevo (2000). These papers and those that followed show that, in order to obtain realistic predictions for differentiated products, it is important to incorporate consumer heterogeneity in the demand systems. Similarly, this paper incorporates consumer heterogeneity into a dynamic demand system.

This paper fits best into the literature on structural dynamic models of storable goods[6]. There are three papers that are closely related to my work: Erdem, Imai, and Keane (2003), Hendel and Nevo (2006), and Hartmann & Nair (2010). Hartmann and Nair study the demand for razor blades. While this product is storable, they assume inventory is observable to a limited degree: When a new brand is purchased, old inventories disappear and the new inventory is the observed purchase. In this paper, however, inventory is unobservable. Erdem, Imai, and Keane (2003) estimate a dynamic inventory model of demand for households' ketchup consumption. They build a model that allows for a limited degree of unobserved consumer heterogeneity. The model developed in this paper is more flexible than theirs in two respects. First, I allow consumption to depend on current and expected future prices. Second, while Erdem, Imai, and Keane (2003) take consumer heterogeneity into account, households are appropriated in a number of types and are identical within each type. This paper takes a random coefficients approach that more flexibly incorporates persistent

[6] There is also a line of literature on dynamic demand models of durable goods. Examples include Gordon (2006), Conlon (2009), and Nair (2007). Some of these papers, like Gowrisankaran & Rysman (2009) incorporate unobservable heterogeneous tastes into dynamic durable goods demand models. However, in most cases, durable goods are purchased once instead of repeatedly. The decisions that consumers make in these models focus on the timing of product replacement instead of stockpiling. Hence these models and methods are not appropriate in studying consumption behaviors of soft drinks.

consumer heterogeneity: Households' preferences for product characteristics follow distributions whose parameters are estimated within the model.

The paper by Hendel and Nevo is closer to my work. They study households' purchases of laundry detergents and build a model that allows households to stockpile. Their model incorporates heterogeneity as a function of observable household attributes. That is, households' tastes depend on their demographics such as the size of the household, race, and income. But they do not allow for persistent random components which could capture households' intrinsic preferences for product characteristics. This means two households within the same demographical group cannot have different persistent brand or diet preferences. While it is important to account for observable household characteristics, it is not obvious that there is a large correlation between households' demographics and their tastes: A poor household may have strong preference for regular Pepsi while another may love diet Coke instead. I build on Hendel and Nevo's framework but incorporate heterogeneity both as a function of household attributes and as a function of unobservable persistent household specific brand-diet random coefficients.

2.3 Data

I use weekly scanner data provided by Information Resources Inc. (IRI) from 2002 to 2004. This data was collected from two cities, one from the New England area and the other from the Great Lakes area. The data comprises of two components: a household panel and a store panel. The households in the dataset are randomly selected[7]. For each household in each week, I observe whether any soda is purchased; if so, I observe which product was purchased, where it was purchased, how much was

[7] IRI randomly selects a sample of households in each of the two cities and requests their participation. While households' participation decision could possibly introduce a selection bias, the participating households represent the local demographics well. There does not seem to have a selection issue in the panel.

bought, and the total dollar amount paid. From the store panel, in each week, I observe the price charged, the total quantity sold, and all promotional activities for each product that is sold in the store. Sales from these stores account for over 97% of all purchases of soft drinks observed from the households in the panel. For each product, I observe its attributes such as the product name, the brand, the packaging and volume, and whether it is regular or diet.

For each observation, from both the store and the household panel, a unique product identifier (UPC code), a unique store ID, and a week ID are provided. Combining these three identifiers allows me to track prices and promotions for all products available to households on each shopping trip. That is, I observe not only information on the purchase itself but also information on each household's complete choice set.

2.3.1 General Description

In this subsection, I display and describe some statistics on the data.

Household Panel The relevant household attributes I observe include total annual income, family size, and race. Since there is very little variation in the households' ethnicity, I don't make use of this information. I separate households by income per capita into 3 groups: The first income bracket consists of households with a per capita income of less than $10K; the households in the second income bracket have a per capita income between $10K and $20K; and the last bracket holds all households with incomes of more than $20K per capita. In table (2.1) below, I report some statistics associated with each income bracket.

Table 2.1: Household Demographics

	Income Bracket 1	Income Bracket 2	Income Bracket 3
Demographics			
bracket definition (K)	$<$ 10K per capita	[10K 20K)per capita	$\geqslant$ 20K per capita
avg Income (K)	7.526	15.952	36.106
avg household size	2.571	2.426	1.865
number of observations	334	272	175
Soda Purchase			
avg weekly vol purchased (liters)	3.42	3.48	3.54
avg weekly dollars spend	3.11	3.14	3.24
avg annual total number of cola trips	12.12	11.46	10.44
Brand Purchase Shares			
Coke market share	50.13%	46.69%	54.14%
Pepsi market share	40.13%	40.80%	41.72%
Store Brand market share	9.74%	12.51%	4.14%
Diet vs. Regular Shares			
diet market share	39.49%	36.32%	51.55%
regular market share	60.51%	63.68%	48.45%

There are a few interesting patterns in the data: 1) On average, the household size is decreasing in income: From table (2.1), we see that on average, low income households have larger family sizes than higher income households. 2) The average weekly household soda purchase is increasing in income. Poor households purchase slightly less soda than rich households. 3) The number of trips where households bought soda is decreasing in income. That is, on average, high income households make fewer purchases of soft drinks than lower income households. Since they also purchase more units, this implies the rich households tend to buy more soda per shopping trip.

There are two ways of reconciling this fact. First, if we assume soda is purchased only for immediate consumption, then rich households do not drink soda as frequently as poorer households. But when they do, they consume more. The other explanation is inventory: Rich households purchase more soda in fewer trips because they stockpile more than poor households. They are more likely to have larger houses, and hence more storage area at home. The second explanation seems more plausible and indeed in the last part of this section, I show that the inventory explanation better fits the data.

I also show a break down in terms of the products households purchase. In table (2.1) I present the shares of Cola drinks (Coke, Pepsi, and store brand) as well as shares of diet and regular drinks within each income bracket. It's clear from the table that high income households purchase Coke more than the other households. Moreover, they also buy significantly more diet soda than poorer households.

Store Panel From the store panel, I observe weekly prices and advertising information of all items sold[8]. For each store in each week, there are over 250 different

[8] I don't directly observe the price of each product. Instead, I observe the weekly total revenue and the total units sold for every UPC (product). I take the average and use it as a proxy for the product price. In terms of promotional activities, for each product in every week, I observe whether

products offered on average, a combination of all brands, regular/diet, packaging, volume, and flavors. In the estimation of the model, prices and promotions of all products are part of the state variables. Carrying prices and promotions of 250 products is infeasible in the dynamic programming problem. Hence, I need to restrict the set of products used in estimation. In this paper, I confine the product space to include all cola-related products. That is, I allow for all regular and diet versions of Coca-Cola, Pepsi Cola, and store brand Cola drinks in different flavors. Cola drinks make a natural choice: While it is a soft drink, its taste is distinct enough such that other products are not good substitutes. Also, it is one of the few sodas that are caffeinated. And lastly, as the table below shows, Cola drinks have, by far, the largest market share among all varieties of soda.

Table 2.2: Leading Soft Drinks Market Share

Type of Drink	Percentage
Cola Drinks	43.02%
Citrus Flavored	14.04%
Root Beer	8.36%
Canada Dry	3.18%
other	31.40%

From the table above, it's easy to see that cola drinks take a definitive lead in market share. They make 43% of all sodas sold in the market. The second largest type of product, citrus flavored products like Sprite, takes only a 14% market share. From here on out I use the word product to indicate a brand and diet combination. For example, a product may be a diet Coke, regular store brand, etc.

Soft drinks come in a variety of packages and sizes. Furthermore, multiple units can be purchased in any given shopping trip. Due computation burden I restrict

it is on display, on feature, and/or have price discounts.

the maximum volume purchased for each shopping trip to be 4 liters[9]. That is, I drop all households who have made any purchase above the 4 liter limit. Additional volume can be added at increased computation time[10]. The computation burden not only comes from added state space in purchase size but also from increased potential consumption.

This constraint is not as restrictive as it may seem. First, in terms of modeling inventories, this assumption does not limit households' inventory buildup. I limit only the amount households can purchase in each trip. They are free to make purchases without consuming much in each period. Over time, this type of behavior will lead to a substantial inventory. Second, limiting households to have a maximum purchase of 4 liters implies that I capture over 85% of all households in the sample. In table (2.3.1), I show the breakdown of the maximum of households' purchases:

Table 2.3: Maximum Household Vol Purchase

Volume	Cumulative
2 liters	61.12%
4 liters	85.35%
6 liters	92.14%
8 liters or above	100.00%

We see that 61.12% of households make a purchase less than or equal to 2 liters and allowing for purchases of 4 liters or less captures 85.35% of the households.

Holiday Effects Soft drinks consumption may be influenced by holidays such as Christmas. In the data, I observe a very small effect. In the following graph I show weekly cola drinks volume sales, as a percentage of the annual total on purchases,

[9] To be more specific, both bottles and cans are included in the final data set. Moreover, I discretize purchases into units of 2 liters.

[10] I am currently working on including 6 and 8 liters in the estimation for robustness checks.

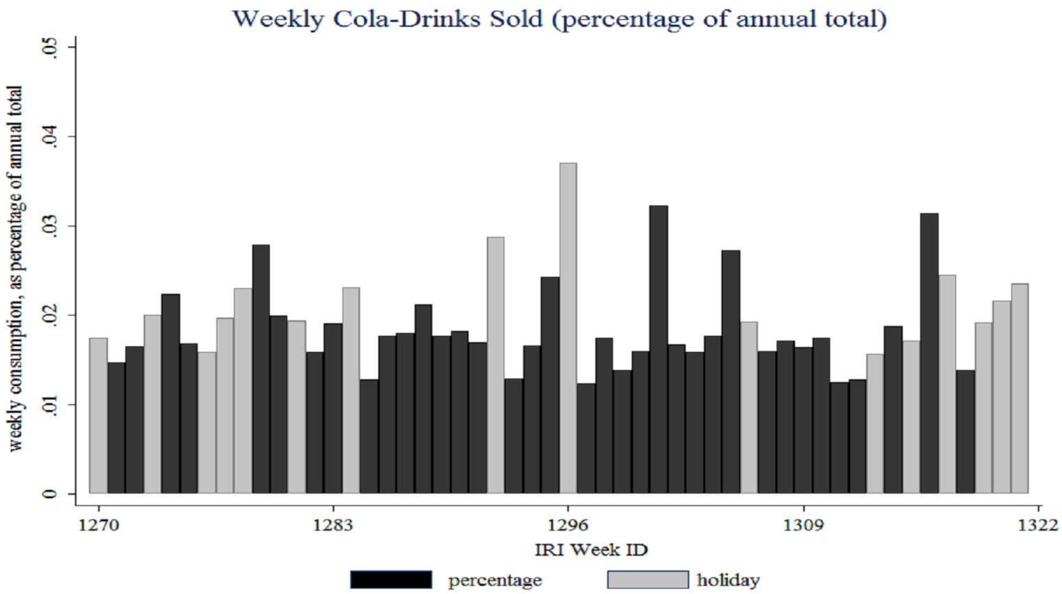

FIGURE 2.1: Volume of cola drink sales by week

observed in stores. The grey bars indicate the presence of a holiday, where holiday is defined very liberally. For example, I count Super Bowl Sunday as a holiday. Here I present the panel from 2004; the graphs for the other two years look very similar.

From the above graph, we see that there are increased sales volumes around some holidays. The dominant ones are around 4th of July and Thanksgiving. In this paper, I do not model holiday effects. I use only data from the 2nd to the 26th week of each year and thus exclude most major holidays such as 4th of July, Thanksgiving, and Christmas. There are some households who are present in the panel for more than one year. I treat them as separate observations. To check, I ran the estimation where each household is counted only once; there is no major difference in the estimates.

Price Variations The identification of parameters comes from variation in relative prices across products and over time; here I show weekly price variations both across brands and within brands. The following table presents the percentage of times I observe different prices offered in the same week between Coke and Pepsi conditioned on the type of product as well as different prices between regular and diet sodas of the same brand.

Table 2.4: Price Variation

Price Difference	Coke vs Pepsi Share	Diet vs Regular Share
no difference	19.63%	72.39%
(\$0 \$0.1)	9.26%	13.27%
[\$0.10 \$0.25)	36.63%	7.56%
[\$0.25 \$0.50)	24.95%	3.96%
[\$0.50 \$0.75)	8.83%	0.94%
[\$0.75 \$1.00)	0.62%	0.94%
[\$1.00 \$2.00)	0.08%	0.94%

note: price differences are in units of 2 liters.

From table (2.4), we see that most of the time Coke and Pepsi offer different prices. In fact I observe the same prices less than 20% of the time. Most of the price differences observed are between 10 to 25 cents. There is less price variation for products of the same brand. However, regular and diet drinks of a given brand are still priced differently about 30% of the time. Most of these price differences are small, mostly just below 10 cents. However, large differences do occur. We see that about 8% of the time regular and diet sodas of the same brands have price differences between 10 to 25 cents. These price variations will help identify households' taste parameters.

2.3.2 Stockpiling Behavior

In this subsection I argue that the inventory model is necessary for modeling consumer behavior. The main competing model is one in which consumption is directly influenced by price and no storage occurs. If this were true, then we should expect to see the following pattern in the data: Sales volume increases when price reductions are available but decreases and remains largely constant when there is no sale. This however is not the case. In the following graph we see that the sales volume drops

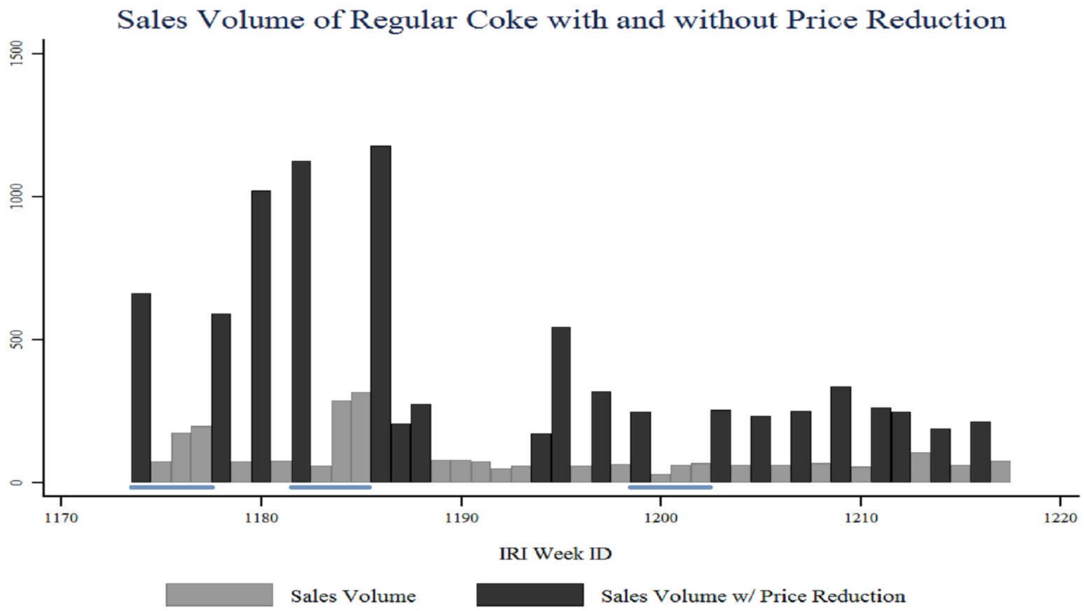

Figure 2.2: Sales Volume of Regular Coke with and without Price Reductions

dramatically immediately after each sale ends; then it slowly grows again. This can be explained by stockpiling: Households fill their inventories during the sale, so that they don't need to purchase much soda immediately afterwards. As time passes after the sale households deplete their inventories. Therefore, more and more households have to restock.

As an example, I use the sale of regular Coke from a store in the data in 2002. This graph shows how sales volume evolves over the duration of a year and how it is influenced by price reductions. Each bar shows the sale volume of a week. The black color indicates price reductions while the grey color is used for weeks without sales.

It's clear that the demand for regular Coke dramatically increases when there is a price reduction. Moreover, we see that the sales volume decreases drastically right after the price reductions; it picks up again in the following weeks. (These occurrences are marked by the blue horizontal lines.) These dips in purchases imply that households stock up on soda when there are price reductions, reducing the need for buying soda immediately afterwards. However, as their stocks become low they make more purchases again. Static models don't account for this effect. They assume all purchased units are immediately consumed.

To further distinguish between the static and the dynamic model I analyze how past prices influence current purchase size choices. To be more precise, I regress the quantity purchased on prices and the time since the last sale ("duration"). In the static model, duration should not have any effect; purchase decisions only depend on current prices and marginal utilities of consumption. However, if the inventory model is correct purchase quantities should be greater if more time passed since the last sale. The reason for this is that households are able to stockpile items not needed for immediate consumption. But when they go through a long time of high prices their inventories are depleted so that it takes a larger purchase to restock.

Table 2.5: Regression of Quantity Purchased

	Coefficient
duration to last sale	10.8128* (3.6670)
price of product	738.137** (14.1617)
constant term	−1630.061** (35.1733)

Note: * at 95% significance level and ** at 99% significance level.

Table (2.5) shows the regression results. Indeed, the coefficient on regression does come out significantly positive. This shows that purchases depend on the past, ruling out any static model.

2.3.3 Persistent Heterogeneous Preferences

Lastly, I present evidence for persistent consumer heterogeneity. In the following table, I show the probability of observing brand and diet purchases conditional on observed last purchases. I display the transition of brand purchases in the first column and the transition of diet soda purchases in the second column:

Table 2.6: Conditional Probability of Purchase

Prob(brand in per. 2 — brand in per.1)		Prob(diet in per. 2 — diet in per.1)	
Prob(coke — coke)	84.81%	Prob(regular — regular)	63.83%
Prob(pepsi — coke)	15.19%	Prob(diet — regular)	36.17%
Prob(coke — pepsi)	26.66%	Prob(regular — diet)	22.04%
Prob(pepsi — pepsi)	73.34%	Prob(diet — diet)	77.96%

If households do not have intrinsic brand or diet preferences, then the product choice in the second period should be independent to that of the first. That is the probability of buying product x in period $t+1$ conditioned buying product x in period t should be roughly the same as that conditioned on buying product y in period t. However, this is not the case, as seen table (2.6). The probability of purchasing Coke in period $t+1$ conditioned on purchasing Coke in period t is 84.81%. The probability of purchasing Coke in period $t+1$ conditioned on purchasing Pepsi in period t, however, is only 26.66%. The same is true for diet preferences. The probability of observing a purchase of a regular soda in period $t+1$ conditioned on purchasing regular in period t is 63.83% while the probability of buying a regular soda in period $t+1$ conditioned on observing a diet purchase in period t is 22.04%. This shows that households have persistent brand and diet preferences.

2.4 Model

2.4.1 Overview

In this model, households' objective is to maximize the expected sum of their discounted future utilities. Households are forward looking: They form expectations over future prices and build inventories accordingly. In each period, they achieve their objective by making three decisions: They choose how much they want to consume, how much they want to purchase, and if they chose to buy anything at all, which product they want to buy. Since the product is storable, households are not forced to consume everything they purchase. They choose the utility maximizing consumption level taking their available inventories and their consumption shocks into account. Any leftover products become part of the storage for future consumption.

Household are each endowed with a set of information. They know their brand and diet preferences, their marginal utilities of consumption, their cost of storage, their marginal utilities of income, and their utilities for promotional activities. In each period, they have information on their beginning of period inventories; they observe all prices and promotional activities in the market[11]; and they know their current period consumption and product shocks. Each household faces two sources of uncertainty for the future: First, they do not observe any future prices or promotional activities. Second, they do not have any knowledge regarding their future shocks.

The timing of events goes as follows. In the beginning of each period, households check their beginning of the period inventory; they obtain information on prices and promotional activities of all products in the market; and they draw shocks associated with consumption and product choice. With the information on current period prices and advertising activities, they form expectations for the next period's prices and promotions. When they visit the store, they simultaneously decide the level of consumption and volume of purchase. If they decide to make any purchases, they will then choose the products conditioned on their preferences, prices, available promotions, and their previously drawn shocks. As they make the purchase, they obtain the utilities associated with the products. They receive the utilities associated with consumption after they consume their predetermined level. At the end of the period, households pay the costs on any leftover inventory.

[11] In this model, I don't specify households' store choices. I assume consumers' store choice does not depend on soda prices. In the estimation, I take store choice as given and use the prices of products available in the store of choice. That is, household's soft drink choice set consists of the items offered in the store visited.

Households are heterogeneous in two ways: First, they are differentiated according to their observable attributes, such as their income, race, etc. Second, they differ by their unobservable persistent preferences for product characteristics. More specifically, households' brand and diet preferences form distributions whose parameters are estimated in the model. As I show in the following subsection, my specification allows for a full correlation structure between brand preferences and diet preferences. It presents a flexible way of incorporating household heterogeneity and provides unrestricted substitution patterns across all products.

2.4.2 Model Setup

Each household, h, in each period, t, maximizes the discounted present value of current and future utility with respect to the level of consumption c^h, the volume of purchase s^h, and the product of choice i^h. For the duration of this paper, I use the word "product", i, to indicate a brand, b, and diet, d, combination. Mathematically, the infinite horizon maximization problem can be expressed as

$$V(\phi_1) = \max_{\{c^h(\phi_t), s^h(\phi_t), i^h(\phi_t)\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[u^h_{tsi} | \phi_1\right] \tag{2.1}$$

where ϕ_t denotes the state variables, which include prices, promotional activities, inventory, and all shocks. I use the standard notation δ for the discount factor. The small letter u denotes the flow utility. Hence, u^h_{tsi} is the utility of household h in period t for product i of size s.

The flow utility is defined as follows:

$$u^h_{tsi} = U\left(c^h_t, \upsilon^h_t | \Theta^h\right) - C\left(inv^h_{t+1} | \Theta^h\right) + U^h_{tsi}\left(p_{tsi}, a_{tsi}, i^h_{tbd}, \varepsilon^h_{tsi} | \Theta^h\right) \tag{2.2}$$

It comprises of three components, the utility of consumption $U(c^h_t, \upsilon^h_t | \Theta^h)$, the cost of inventory $C(inv^h_{t+1} | \Theta^h)$, and the utility of purchase $U^h_{tsi}(p_{tsi}, a_{tsi}, i^h_{tbd}, \varepsilon^h_{tsi} | \Theta^h)$, where Θ^h denotes the household specific taste parameters.

Utility of Consumption: The utility of consumption depends on the amount c_t^h household h consumed and a stochastic shock υ_t^h. The consumption is equal to the total of all product varieties consumed by household h in period t. That is, $c_t^h = \sum_i c_{ti}^h$. In each period, each household receives a stochastic shock υ_t^h. This allows households' consumption to depend on factors that are unobserved by the researcher. For simplicity, the utility of consumption takes the functional form:

$$U(c_t^h, \upsilon_t^h | \Theta^h) = \alpha^h \log(c_t^h + \upsilon_t^h) \tag{2.3}$$

where the consumption level and the stochastic shock enter additively into the utility. The household specific marginal utility of consumption is denoted by α^h.[12] Naturally, consumption is restricted to be non-negative. The consumption shock is also restricted to be non-negative. I assume it is distributed i.i.d over time and households and follow a log-normal distribution with mean μ_c and standard deviation σ_c.[13] This stochastic component influences the utility in the following way: If household h draws a high realization of υ_t^h, then its marginal utility of consumption would be smaller than if a low realization were drawn. Therefore, a high realization decreases households' need to consume and hence make demand more elastic.

Cost of Inventory: Household's cost of inventory follows a quadratic functional form:

$$C\left(inv_{t+1}^h | \Theta^h\right) = \left(\beta_1^h inv_{t+1}^h + \beta_2^h \left(inv_{t+1}^h\right)^2\right) \tag{2.4}$$

where β_1^h and β_2^h are household specific linear and quadratic costs. The end of period inventory, inv_{t+1}^h, is equal to the beginning of period inventory, inv_t^h, plus the total current purchase, s_t^h, minus the total consumption, c_t^h. The total inventory

[12] Technically, the marginal utility of consumption is $\alpha^h / \left(c_t^h + \upsilon_t^h\right)$, which depends on both the consumption level and the consumption shock. However, α^h is the only estimable part of this expression. In order to avoid complicated terminology, I call α^h the marginal utility of consumption.

[13] I discuss all assumptions in detail later in this section.

is restricted to be non-negative, $inv_t^h \geqslant 0$. That is, households are not allowed to consume more than what is available in their storage. Mathematically, this translates to $inv_{t+1}^h = inv_t^h + s_t^h - c_t^h \geqslant 0$. Moreover, households cannot make negative purchases, hence $s_t^h \geqslant 0$.

Utility of Purchase: The last component of the flow utility is associated with household h's purchase in period t. It is a function of prices, promotional activities, product preferences, and a random shock to household h's choice. More specifically, the purchase utility can be represented as:

$$U_{tsi}^h \left(p_{tsi}, a_{tsi}, i_{bd}^h, \varepsilon_{tbs}^h | \Theta^h\right) = -\gamma_1^h p_{tsi} + \gamma_2^h a_{tsi} - \sum_b \sum_d 1_{t,-b,-d}^h \left(\zeta_b^h + \psi_d^h\right) + \varepsilon_{tsi}^h \quad (2.5)$$

It contains of four pieces. The first piece is household h's disutility of price. It is equal to the negative of household h's marginal utility of income, γ_1^h, multiplied by the price, p_{tsi}, paid for product i of volume s. The second part is the utility associated with advertising, which captures the advertising effects on demand. This is equal to the household specific utility for advertising γ_2^h if any promotional activities are available for product i of volume s. a_{tsi} is an indicator: It denotes the presence of any features or display. The next component represents household h's intrinsic preferences for brand b and diet drinks d, denoted by ζ_b^h and ψ_d^h. These household specific brand and diet random coefficients can be seen as disutilities household h incur if its preferred product, $i_{bd}^h = \left\{b^h, d^h\right\}$, is not purchased.

More specifically, I allow for 9 different types of households. Households of each type, except one, have preferences for a brand and diet combination. That is, households may prefer diet Coke, regular Coke, or simply Coke. The same holds true for Pepsi. Households may prefer to drink regular sodas or diet ones regardless of the brand or they may have no special preference for any products. The disutility enters household's product purchase utility as follows: Assume household h is a diet Coke lover. Then it incurs no disutility in period t if diet Coke is purchased. However, if it makes a purchase of regular Coke, then it loses utility equal to its household specific preference for diet soft drinks, $-\psi_D^h$. Similarly, if it buys diet Pepsi, then it loses an amount of utility that is equal to its preference for Coke, $-\zeta_C^h$. Now suppose this household buys regular Pepsi instead of its preferred diet Coke, then it incurs disutility equalling the sum of its preference for Coke and diet drinks, which is $-\psi_D^h - \zeta_C^h$. The table below shows in detail how purchases of different items impact the utilities of households of each type:

Table 2.7: Specification of Product Utilities for Different Household Types

	Item(s) of Purchase					
HH Type	Diet C	Regular C	Diet P	Regular P	Diet G	Regular G
Diet Coke	0	$-\psi_D^h$	$-\zeta_C^h$	$-\psi_D^h - \zeta_C^h$	$-\zeta_C^h$	$-\psi_D^h - \zeta_C^h$
Regular Coke	$-\psi_R^h$	0	$-\psi_R^h - \zeta_C^h$	$-\zeta_C^h$	$-\psi_R^h - \zeta_C^h$	$-\zeta_C^h$
Neutral Coke	0	0	$-\zeta_C^h$	$-\zeta_C^h$	$-\zeta_C^h$	$-\zeta_C^h$
Diet Pepsi	$-\zeta_P^h$	$-\psi_D^h - \zeta_P^h$	0	$-\psi_D^h$	$-\zeta_P^h$	$-\psi_D^h - \zeta_P^h$
Regular Pepsi	$-\psi_R^h - \zeta_P^h$	$-\zeta_P^h$	$-\psi_R^h$	0	$-\psi_R^h - \zeta_P^h$	$-\zeta_P^h$
Neutral Pepsi	$-\zeta_P^h$	$-\zeta_P^h$	0	0	$-\zeta_P^h$	$-\zeta_P^h$
Diet Neutral	0	$-\psi_D^h$	0	$-\psi_D^h$	0	$-\psi_D^h$
Regular Neutral	$-\psi_R^h$	0	$-\psi_R^h$	0	$-\psi_R^h$	0
No Preference	0	0	0	0	0	0

Note: C denotes Coke; P denotes Pepsi; and G denotes Generic / Store Brand.

HH is a shorthand for household.

This table not only shows how households vary by their types, it also shows how households within the same type differ from each other: The brand and diet random coefficients are household specific; therefore, households in each type have differing degrees of tastes for each product. That is, some households may have stronger preferences for a product than others within their type. This specification of consumer heterogeneity allows for a very flexible correlation structure across all products. Not only do households have differing preferences for different products; their brand choices may also be correlated to their diet preferences.

The last component of the brand purchase utility, ε^h_{tsi}, is a random product choice shock. It influences households' idiosyncratic per-period product taste. This shock is distributed i.i.d. across all products and sizes. I assume it follows a type I extreme value distribution.

In this model, the utility from product differentiation is enjoyed at the moment of purchase. While households have intrinsic preferences for different brands and diet drinks, these preferences and the differences between products affect households' behaviors exclusively at the store. Once the product is bought, they are no longer differentiated. That is, all products in storage provide the same level of utility during consumption for household h. Since consumptions and inventories are not observed, this specification avoids putting extra structure on consumption rates of different items in inventory. Moreover, it reduces the state space. Instead of keeping track of all inventories of all products in store, I need to take care of only the total quantity.

There are a few distributional assumptions in this model. They arise from two sources: uncertain future faced by consumers and household-specific information unobservable to the researcher. From the households' perspective, uncertainty comes from future utility shocks and future prices and promotional activities. Recall, consumers have two utility shocks: consumption shocks and unobserved product choice shocks for each brand-diet-size combination. Both of these shocks follow a pre-specified distribution. Furthermore, households also face uncertain future prices and promotions. Since they are forward looking, they form expectations over those prices and advertising activities. From the researcher's perspective, households' persistent preferences and their initial inventories are unobservable. Hence both are assumed to follow some distribution where the parameters that govern these distributions are estimated. Below I discuss all the assumptions and their implications.

A1. Consumption Shocks The consumption shocks, v^h_t, are distributed independently and identically across households and over time. They follow a log-normal distribution with mean μ_c and standard distribution σ_c. These consumption shocks introduce randomness in households' need to consume. Suppose a high realization of υ_t^h is drawn, then household h will derive less utility from each unit of product. Hence it would reduce this household's current period consumption and make its demand for soda more elastic.

A2. Product Shocks The product choice shocks, ε_{tsi}^h, follow a type I extreme value distribution and are independent and identically distributed across each household, period, product, and the size of purchase. This assumption is standard for discrete choice problems. Here it introduces randomness not only in households' product purchases but also in volumes of purchases.

A3. Prices and Promotions Prices and advertising activities follow an exogenous first order Markov process. This assumption is often employed in the empirical literature. It helps to reduce the state space in the dynamic programming problem. This is a simplifying assumption on households' expectations for future prices and promotions: It implies that consumers use only the current period prices and promotions in predicting future prices and promotion. This assumption seems to be a reasonable approximation of the formation of consumers' price expectations. Higher order processes would require households to recall more prices from previous visits. The main concern might be seasonal price fluctuations, when the probability of advertising increases. However, my data does not show significant seasonal effects. Furthermore, since I use only data from the 2nd to the 26th week of each year, I exclude most holidays.

A4. Household Specific Preferences The random coefficients

$$\Theta^h = \left\{\alpha^h, \beta_1^h, \beta_2^h, \gamma_1^h, \gamma_2^h, \zeta_C^h, \zeta_P^h, \psi_D^h, \psi_R^h\right\}$$

are distributed jointly log-normal with mean μ_Θ and variance-covariance matrix Σ_Θ. They capture households' persistent heterogeneous tastes. Since the econometrician does not observe them, the joint distribution of these preferences is estimated. Also, these parameters are naturally bounded on either side of the real line. For example, the marginal utilities of consumption are non-negative. Hence, the log normal distributions are fitting. I further assume that the off diagonal elements of the variance-covariance matrix are zero. This restriction can be lifted at the cost of a slightly increased computation burden.

A5. Initial Inventory Initial inventory inv_0^h is distributed log-normal with mean μ_{inv} and standard deviation σ_{inv}. This is an initial conditions problem: Households' initial inventories are unknown to the econometrician and hence need to be estimated. One way of estimating them is to start at an arbitrary level and use the first few weeks of observations to generate the distribution of inventories. To obtain a good estimate using this method, the panel has to be relatively long. For example, Hendel and Nevo (2006B) take this approach. Their panel spans 2 years and they use the first 11 weeks to recover estimates for the initial inventories. In this paper, however, the length of panel is relatively short, 25 weeks. Hence, I adopt a different method. I use the model and observed households' purchase behaviors to explicitly estimate a distribution over the initial inventory. I assume initial inventories follow a log-normal distribution, since they are bounded below at zero.

2.5 Estimation

To estimate this model, I need to solve the infinite horizon maximization problem faced by each household in each period. For a large state space, this process can be time consuming. Several key factors of the model further complicate this process:

First of all, current period inventories are unknown since neither consumption levels nor the initial inventories are observed. To calculate the current inventory, both the current consumption and the initial inventory have to be estimated. Secondly, since all parameters are random coefficients, the likelihood has to be integrated over the joint distribution of these parameters. Lastly, households' persistent heterogeneous brand and diet tastes are unobservable to the researcher. Therefore, for every household, the maximization process along with the integration over random parameters have to be calculated repeatedly for all nine household types. This presents a significant computation burden. I adopt a simulated maximum likelihood approach and adapt Ackerberg (2009)'s importance sampling method to reduce the computation time. I start the discussion of the estimation by providing an overview and then move on to a more detailed discussion of the technical details.

2.5.1 Overview

Recall from the data, we observe only the purchases of all households over time. In each period, we know whether household h made any purchases. If a purchase was made, then we observe which product was purchased and how many units were bought. However, no information is available on households' consumptions, inventories, or their intrinsic tastes. The goal is to maximize the probability of observing this sequence of purchases with respect to the parameters that govern the distribution of the random coefficients and the initial inventory along with the shares of each household type.

Assume for now that we are in a world where the econometrician observes not only the purchases but also both the households' types, τ^h, and their initial inventories, inv_0. Then for a given set of parameters, the probability of observing this sequence of purchase decisions across all households and over time,

$\left(pur_1^1 \cdots pur_T^1 \ldots\ldots pur_1^H \cdots pur_T^H\right)$, as a function of the state variables, $\left(\phi_t^1 \cdots \phi_t^H\right)$, is given by

$$L = \int_{\Theta^h} \Pr\left(pur_1^1 \cdots pur_T^1 \ldots pur_1^H \cdots pur_T^H | \phi_t^1 \cdots \phi_t^H, inv_0^1 \cdots inv_0^H, \tau^1 \cdots \tau^H, \Theta^h\right)$$

$$\cdot dF\left(\Theta^h | \mu_\Theta, \Sigma_\Theta\right)$$

$$\text{where} \;\; \phi_t^h = \left\{p_t,\, \upsilon_t^h,\, inv_t^h\left(pur_1^h \cdots pur_{t-1}^h,\, \upsilon_1^h \cdots \upsilon_{t-1}^h,\, inv_0^h\right)\right\}$$

Here I use pur_t^h denote the purchase decision of household h in period t. The state variable ϕ_t^h includes the vector of prices and promotions, p_t; the current period consumption shock, υ_t^h; and the beginning of the period inventory, inv_t^h. This inventory is a function of previous decisions and states. It depends on last period's consumption and purchase decisions, which in turn depend on consumption shocks and available inventory.

I assume households are independent from each other; therefore, the probability of observing a sequence of purchases from household h is independent from observing the purchases of household h'. Hence, the probability of observing the entire purchase sequences of all households can be re-written as:

$$L = \prod_{h=1}^{H} \int_{\Theta^h} \Pr\left(pur_1^h, \cdots, pur_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right) \cdot dF\left(\Theta^h | \mu_\Theta, \Sigma_\Theta\right) \tag{2.6}$$

Furthermore, recall that purchases can be decomposed into product purchases, i_t^h, and volume purchases, s_t^h. That is

$$\Pr\left(pur_1^h, \cdots, pur_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right) = \Pr\left(s_1^h \cdots s_T^h, i_1^h \cdots i_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right).$$

Since the product choice shocks are i.i.d., households' brand purchases are independent over time. Therefore rewriting the probability of observing the sequence of purchases once again, we arrive at:

$$L = \prod_{h=1}^{H} \int_{\Theta^h} \left[\prod_{t=1}^{T} \Pr\left(i_t^h | s_t^h, \phi_t^h, \tau^h, \Theta^h\right) \right] \Pr\left(s_1^h \cdots s_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right) dF\left(\Theta^h | \mu_\Theta, \Sigma_\Theta\right) \tag{2.7}$$

From this equation, we can see that household's two purchase decisions, product and size, are treated differently. The product choices, i_t^h, are uncorrelated across periods, hence the decision making process is static. That is, conditional on households' intrinsic preferences, their brand and diet choices in period t are independent from the ones in period $t+1$. Given that product shocks are distributed according to a Type I extreme value distribution, the probability of seeing any particular brand-diet purchase follows the standard logit formula:

$$\Pr\left(i_t^h | s_t^h, \phi_t^h, \tau^h, \Theta^h\right) = \frac{\exp\left[\left(-\gamma_1^h p_{tsbd} + \gamma_2^h a_{tsbd} - \sum_b \sum_d 1_{t,-b,-d}^h \left(\zeta_b^h + \psi_d^h\right)\right)\right]}{\sum_b \sum_d \exp\left[\left(-\gamma_1^h p_{tsbd} + \gamma_2^h a_{tsbd} - \sum_b \sum_d 1_{t,-b,-d}^h \left(\zeta_b^h + \psi_d^h\right)\right)\right]} \tag{2.8}$$

The volume purchase decisions, s_t^h, on the other hand, depends on current consumption as well as past consumption levels and purchase decisions. Consumptions and volume purchases are determined from the households' utility maximization and are correlated over time. Hence, the probability of observing a sequence of size purchases depends on households' optimal policy rule, which is a dynamic programming problem, and hence cannot be estimated separately for each period.

However, households' consumption shocks υ_t^h, types τ^h, and initial inventories inv_0 are not observed. Therefore, to estimate the model, we need to integrate equation (2.7) over the consumption shocks, the possible household types, and the initial inventory. Mathematically, this can be expressed as:

$$L = \prod_{h=1}^{H}\left\{\int_{\Theta^h}\left[\int_{\upsilon}\int_{inv_0}\sum_{\tau}\left(\prod_{t=1}^{T}\Pr\left(i_t^h|s_t^h,\phi_t^h,\tau^h,\Theta^h\right)\Pr\left(s_1^h\cdots s_T^h|\phi_t^h,inv_0^h,\tau^h,\Theta^h\right)\right)\right.\right.$$
$$\left.\left.\Pr\left(\tau^h\right)dF\left(inv_0^h\right)dF\left(\upsilon_1^h\cdots\upsilon_T^h\right)\right]dF\left(\Theta^h|\mu_\Theta,\Sigma_\Theta\right)\right\}$$

The calculation of this likelihood function involves the following steps:

1. Solve the model and compute the value function: find optimal consumption & purchase
2. calculate the choice probabilities

This process has to be repeated for each household, each household type, initial inventory level, and parameter try. Hence it is clear where the computation burden arises. To be more specific, there are three hurdles involved in estimating this model. The first hurdle comes from estimating unobserved inventory levels; the second from carrying a large state space; and the last from estimating households' persistent product preferences. Each is discussed in detail below.

First, since I observe neither initial inventories nor consumption decisions, households' per-period inventory levels are unknown. I estimate these inventory levels by deriving households' optimal consumption decisions from the model and estimating a distribution over the initial inventories. Assume that we observe households' initial inventory levels. Then, conditional on the consumption shocks, households' utility maximizing behaviors determine the optimal consumption and purchase choices and therefore the end of period inventories for the first period. We can roll this process forward and obtain all inventories for all periods. To get this process started, I estimate a distribution of initial inventories.

The second hurdle is associated with having a large state space. For each household in each period, the state space encompasses all prices and promotional activities of all products of all sizes in the store, the consumption shock, and the current inventory. This is a large set of state variables. Since I use the policy function iteration approach to compute the value function[14], I have to discretize the state space. I allow for 30 levels of inventory and discretize the consumption shock into 15 points. Thus, there are 10,800 possible states. Since the transition probabilities of the state, $10{,}800^2$ elements[15], has to inverted, it is time consuming to compute the value function and to solve the model. Therefore, it's computationally infeasible to allow for a full state. To reduce the computation burden, I adopt the standard inclusive value approach to shrink the state space.

Lastly, households' persistent brand and diet preferences are unobserved. To obtain these household specific preferences, I estimate the distribution of the random coefficients along with the shares of household types in the market. This also implies that the dynamic programming problem and the choice probabilities need to be calculated for each household, each household type, each initial inventory, and each parameter try. It takes roughly 10 seconds to solve the model, calculate the value function, and compute the choice probability once on an 8 core 2.80GHz computer. However, given that I observe about 800 households and allow for 9 household types, and that the maximization generally takes at least 3000 iterations[16] to solve this model, this implies I need to repeat the process at least 21 million times. This makes a total 216 million seconds, which is about 6 years, if a standard maximum likelihood estimator is used. I adopt a simulated maximum likelihood approach and build on the important sampling method proposed by Ackerberg (2009) to significantly cut

[14] I also experimented with parametric value function iteration and found the result less accurate and less stable. This is due to the fact that the state space is large and the number of parameters is big.

[15] There are 6 products and each of which contains 2 sizes. Hence, there are 12 different prices and 2 different advertising indicators. Along with 30 inventory levels and 15 consumption shock, the state space has 10,800 ($12\times2\times30\times15$) elements.

[16] This estimate of the iteration number assumes a convergence criterion of 1e-3.

down computation time. In estimating this model, allowing for parallel computing on 8 cores, the entire estimation process takes about 3 days.

The advantage of this estimation approach is threefold: 1) The model and the value function need to be calculated once for each household type and parameter try but not for every household, saving time by a factor of 800 (number of households in the sample). 2) The model, value function, and choice probability calculations can be performed outside the maximization routine. 3) This method allows for easy parallel computing. Together, these points significantly cut down computing time. I discuss the estimation procedure in detail in the section below.

2.5.2 Estimation Procedure

I break the estimation into five steps. Here I discuss each step in more detail:

Step 1 - Initial Parameter Distribution To start the estimation process, I define the initial distributions of all random coefficients and the distribution over the initial inventory. I assume these parameters are distributed according to log-normal distributions. I arbitrarily choose the parameters that govern these distributions, $(\mu_\Theta^0, \Sigma_\Theta^0)$ and $\left(\mu_{inv_0}^0, \sigma_{inv_0}^0\right)$.[17] That is, I define

$$g\left(\widetilde{\Theta}^h \left| \mu_\Theta^0, \Sigma_\Theta^0\right.\right)$$

the joint distribution of all random parameters and

$$g\left(\widetilde{inv}_0 \left| \mu_{inv_0}^0, \sigma_{inv_0}^0\right.\right)$$

the distribution of the initial inventory. I draw N sets of parameters and initial inventories according to these distributions. I denote these drawn parameters by $\widetilde{\Theta}^h$ and drawn initial inventory by $\widetilde{inv}_0$.

[17] It's important to choose distributions that have a sufficiently large variance. Otherwise, certain parameters are virtually excluded from being drawn. This can lead to misestimation if the true parameters lay outside of the range of the initial parameter draws.

Step 2 - Product Choice Recall that the product choices are static in that they are not correlated over time. Hence, the product choice conditional on quantity purchased follows the standard logit formula. For each household, household type, and parameter draw, I calculate the following:

$$\begin{aligned}
&\Pr\left(i_t^h|\phi,\tau^h,s_t^h,\widetilde{\Theta}_{sim}^h\right)\\
= &\frac{\exp\left[\left(-\widetilde{\gamma}^h p_{tsbd}+\widetilde{\gamma}^h a_{tsbd}-\sum_b\sum_d 1_{t,-b,-d}^h\left(\widetilde{\zeta}_b^h+\widetilde{\psi}_d^h\right)\right)\right]}{\sum_b\sum_d \exp\left[\left(-\widetilde{\gamma}^h p_{tsbd}+\widetilde{\gamma}^h a_{tsbd}-\sum_b\sum_d 1_{t,-b,-d}^h\left(\widetilde{\zeta}_b^h+\widetilde{\psi}_d^h\right)\right)\right]}
\end{aligned}$$

Furthermore, to reduce the state space, we can combine the utilities for all products of a given size into a measure called the inclusive value. An inclusive values is basically a weighted average of the preferences across all products of a particular size. This method avoids keeping track of all the prices and promotions of all the brand-diet-size combinations. Instead, I only need to follow a single inclusive value for each size. Mathematically, it is defined as:

$$\omega_{ts}^h=\log\left\{\sum_b\sum_d\exp\left(-\widetilde{\gamma}^h p_{tsbd}+\widetilde{\gamma}^h a_{tsbd}-\sum_b\sum_d 1_{t,-b,-d}^h\left(\widetilde{\zeta}_b^h+\widetilde{\psi}_d^h\right)\right)|\widetilde{\Theta}_{sim}^h\right\}$$

These inclusive values, ω_{ts}^h, can be interpreted as quality adjusted price measures for each available size. They are household and time specific; hence they are calculated for each household, household type, period, size, and parameter draw. Lastly, recall that households assume prices and promotions follow an exogenous first order Markov process. Therefore, I need to compute the transition probabilities of these inclusive values.

Step 3 - Dynamic Programming Problem Households' optimal purchase size choices depend on their current level of inventory, which in turn depends on past consumption and purchase decisions. Hence these size choices are correlated over time. Households' optimal controls, the consumptions and purchases, are solutions to their dynamic programming' problem. The Bellman equation equivalent to households' infinite horizon maximization problem is:

$$
\begin{aligned}
& V\left(\phi_1^h\right) \\
= \; & \max_{\{c,s,i\}} \left\{ \left[\widetilde{\alpha}^h \log\left(c + \upsilon^h\right) - \left(\widetilde{\beta}_1^h inv' + \widetilde{\beta}_2^h inv'^2 \right) \right.\right. \\
- \; & \left. \widetilde{\gamma}_1^h p_{si} + \widetilde{\gamma}_2^h a_{si} - \sum_b \sum_d 1_{ti}^h \left(\widetilde{\zeta}_b^h + \widetilde{\psi}_d^h \right) + \varepsilon_{si}^h \right] \\
& \left. + \delta E \left[V\left(\left(\phi^h\right)' \right) | \phi_1^h, c, s, i \right] \right\}
\end{aligned}
\tag{2.9}
$$

where households maximize the sum of their current period flow utilities plus the discounted expected future values with respect to consumption, product and size purchases, conditional on current inventories, prices and promotions, and shocks. Incorporating the inclusive values, the simplified Bellman equation is:

$$
\begin{aligned}
& V\left(\phi^h\right) \\
= \; & \max_{\{c,s,i\}} \left\{ \left[\widetilde{\alpha}^h \log\left(c + \upsilon^h\right) - \left(\widetilde{\beta}_1^h inv' + \widetilde{\beta}_2^h inv'^2 \right) + \omega_s^h + \varepsilon_s^h \right] \right. \\
& \left. + \delta E \left[V\left(\left(\phi^h\right)' \right) | \phi^h, c, s, i \right] \right\}
\end{aligned}
$$

I solve for the optimal consumptions and purchases for each possible state using the policy function iteration method. Recall that I discretize all states, hence the value function for each possible state of the world is calculated exactly.

With the optimal consumption and size purchase decisions computed, we can then calculate the probability of observing the sequence of size purchases made by each household, for the given household type and parameter draw:

$$
\begin{aligned}
& \Pr\left(s_1^h \cdots s_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right) \\
= & \prod_t \Pr\left(s_t^h | p_t,\, \upsilon_t^h,\, inv_t^h\left(s_1^h \cdots s_{t-1}^h,\, \upsilon_1^h \cdots \upsilon_{t-1}^h,\, inv_0^h\right), \tau^h, \Theta^h\right) \\
& \text{where } \Pr\left(s_t^h | p_t,\, \upsilon_t^h,\, inv_t^h\left(s_1^h \cdots s_{t-1}^h,\, \upsilon_1^h \cdots \upsilon_{t-1}^h,\, inv_0^h\right), \tau^h, \Theta^h\right) \qquad (2.10) \\
= & \frac{\max\limits_{c_t^h}\left\{U\left(c_t^h, \upsilon_t^h\right) - C\left(inv_{t+1}^h\right) + \omega_s^h + \delta E\left[V\left(\phi_{t+1}^h\right) | \phi^h, c_t^h, s_t^h\right]\right\}}{\sum_{\widetilde{s}}\left\{\max\limits_{c_t^h}\left\{U\left(c_t^h, \upsilon_t^h\right) - C\left(inv_{t+1}^h\right) + \omega_s^h + \delta E\left[V\left(\phi_{t+1}^h\right) | \phi^h, c_t^h, \widetilde{s}_t^h\right]\right\}\right\}}
\end{aligned}
$$

Having determined both the probability of brand purchases and size purchases, we can form households' choice probabilities.

Step 4 - Choice Probability For each household and parameter draw, I calculate the choice probability conditioned on household h's type and its beginning of the period inventory:

$$
\begin{aligned}
& \widetilde{L}_{sim}^h\left(pur_1^h \cdots pur_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right) \\
= & \left[\prod_t \Pr\left(i_t^h | \phi^h, \tau^h, s_t^h, \widetilde{\Theta}_{sim}^h\right)\right] \Pr\left(s_1^h \cdots s_T^h | \phi_t^h, inv_0^h, \tau^h, \widetilde{\Theta}_{sim}^h\right) \qquad (2.11)
\end{aligned}
$$

This conditional choice probability of household h is equal to the probability of observing a sequence of product choices conditioned on the the sequence of size choices times the probability of observing that sequence of size choices. To remove the conditioning arguments, I integrate the above equation (2.11) over all household types, initial inventories, and consumption shocks for each parameter draw:

$$
\begin{aligned}
& \widetilde{L}_{sim}^h\left(pur_1^h \cdots pur_T^h | \phi_t^h, \widetilde{\Theta}_{sim}^h\right) \\
= & \int_\upsilon \int_{inv_0} \sum_\tau \widetilde{L}_{sim}^h\left(pur_1^h \cdots pur_T^h | \phi_t^h, inv_0^h, \tau^h, \Theta^h\right) \\
& \Pr\left(\tau^h\right) dF\left(inv_0^h\right) dF\left(\upsilon_1^h \cdots \upsilon_T^h\right) \qquad (2.12)
\end{aligned}
$$

This gives me a set of choice probabilities for each household. To be more precise, each household has N choice probabilities[18]. With this large set of choice probabilities, I compute and maximize the total log likelihood.

Step 5 - Log Likelihood Function To compute the total likelihood function, for each household I randomly select a subset of the parameter draws along with their corresponding choice probabilities. Then the simulated maximum likelihood function is a weighted average of all choice probabilities:

$$\widetilde{L} = \prod_{h=1}^{H} \left\{ \frac{1}{\#\text{draws}} \sum_{\#\text{draws}} \left[\widetilde{L}_{sim}^{h} \left(pur_1^h \cdots pur_T^h | \phi_t^h, \widetilde{\Theta}_{sim}^{h} \right) \frac{h\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta, \Sigma_\Theta\right)}{g\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta^0, \Sigma_\Theta^0\right)} \right] \right\} \tag{2.13}$$

where $h\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta, \Sigma_\Theta\right)$ denotes the estimated distribution over random coefficients and over initial inventories. The component $h\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta, \Sigma_\Theta\right) \Big/ g\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta^0, \Sigma_\Theta^0\right)$ can be seen as a weight on the likelihood: Parameters are allocated more weight if they explain households' decisions well. Note, since all the choice probabilities are calculated prior to maximization, $\widetilde{L}_{sim}^{h}\left(pur_1^h \cdots pur_T^h|\phi_t^h, \widetilde{\Theta}_{sim}^{h}\right)$ is nothing more than a large matrix of values. Moreover, since $g\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta^0, \Sigma_\Theta^0\right)$ is pre-defined, the maximization routine updates only the weight $h\left(\widetilde{\Theta}_{sim}^{h} \,|\mu_\Theta, \Sigma_\Theta\right)$.

This method significantly reduces computation time. The reduction comes from several factors: 1) Calculations of the value functions and the choice probabilities can be conducted outside of the maximization routine. Thus, parameter search costs very little time. 2) Value functions depend on parameter draws but not households'

[18] Recall, N is the number of parameter draws specified by the econometrician. In the estimation, I use $N = 5000$.

observable purchase choices. Therefore, I can parallel compute all value functions for different parameter draws. Once this process is completed, I can then use the value functions to compute the choice probabilities for each household. 3) Moreover, households are independent from each other. Hence, the set of choice probabilities for all households and all parameter tries can also be computed simultaneously. Allowing for parallel computing in the value function and choice probability calculations significantly decreases computation time.

Ackerberg (2009) proposes to use all parameter tries in the calculation of each household's choice probabilities, i.e. the same set of value function calculations is applied to each household. However, I find that the accuracy of the estimate depends on the number of shared parameter draws across households. That is, estimates become less accurate when a single parameter draw is used for all households. Hence, instead I randomly draw a subset of the value functions for each household and compute its choice probabilities using that subset. Similar to all simulated maximum likelihood methods, the accuracy of the estimates depends on the number of the parameters drawn. In the estimation, I draw 5000 sets of parameters and randomly choose 50 corresponding choice probabilities for each household. I test these choices in Monte Carlo experiments and find that they give accurate results.

2.6 Results

To control for observable household characteristics, recall that I divide households into three different income brackets: households with income below \$10K per capita (low income), households with income between \$10K and \$20K per capita (middle income), and households with income of or above \$20K per capita (high income). This allows me to obtain more precise estimates of consumer behaviors. The model and method can accommodate large variations in observable household characteristics. Here I account for households' income and size; but since households in this data set are largely homogeneous in term of ethnicity I disregard their race information. Below I first present the parameter estimates and then discuss the model fit.

2.6.1 Parameter Estimates

The parameter estimates are presented in tables (2.8) and (**??**). In table (2.8), I present the mean, the standard deviation, and the standard errors of the distributions of the random parameters along with those of the initial inventories. The estimates for each income bracket are listed in a separate column:

Table 2.8: Mean and Standard Deviations of the Distribution of Parameters

	Income Bracket					
	Low Income		Middle Income		High Income	
	mean	std. dev.	mean	std. dev.	mean	std. dev.
marg. util. of consumption	4.3599 (0.1519)	1.2204 (0.0651)	3.3021 (0.1464)	1.2180 (0.0460)	3.2864 (0.1516)	1.1499 (0.0374)
linear cost of inv.	1.4666 (0.1265)	1.0979 (0.0207)	1.3141 (0.1343)	1.1145 (0.0286)	1.0086 (0.1343)	1.1090 (0.0219)
quadratic cost of inv.	1.5864 (0.1384)	1.1267 (0.0368)	1.3078 (0.1337)	1.0869 (0.0221)	0.9293 (0.1453)	1.1020 (0.0320)
marg. utility of income	1.3968 (0.1572)	1.3394 (0.0594)	1.2279 (0.1568)	1.3081 (0.0554)	0.8768 (0.1520)	1.2854 (0.0520)
utility of promotions	1.5361 (0.1628)	1.3585 (0.0670)	0.9539 (0.1551)	1.2461 (0.0654)	0.9843 (0.1744)	1.2734 (0.1086)
Coke taste par.	2.2073 (0.1422)	1.2509 (0.0586)	2.3350 (0.1515)	1.3029 (0.0519)	3.8073 (0.1492)	1.2294 (0.0443)
Pepsi taste par.	2.5985 (0.1371)	1.2357 (0.0437)	2.3985 (0.1410)	1.2638 (0.0542)	3.1985 (0.1523)	1.2645 (0.0829)
diet taste par.	0.6032 (0.1348)	1.3076 (0.0653)	0.5084 (0.1495)	1.2965 (0.0531)	0.6976 (0.1472)	1.2765 (0.0608)
regular taste par.	1.7035 (0.1376)	1.2706 (0.0564)	1.3120 (0.1501)	1.3365 (0.0636)	1.1035 (0.1621)	1.2822 (0.0678)
mean initial inv. (2L)	1.15 (0.1833)	1.8935 (0.1814)	2.04 (0.2577)	1.9269 (0.1441)	2.96 (0.2955)	2.1556 (0.2268)

Note: The estimates are presented on top and the standard errors are in the parenthesis.

From the above table, the differences between household types are clear. In terms of the marginal utility of income, we see that households' values are around 1; however, poor households value money significantly more than rich households. This result is consistent with decreasing marginal utility of income: As households' income per capita increases, money becomes less important for them. The same pattern holds for the utility of promotions: low income households have a much larger affinity for promotional activities than either of the other two income groups. In terms of the marginal utility of consumption, middle income and high income households are roughly the same with values around 3.3; low income households, on the other hand, value consumption of soft drinks much more. They hold a mean marginal utility at just below 4.4. This implies that holding inventory constant across all income brackets, poor households would consume more soda than either middle or rich households. The cost of inventory has an opposite story. As income increases, households' cost of inventory decreases: It's more costly for poor households to store products at home than for middle income households and similarly it's more costly for middle income households to store products than for rich households.

Households from different income brackets also differ by their persistent brand and diet preferences. High income households value Coke brand drinks more than either middle or low income households, both of whom value Coca-Cola at about the same level around 2.3. The same story holds true for Pepsi as well, although to a lesser degree. While rich households value Coke more than poor households by over 70%, they value Pepsi only about 23% more than their poor counterparts. In terms of preferences for sugar contents, rich households have the strongest preferences for diet drinks (0.6976) and the weakest preferences for regular drinks (1.1035). Poor households value regular soda the most among all households. The average poor household's preference for regular soda is 30% higher than the average middle income household's and 54% more than the average high income household's. Poor households also value diet drinks less than rich households; however, they value it a bit more than middle income households. This conforms to the fact that in the data (see table 2.1) we observe that middle income households have the smallest diet market share compared to the other two groups.

Lastly, the mean of the initial inventory distribution is increasing in income. Poor households, on average store 1.15 bottles of soda at home, middle income households store just about 2 bottles, and rich households just under 3 bottles. There are several factors that influence households' inventory behavior. First, the marginal utility of income: Since poor households have a higher marginal utility of income, they are more sensitive to price changes. This implies that they would want to buy more when a sale is offered. Rich households, on the other hand, are less responsive to price changes; hence sales are less important to them and they would not need to make big purchases when the price is low. Rich households also have lower marginal utilities of consumption, meaning they don't derive as much utility from drinking soda as poorer households. While poor households have higher marginal utilities of consumption and may want to take advantage of sales more, they are also more constrained by their storage, as seen from their estimated cost of inventory. It's much more costly for low income households to store extra bottles at home, presumably because they are more likely to have smaller houses. Hence they cannot have a large quantity stacked at home for future consumption. Rich households, on the other hand, have a low cost of inventory. Hence, they store more drinks at home but consume at a lower rate. Poor households drink more soda, but due to their storage limit cannot store much and hence have to make more frequent purchases. To summarize, many factors influence households' inventory; we see there are many coefficients pointing in different directions and it is not clear exactly which effect is dominant. Below I show the shares of household types:

Recall, there are nine different types of households. Each represents a type of unobserved persistent product preference. The types range from tastes for specific brand-diet combinations to no preferences at all. From table (2.9), we see that the shares of household types differ by income brackets. High income households have a much larger share of diet Coke lovers than low income households. More

Table 2.9: Shares of Each Household Type

	Income Bracket		
	Low Income	Middle Income	High Income
share of diet Coke households	26.77% (0.91%)	27.25% (0.83%)	35.95% (1.28%)
share of regular Coke households	20.69% (0.71%)	17.58% (0.69%)	10.04% (0.69%)
share of Coke only households	7.75% (0.58%)	7.04% (0.48%)	6.48% (0.65%)
share of diet Pepsi households	15.41% (0.54%)	18.31% (0.71%)	20.33% (1.14%)
share of regular Pepsi households	9.55% (0.58%)	13.24% (0.86%)	9.80% (0.64%)
share of Pepsi only households	4.36% (0.53%)	5.74% (0.55%)	4.16% (0.54%)
share of diet only households	7.20% (0.63%)	5.44% (0.59%)	7.28% (0.70%)
share of regular only households	6.77% (0.59%)	5.14% (0.50%)	5.23% (0.68%)
share of neutral households	1.49% (0.19%)	0.25% (0.03%)	0.72% (0.12%)

Note: The estimates are presented on top and the standard errors are in the parenthesis.

specifically, 35.95% of rich households prefer diet Coke, a 9 percentage points lead to poor households. The same is true for diet Pepsi. The share of poor households who like diet Pepsi is much smaller than that of rich households, 15.41% versus 20.33%. For households who only have preferences for diet but not brand, however, there is very little difference between poor households and rich households. For both groups about 7% of households have this type. For middle income households, on the other hand, this type of households constitutes a smaller share.

It's not surprising then to see that low income households prefer regular drinks more than their richer counterparts. To be more specific, 20.69% of poor households prefer regular Coke. Comparing this to rich households' 10.04%, we see that the share of poor regular Coke lovers doubles that of the rich ones. For regular Pepsi, the share of poor households who have affinities for it is roughly the same as that of the rich households; both at around 10%. In contrast, middle income households have the largest shares of regular Pepsi lovers, at about 13%. The shares of households who prefer regular drinks but don't have intrinsic preferences for the brand is larger for low income households than for the other two income brackets: 6.77% of poor households, but only about 5% of middle and high income households, prefer any regular drinks.

As table (2.9) shows, for all income brackets, the shares of households who have no preferences for specific soda at all, called neutral households, are very small. Households of this type are ones who would purchase predominantly store brand items and mix the regular and diet soda purchases. Since a very small percent of households who engage in this behavior are seen the data, it's not surprising that these estimated shares are small. Below I discuss how well the model fits the data.

2.6.2 Model Fit

Taking the observed price paths as given, I simulate households according to the distribution of the parameter estimates. I compare the resulting market shares of different brands and diet types with the corresponding ones from the data. Overall, they match very well. For example, I observe a 63.68% market share for regular drinks for middle income households. In the simulation, the market share is predicted to be about 63.04%, only a 0.64 percentage points difference. Recall that I adopt a simulated maximum likelihood approach instead of a simulated method of moments one. Hence, the observed and predicted market shares don't have to match.

Table 2.10: Observed and Predicted Market Shares

	Low Income		Middle Income		High Income	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
Diet Soda	39.49%	39.08%	36.32%	36.96%	51.55%	51.70%
Regular Soda	60.51%	60.92%	63.68%	63.04%	48.45%	48.31%
Coca Cola	50.13%	50.16%	50.13%	50.16%	54.14%	55.08%
Pepsi Cola	40.13%	39.89%	40.13%	39.89%	41.72%	40.19%
Store Brand	9.74%	9.95%	9.74%	9.95%	4.14%	4.73%

I use the distributions estimated from the model to simulate the households of each type and income bracket. With these simulated households, I analyze the effects of several sugar tax proposals. In the next section, I discuss the effects of the policy changes.

2.7 Policy Study

There are a large variety of taxes in debate. Some are percentage sales taxes and others are flat per ounce taxes. Here, I follow a previous study conducted by Brownell and Frieden (2009) from the Yale Rudd Center for Food Policy and Obesity. They analyze two different taxes. One is a 10% sales tax while the other is a penny per ounce tax. They predict the total decrease in sugary soda consumption assuming a 100% pass-through. That is, they assume the entire burden of the tax will fall on consumers. Here I analyze both taxes for each income bracket assuming four levels of pass-through: 100%, 75%, 50%, and 25%. For each tax, income bracket, and pass-through level, I present and discuss the predicted total decrease in consumption, the estimated total welfare loss, and the projected post-tax market share for each brand.

Furthermore, I compare the effects predicted by this model against two benchmark studies. In the first benchmark study, I ignore households' stockpiling behavior and specify a static demand model that allows for household heterogeneity in unobserved product characteristics. In the second benchmark study, I model households' stockpiling behavior but do not account for their persistent heterogeneous tastes for product characteristics.

2.7.1 Price Elasticities

Long run price elasticities measure consumers' responsiveness to permanent price changes. Hence, their estimates directly influence the predictions of post-tax consumption reductions in sugary soda and the resulting welfare loss. Therefore, I start

Table 2.11: Long Run Price Elasticities

	Long Run Own Price Elasticity
All Regular Drinks	**-0.4730**
Regular Coke	-1.1916
Regular Pepsi	-1.2061
Regular Store Brand	-1.0530
	Long Run Cross Price Elasticity
Diet Coke	0.9760
Diet Pepsi	0.8871
Diet Store Brand	0.8749

the discussion of the policy analyses by talking about price elasticities. In the following table, I present the long run own price elasticity of all regular soda. I also present the long run own price elasticity estimates of regular soda for each brand and the long run cross price elasticities of diet soda for each brand:

From this table, we see that the long run own price elasticity of regular soda is -0.4730. This means if the price of regular soda increases by 1% its demand will fall by only 0.47%. In another words, the demand of regular soda is inelastic. This implies that households will not likely switch to diet sodas after permanent price increases in regular ones. In contrast, the price elasticity estimate used in Brownell and Frieden (2009) is a -1.2. That is, if the prices of all regular sodas increase by 1%, the demand will fall by 1.2%. This indicates a much more elastic demand. As I show in the benchmark studies section, the price elasticity estimated using a static demand model is similar to the one used in Brownell and Frieden (2009). These estimates do not take consumers' stockpiling behavior into account and hence are overestimated. Once intertemporal substitution is properly treated, the own price elasticity dramatically decreases.

In the above table, for each brand I report the long run own price elasticity of regular soda. The demand for each brand is elastic: For example, the long run own price elasticity of regular Coke is -1.19. That is, if the price of regular Coke increases by 1% while holding all other prices constant, then the demand of regular Coke will fall by 1.19%. This indicates that, on average, consumers are somewhat likely to substitute out of regular Coke and into other products if the price of regular Coke goes up. In comparison, regular Pepsi is a bit more elastic than regular Coke but regular store brand is less elastic than regular Coke.

Moreover, I also report the long run cross price elasticity of demand for diet sodas when the price of the regular one of the same brand increases. The long run cross price elasticity measures the responsiveness of the demand for a good to a change in the price of another good. Here, we see that the demand of diet Coke increases by 0.98% if the price of regular Coke increases by 1%. In contrast, the demand of diet Pepsi and diet store brand would not increase as much if their corresponding regular drinks experience price increases.

The important number to take away here is the long run price elasticity of demand for all regular sodas. Since its estimate is -0.47, we should not see a large decrease in the consumption of all regular drinks post tax. In the next section, I present and discuss the predicted consumption reduction of each tax proposal for all income groups and at all four pass-through levels.

2.7.2 Reduction in Consumption

Table (12) below shows the estimated average reduction in households' consumption of regular soft drinks for both the 10% sales tax and the penny per ounce tax. In monetary terms, the 10% sales tax ranges from around 9 cents to about 17 cents per 2 liters. The penny per ounce tax, on the other hand, is independent of the price and is just under 68 cents for every 2 liters of regular soda.

For each tax and pass-through combination, I report the average estimated total reduction in consumption for each household income bracket. I also report the decomposition of these reductions. I decompose the total reduction in consumption into two parts: 1) substitution from regular to diet sodas and 2) substitution from regular sodas to no purchase (or outside option).

Table 2.12: Post-Tax Reduction in Consumption of Both Taxes for All Pass-Through Levels and Income Brackets

	10% Sales Tax			Penny per Ounce Tax		
	100% Pass-Through			100% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
consump. reduction	-3.28%	-3.05%	-1.79%	-12.92%	-12.31%	-6.13%
sub. to diet	71.58%	69.94%	80.84%	73.15%	71.53%	83.03%
sub. to no purchase	28.42%	30.06%	19.16%	26.85%	28.47%	16.97%
	75% Pass-Through			75% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
consump. reduction	-2.36%	-2.26%	-1.30%	-10.17%	-9.64%	-4.39%
sub. to diet	74.56%	65.73%	77.52%	71.84%	71.80%	82.92%
sub. to no purchase	25.44%	34.27%	22.48%	28.16%	28.20%	17.08%
	50% Pass-Through			50% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
consump. reduction	-1.68%	-1.38%	-0.79%	-6.15%	-6.00%	-2.57%
sub. to diet	72.87%	65.41%	77.37%	71.78%	71.07%	82.85%
sub. to no purchase	27.13%	34.59%	22.63%	28.22%	28.22%	17.15%
	25% Pass-Through			25% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
consump. reduction	-0.77%	-0.71%	-0.37%	-2.46%	-2.46%	-0.35%
sub. to diet	71.56%	67.07%	76.39%	70.88%	67.77%	81.61%
sub. to no purchase	28.44%	32.93%	23.61%	29.12%	32.23%	18.39%

For all taxes at all pass-through levels, poor households experience the largest reduction in consumption while rich households experience the least. Recall from table (2.8), poor households have the highest preferences for regular drinks and rich households have the lowest. This seems to be a contradiction: Since low income households have stronger preferences for regular soda, one might expect to see a smaller decrease in consumption of regular soda. However, since households' taste preferences are not the only factors that influence their consumption behavior, the result here is not a contradiction.

Recall, poor households have the highest marginal utility of income. That is, they react most strongly to price increases. Without their intrinsic tastes, we would expect them to decrease their consumption by the highest margin among all household income brackets. Their intrinsic taste for regular drinks decreases that margin but is not enough to offset the entire burden of the increase price. Rich households, on the other hand, have the lowest marginal utility of income and the highest share of diet-loving households (63.56%). Not only do they have smaller responsiveness to price changes, most of these households already consume a large quantity of diet drinks. Therefore, the tax does not impact them as much. We see the reduction in sugar-sweetened soft drinks is the highest for low income households and the lowest for high income households.

For the 10% sales tax at 100% pass-through level, the largest predicted decrease in consumption of regular soda is 3.28%. According to Brownell and Frieden (2009) however, the expected decrease in the consumption is 7.9%. This more than doubles the estimates I obtained from my model. For the penny per ounce tax at 100% pass-through, the largest predicted decrease in consumption of regular soda is 12.92%. This tax imposes about 68 cents for every 2-liter bottle of regular soda. This is an aggressive tax, considering a bottle of regular soda is around 1.5 dollars. Thus, its impact is also large compared to the 10% sales tax. For this tax proposal, Brownell and Frieden (2009) provides a somewhat vague estimate. They predict the reduction in consumption, as a result of this tax, is above 10%. My estimates conform to this finding.

For each of the taxes and pass-through levels, I decompose the reduction in regular soda consumption into two parts. The reduction comes from substitutions to diet drinks or substitution to no purchase (or outside good). The share of both of these factors across all household income brackets, taxes, and pass-through levels are roughly the same. The largest share of the reduction, around 70%, is replaced by diet drinks. The rest is replaced by no purchases (or outside options). In the following section, I discuss consumers' welfare loss as a result of the proposed taxes.

2.7.3 Welfare Loss

I use compensating variation to measure households' welfare loss. Compensating variation can be interpreted as the dollar amount the government needs to pay to compensate a household in order for the household to reach its pre-tax level of utility. In the table (13) below, I show the compensating variation for each income bracket for both taxes and all pass-through levels. I further decompose the compensating variation into the dollar amount paid in taxes, compensation for utility loss in consumption decreases, and compensation for product switching. Note that the sum of the decomposition does not equal the total compensating variation. I do not present the compensating variation associated with different advertising schemes or the cost associated with inventory costs.

For the 10% sales tax, the overwhelming portion of compensating variation comes from the tax paid. For example, for poor households, the tax makes 89.85% of the total compensating variation at 100% pass-through. This result is intuitive. Since households' demand for regular soda is not elastic, they do not switch to diet drinks very much. Hence, they do not incur large deadweight losses. Indeed, we see that the non-tax payment portion of the compensating variation is only about 12 cents for poor households at 100% pass-through. The highest compensating variation comes from the high income households. Recall, these households switch to other products the least. The reason that they have to be compensated the most is because they are the least price sensitive. They don't tend to wait for sales and purchase products from more expensive stores. Therefore, in a percentage tax, they end up paying more into the tax.

Indeed, from the penny per ounce tax, we see that rich households have to be compensated the least among all households. The difference between these two taxes is that the penny per ounce tax is independent of the prices. Hence, since rich households switch to diet drinks less than their poor counterparts, they are compensated less. Poor households change their behavior more and hence have to be compensated for the deadweight loss created. Even so, the largest share of the compensating variation still comes from the taxes paid. Overall, this implies that the taxes are largely efficient in that they don't create large deadweight losses in the process. Hence, they are well suited for tax revenue collection purposes.

Table 2.13: Compensating Variation Both Tax Proposals for All Pass-Through Levels and Income Brackets

	10% Sales Tax			Penny per Ounce Tax		
	100% Pass-Through			100% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
Avg. Comp. Variation	$1.38	$1.34	$1.46	$7.98	$7.16	$5.76
- tax paid	$1.24	$1.23	$1.39	$4.48	$4.51	$3.83
- change in cons. util.	$0.04	$0.03	$0.02	$1.04	$0.87	$0.65
- change in brand-diet util.	$0.08	$0.06	$0.04	$2.38	$1.72	$1.22
	75% Pass-Through			75% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
Avg. Comp. Variation	$1.06	$1.05	$1.12	$6.54	$5.99	$4.67
- tax paid	$0.96	$0.96	$1.06	$3.93	$3.93	$3.22
- change in cons. util.	$0.03	$0.02	$0.01	$0.83	$0.69	$0.49
- change in brand-diet util.	$0.06	$0.05	$0.03	$1.76	$1.34	$0.93
	50% Pass-Through			50% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
Avg. Comp. Variation	$0.73	$0.71	$0.75	$4.76	$4.41	$3.40
- tax paid	$0.65	$0.66	$0.72	$3.06	$3.03	$2.38
- change in cons. util.	$0.02	$0.01	$0.01	$0.56	$0.44	$0.33
- change in brand-diet util.	$0.04	$0.04	$0.01	$1.12	$0.91	$0.65
	25% Pass-Through			25% Pass-Through		
	Low Income	Middle Income	High Income	Low Income	Middle Income	High Income
Avg. Comp. Variation	$0.38	$0.37	$0.39	$2.67	$2.43	$1.81
- tax paid	$0.34	$0.33	$0.37	$1.76	$1.74	$1.32
- change in cons. util.	$0.01	$0.01	$0.01	$0.32	$0.22	$0.16
- change in brand-diet util.	$0.02	$0.02	$0.01	$0.56	$0.44	$0.31

Table 2.14: Tax as a Percentage of Total Spending

10% Sales Tax at 100% Pass-Through			
	Low Income	Middle Income	High Income
$\frac{\text{(total tax paid)}}{\text{(total price paid)}} \cdot 100$	5.41%	5.52%	4.41%

It's important to note that while rich households pay more into the tax in absolute terms, they pay less in proportion to the total dollar amount spent on soft drinks. As an example, I show the relative percentage for the 10% sales tax at 100% pass-through in the following table. The same holds for all other tax and pass-through combinations.

This table shows that poor households pay more into the tax as a percentage of their total soda consumption than rich households. On average, 5.41% of the total dollar amount spent on soda goes to pay taxes for low income households. In contrast, for rich households only 4.41% of the total dollar amount is spent on taxes. This difference can be explained by the fact that rich households have a large share of diet soda consumers. Therefore, on average, the taxes paid on regular sodas take a smaller share of the total amount paid.

In combination, the two previous tables show that the taxes proposed do not create large deadweight losses, but are regressive. Most of the compensating variation comes from tax payment. However, poor households pay more in tax than rich households in proportion to their respective total soda consumptions.

2.7.4 Post Tax Market Shares

Lastly, in this subsection I briefly discuss the post-tax market shares for different brands. In the table below I present the market share of each brand, with and without tax for each income bracket. Here I again use the 10% sales tax with a 100% pass-through level as an example. The information shown in this table is representative for all other taxes and pass-through levels.

Table 2.15: Market Shares by Brand (10% Sales Tax)

Low Income Households		
	no tax	100% pass-through
Coke market share	50.16%	49.65%
Pepsi market share	39.89%	39.54%
store brand market share	9.95%	10.81%
Middle Income Households		
	no tax	100% pass-through
Coke market share	47.21%	46.79%
Pepsi market share	42.13%	42.01%
store brand market share	10.68%	11.20%
High Income Households		
	no tax	100% pass-through
Coke market share	55.08%	54.98%
Pepsi market share	40.19%	40.05%
store brand market share	4.73%	4.97%

It's clear from the above table that the proposed taxes have little impact on the market shares for each brand. Each brand still occupies mostly its pre-tax share in the market. Store brands do see an increase in market share; but the gains are modest, less than 1 percentage point. The intuition behind this is that even when households substitute, they do so within brands, i.e. households mostly switch from the regular Coke to diet Coke but not to diet Pepsi or diet store brand. In other words, households have strong brand preferences. In the following subsection, I discuss the two benchmark studies conducted and report my findings.

2.7.5 Benchmark Studies

The rationale for my model is that it is important to account for two aspects of soft drinks: storability and differentiation. Soda is a storable good. Previous literature on storable goods, such as Hendel and Nevo (2006 A&B) and Erdem, Imai, and Keane (2003), has shown that a dynamic demand model of inventory is necessary to account for intertemporal substitution. Soda is also a differentiated product. From papers like Berry, Levinsohn, and Pakes (1995) and Gowrisankaran & Rysman (2009), we know that to obtain realistic substitution patterns for differentiated products it is important to incorporate consumer heterogeneity in the demand system. In this subsection, I show how each of these aspects influences consumption change predictions and welfare loss estimates.

I conduct two benchmark studies. In the first benchmark study below, I estimate the demand of soda using a static model. This model accounts for consumer heterogeneity but ignores households' stockpiling behavior. I compare the results obtained by my dynamic model and the ones predicted by this model to show how ignoring dynamics affects the predicted post tax consumption change. In the second benchmark study, I estimate a dynamic demand model. This model accounts for intertemporal substitution but ignores persistent consumer heterogeneity. I use this model to show how heterogeneity affects estimates of consumer welfare loss.

Benchmark I - Static Model with Consumer Heterogeneity

In this benchmark study, I specify a static demand model that allows for consumer heterogeneity. Households in this model do not accumulate inventory. They maximize their per-period utility with respect to the amount and the product they want to consume. Notice that since stockpiling is not accounted for, households consume the exact amount they purchase; nothing is left at the end of the period for future consumption. Households in this model have heterogeneous preferences for product tastes. I maintain the structure on consumer heterogeneity as before. That is, I allow for household specific brand and diet random coefficients. This specification of the static model enables me to isolate the effect of ignoring inventory. The mathematical formulation of the model is as follows:

$$
\begin{aligned}
& \max_{\{s^h, i^h\}} U\left(s^h|\Theta^h\right) + U^h_{si}\left(p_{si}, a_{si}, i^h_{bd}, \varepsilon^h_{si}|\Theta^h\right) \\
= & \max_{\{s^h, i^h\}} \alpha^h \log(s^h) - \gamma^h_1 p_{si} + \gamma^h_2 a_{si} - \sum_b \sum_d 1^h_{-b,-d}\left(\zeta^h_b + \psi^h_d\right) + \varepsilon^h_{si} \qquad (2.14)
\end{aligned}
$$

Since consumption is equivalent to the volume purchased in this model, I use s^h to denote consumption of household h. This model does not include inventories and hence does not require solving a dynamic programming problem. This makes estimation much easier allowing the use of a standard fixed point approach. Nonetheless, I again use the importance sampling method proposed in Ackerberg (2009); hence the method used here is similar to that used in estimating the main model.

Using the estimates of this model, I calculate the price elasticity of sugar-sweetened soft drinks and again simulate households for each income bracket. I measure the effects of a 10% sales tax with 100% pass-through level on consumers' sugary soda consumption.

In line with previous literature, I find that failing to account for inventory behavior overestimates the price elasticity of demand. More specifically, using the static model specified above, I estimate that the price elasticity of demand for sugary soda is about -1.06. This is much closer to the elasticity used in Brownell and Frieden (2009), -1.2. Recall, the long run own price elasticity of demand estimated using the dynamic demand model is around -0.47. This implies that ignoring dynamics in modeling demand for soda dramatically overestimates the long run own price elasticity. This result is intuitive. Since static models do not allow for intertemporal substitution, any increase in demand during a temporary price reduction is attributed to increases in immediate consumption instead of stockpiling. Using a dynamic model, it becomes clear that a large portion of the additional purchase goes to inventories. Thus if the price decrease becomes permanent, households would not react as strongly to it. This implies that the price elasticity of demand for soda is overestimated if a static model is adopted.

Long run price elasticities have significant implications for predicting the reduction in post-tax soda consumptions. Overestimated price elasticities will lead to overestimated reduction in soda consumption post tax. In the table below, I compare

Table 2.16: Predicted Reduction in Consumption

	Low Income	Middle Income	High Income
Static Demand Model	-5.81%	-5.35%	-3.29%
Dynamic Demand Model	-3.28%	-3.05%	-1.79%

the reduction in sugary soda consumption predicted by the static demand against that of the dynamic one for the 10% sale tax with 100% pass-through:This table shows that the static demand model overestimates the reduction in consumption for all income brackets by roughly 75%. For example, under the dynamic demand model, poor households are estimated to reduce their sugary soda consumption by 3.28%. However, under the static model, the reduction is about 5.81%, a 77% overestimation. Recall, in the study conducted in Brownell and Frieden (2009), the predicted reduction in consumption of sugar sweetened soda under the 10% sales tax is 7.8%. The estimate obtained in the static model much more conforms to this estimate.

Not accounting for households' inventory behavior overestimates the reduction in soda consumption post tax. Below, I discuss the impact of not including persistent household preferences in the dynamic demand model of storable goods.

Benchmark II - Dynamic Model without Persistent Consumer Heterogeneity

In this benchmark study, I specify a dynamic demand model of storable goods[19]. This model allows for heterogeneity as a function of observable household characteristics but not as persistent household specific product tastes. More specifically, households differ by their income group, but do not have persistent brand and diet preferences. Similar to my model, households may engage in stockpiling. Their objective is to maximize the sum of discounted expected future flow utility with respect to their current consumptions and current purchases. Mathematically, households' objective

[19] This model is structured after Hendel and Nevo (2006B).

can be expressed as:

$$
\begin{aligned}
V\left(\phi_1^h\right) &= \max_{\{c^h,s^h,i^h\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[U\left(c_t^h, \upsilon_t^h|\Theta\right) - C\left(inv_{t+1}^h|\Theta\right) + U_{si}^h\left(p_{si}, a_{si}, i_{bd}^h, \varepsilon_{si}^h|\Theta\right)\right] \\
&= \max_{\{c^h,s^h,i^h\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[\alpha \log(c_t^h + \upsilon_t^h) - \beta_1 inv_{t+1} - \beta_2 \left(inv_{t+1}\right)^2 \right. \qquad (2.15) \\
&\qquad \left. -\gamma_1 p_{tsi} + \gamma_2 a_{tsi} - \xi_i^h + \varepsilon_{tsi}^h|\phi_1\right]
\end{aligned}
$$

The main difference between this model and mine lies in households' persistent brand and diet preferences. In this benchmark model, households are not allowed to have intrinsic tastes for specific products. For example, household h cannot have high unobserved preferences for regular Coke. Hence, when there are sales for other products, he would make purchases instead of waiting for regular Coke to go on sale. In contrast, in my model he is not likely to react to those sales and instead waits for deals on his preferred product. This implies that the benchmark model might overstate households' switching behavior. Also, since households don't have persistent brand and diet tastes, their switching cost would be underestimated. That is, when household h switches from his favorite drink to something else, his disutility of not consuming his favorite product is not accounted for.

I estimate this model using a maximum likelihood estimator. Since there are no random coefficients in this model, no integrations over taste parameters are necessary. I broadly follow the three step estimation technique introduced in Hendel and Nevo (2006B). Since there are no persistent random components in this model, parameters associated with the product purchase choice and the size purchase choice can be estimated separately. In the first stage, I estimate the product purchase choice parameters using a conditional multinomial logit. These parameters are then used in the second stage to calculate the inclusive values and their transitions. In the last stage, I estimate the size purchase choice and recover parameters associated

Table 2.17: Predicted Reduction in Consumption

	Low Income	Middle Income	High Income
Benchmark Model	-3.94%	-3.86%	-2.15%
Main Model	-3.28%	-3.05%	-1.79%

Note: Main model refers to the model with persistent consumer heterogeneity developed in this paper.

Table 2.18: Average Compensating Variation

	Low Income	Middle Income	High Income
Benchmark Model	$0.97	$0.94	$1.09
Main Model	$1.38	$1.34	$1.46

Note: Main model refers to the model with persistent consumer heterogeneity developed in this paper.

with consumption and inventory. To calculate the dynamic programing problem, I again use a policy function iteration approach instead of parametric value function iteration.

Using the estimates of this model, I again predict the effects of the policy changes. In the table below, I compare the reduction in consumption as a result of a 10% sale tax at 100% pass-through between the two models:From the above table, we see that the benchmark model overestimates the predicted reduction in sugary soda consumption post tax. For example, the estimated reduction in consumption from the demand model with persistent consumer heterogeneity (main model) for low income households is 3.28%. The one estimated using the model without persistent consumer heterogeneity is 3.94%, a 20% overestimation. This result is intuitive: Since households in the benchmark model do not have intrinsic preferences for specific products, they are more prone to switching to diet sodas after the tax. Therefore, not accounting for households' persistent brand and diet tastes overestimates the reduction in the consumption of sugar-sweetened soft drinks. In the next table, I present the compensating variation estimated using both models:

This table shows that the average compensating variation estimated using the benchmark study is underestimated. For example, for the 10% sales tax, the estimated compensating variation for low income households is $0.97 under the benchmark model but is \$1.38 under the model with persistent household heterogeneity. That is, the benchmark study underestimated the average compensating variation by just below 30%. This difference comes from the fact the benchmark model does not allow households to have persistent brand or diet preferences. Hence, they don't incur the disutility associated with not consuming their preferred product. Therefore, the benchmark model overestimates the reduction in consumption while underestimating the compensating variation.

Together these two benchmark studies show that a dynamic demand model of storable goods with persistent household-specific product tastes is necessary to obtain accurate estimates for welfare analyses of sugar taxes. Lacking dynamics in the model results in dramatically overestimated reductions in sugary soda consumption; not accounting for consumers' persistent brand and diet preferences leads to overestimated consumption reductions and underestimated compensating variation.

2.8 Concluding Remarks

In this paper, I specify and estimate a structural dynamic demand model of storable goods. I flexibly incorporate persistent consumer heterogeneity into this model as household specific brand and diet random coefficients. To mitigate the computation burden, I adapt Ackerberg (2009) and develop a computationally attractive method for estimating the model. The model and method developed here are readily applicable to many studies involving storable goods. Here I employ them to analyze the distributional impact of sugar taxes.

To accurately analyze the effects of these policies, I take two specific aspects of soft drinks into account: Storability and differentiation. I compare the results from my model against two benchmark studies: a static model that ignores intertemporal substitution but allows for consumer heterogeneity and a dynamic model that allows households to stockpile but does not incorporate persistent household heterogeneous tastes. I find that failing to account for dynamics results in overestimated reduction in consumption and failing to account for persistent heterogeneous preferences results in overestimated consumption reduction and underestimated welfare loss.

I use the estimated distribution of household preference to perform policy analyses. Following a study conducted by the Yale Rudd Center for Food Policy and Obesity, I analyze two specific tax proposals: a 10% sales tax and a penny per ounce tax. I divide households into three income brackets. For each of them, I calculate the welfare effects of the sugar taxes at four pass-through levels: 25%, 50%, 75%, and 100%. I show that the reduction in sugared soda consumption is at most half of what was previously predicted. Moreover, I find that the policies generate a small deadweight loss but tax poor households more than their rich counterparts.

This paper neglected the supply side of the market. I assume prices evolve according to an exogenous first order Markov process and conduct policy analyses assuming four levels of arbitrary pass-through levels. In future work, I plan to use the demand model developed here to examine firms' pricing strategies, especially in terms of the profit maximizing pass-through level.

3

Incidence of Taxes on Retail Soft Drinks: Theory & Evidence from Scanner Data

3.1 Introduction

With obesity and diabetes numbers on the rise, soft drinks have recently become more and more the focus of health politics. The US senate, the District of Columbia, and twelve states have drafted, and in some cases passed, a variety of sugar tax proposals meant to decrease Americans' consumption of sugary soda. However, it is not clear how effective such measures can be. For instance, if the tax is not high enough then consumers might not react strongly to it. In addition calories saved from soft drinks might be replaced by other unhealthy dietary choices. In this paper we are looking at a third problem: It is not clear yet how much of a soda tax would be passed through to consumers and how much would be borne by producers and retailers.

This question is relevant because a low pass-through would mean that consumer choices and therefore consumer behavior don't change very much. Thus, they would require different ways of influencing the national health.

We find that generally speaking pass-through levels highly depend on market specifics and that in certain situations the whole tax is passed through while in other situations virtually no tax is passed through. In still other situations a small portion of the tax is passed on to consumers. The goal of this paper is to identify the factors that determine the level of pass-through. This would allow us to make more accurate predictions on how sales taxes influence consumer behavior.

The rest of the paper will proceed as follows. The next section provides a detailed description of the datasets used along with relevant summary statistics and institutional details. In section 3, we present our empirical findings along with theoretical models and their predictions. In section 4 we conclude.

3.2 Data

In this paper, we use make use of three sets of data. The primary data set is a weekly scanner data provided by Information Resources Inc. (IRI) from 2001 to 2007. The two auxiliary data sets come from ImpacTeen, a policy research partner of Healthier Youth Group, and the census QuickFacts.

3.2.1 IRI Data

The scanner data provides a panel of weekly retail sales information on all carbonated beverages for each grocery store in the sample. The IRI sample contains over 2,800 stores the across continental United States. For each and every item sold in each week and each store, we observe information on its total revenue collected, total units sold, and any promotional activities. We use measurements on the revenue and units sold to construct a proxy for the unit price.

There are close to 15,000 unique soda items offered across different locations and different time periods. Products differ in their brand, their taste, and the container's size and shape. IRI provides product identifier information that allow us to know which brand each item is, what flavor each item have, and what kind of container each item is bottled in. Thus, we can narrow products specification for easier and more accurate comparison of prices. For example, in one specification, we choose to concentrate on 2 liter bottled Classic Coke.

For each store, we observe when it starts being included in the data and when it stops being included. Furthermore, for each store we observe the state in which it is located. We use this location information to control for time variation in price changes. That is, for a state that experiences an introduction or removal of a soda tax, we control for the natural changes in prices by using the pricing information in a state that has similar demographical characteristics but did not experience such changes.

3.2.2 ImpacTeen Data

Soda tax laws vary, occasionally with significant changes, over time and over states. This information is generally obtainable by going through past state tax regulations as most states keep a record of this information. However, most states do not have easily accessible tax information. The ImpacTeen program keeps a record of all taxes related to soft drinks and salty snacks. Here we have available tax rates across all US states from 2001 to 2007, which covers all years of the store panel data we have from IRI.

This data set shows, for each of the seven years, what the general sales tax of each state is; whether that state has any tax on food products; if the state has a general food tax, what is its tax rate; whether that state taxes soft drinks; and if soda is being taxed, then what its rate is. Out of the 50 states and the district of Colombia, 37 states have experienced a soda tax at least in one of the years. Out of these 37 states, 22 have a "disadvantaged" soda tax. That is, in these 22 states, soda is taxed at a higher rate than general food items. For instance, in Ohio, soda is not listed in the definition of food; hence it is taxed as a non-food item.

Out of all the states that have ever had a soda tax between 2001 and 2007, 15 experienced a change in soda tax rates. Most of these tax changes are between 0.5% and 1%. However, there are seven states that have had a change in soda tax of at least 2%. Table 3.1 lists these states and their tax change information.

Table 3.1: Significant Soda Tax by State

State	Tax Rate Before (%)	Tax Rate After (%)	Year Instituted
DC	5.75	0	2002
North Carolina	4.00	0	2002
North Carolina	0	4.5	2004
New Mexico	5.00	0	2005
Ohio	0	5.00	2003
South Carolina	5.00	3.00	2007
Utah	4.75	2.75	2007
Wyoming	4.00	0	2007

In this paper, we make use of three of these state tax changes: changes in the DC soda tax from 2001 to 2002, changes in the North Carolina soda tax from 2003 to 2004, and changes in the Ohio soda tax from 2002 to 2003.

3.2.3 Census Data

The last set of data we use is the Census QuickFact data. We use this data to ensure that any two states that are being compared against each other have similar demographical characteristics. Since we use states that do no experience soda tax changes as a control for natural price changes over the years, it is important that these "control" states are not too different from the "experiment" states.

3.3 Empirical Analysis

We focus on the determinants of the pass-through level of soda taxes. Currently, we conduct two sets of analyses. In the first set, we study how the size of the stores influences the level of pass through, where the size of a store is represented by its capacity. We set up a basic model of firm behavior with differing capacity constraints and compare the results predicted by this simple theory model to the empirical pass-through seen in the data. In the second set of analysis, we examine how the location of a supermarket influences its pass-through level. We again set up a theoretical model of firm behavior with consumers of differing travel cost and compare the results predicted by this model to empirical results seen in the data. In all empirical analyses, we employ the differences-in-differences method to control for any possible time trends. Each of these analyses are described in detail in the follow sections.

3.3.1 Large vs. Small Stores

We set up a simple model with of monopolistic firms facing a capacity constraint. The capacity constraint is used to represent firms of different sizes. For large, big-box supermarkets the constraint does not bind; while the smaller, more local, stores are assumed to be capacity constrained. Each is assumed to be a monopolist in its market. This capacity can be thought of as the available inventory space of the store. Here, we use the most fundamental demand function and derive pass-through level by profit maximization. While this model is far too simplistic for the complex reality of the US supermarket industry, it does shed some light on the pass-through and its determinants.

Assume we have a monopolist with constant marginal cost c and capacity k. Furthermore, the monopolist faces demand: $q = \frac{a-p}{b}$. Or, $p = a - b * q$. The goal of the monopolist, as always, is to maximize its profit. Its objective function is:

$$\max_q \pi = (a - bq - c)\, q$$

Recall that in this setting, maximizing with respect to quantity is equivalent to maximizing with respect to price. Following standard procedures, we obtain the first order condition, $a - 2bq - c = 0$. Solving the first order condition for quantity, we obtain the standard result:

$$q = \frac{a-c}{2b}$$

Hence, the optimal quantities and prices, depending on capacity constraint, can be written as:

$$q^* = \begin{cases} k & \text{if } k < \frac{a-c}{2b} \\ \frac{a-c}{2b} & \text{otherwise} \end{cases} \quad \text{and } p^* = \begin{cases} a-bk & \text{if } k < \frac{a-c}{2b} \\ \frac{a+c}{2} & \text{otherwise} \end{cases}$$

Now assume an ad valorem tax, t, is introduced. The demand becomes $q = \frac{a-p(1+t)}{b}$. Or, $p = \frac{a-bq}{(1+t)}$. The new optimal quantities and prices are:

$$q^*_{\text{tax}} = \begin{cases} k & \text{if } k < \frac{a-(1+t)c}{2b} \\ \frac{a-(1+t)c}{2b} & \text{otherwise} \end{cases} \quad \text{and } p^*_{\text{tax}} = \begin{cases} \frac{a-bk}{1+t} & \text{if } k < \frac{a-(1+t)c}{2b} \\ \frac{a+(1+t)c}{2(1+t)} & \text{otherwise} \end{cases}$$

To determine the pass through level, we simply compare the optimal prices before and after tax for each of the two types of monopolists. Assume the firm is capacity constrained, that is, $k < \frac{a-(1+t)c}{2b}$. Then comparing the prices paid by consumers under p^*_{tax} with p^* we see:

$$\begin{aligned} (1+t)p^*_{\text{tax}} - p^* &= (1+t)\frac{a-bk}{1+t} - (a-bk) \\ &= 0 \end{aligned}$$

Therefore, under this setting, the capacity constrained supermarket will have a pass-through level of 0%. It will absorb all of incidence of the tax and pass no tax burden onto the consumers. On the other hand, assume the firm is not capacity constrained, $k \geqslant \frac{a-c}{2b}$. And again comparing the prices paid by consumers under p^*_{tax} with p^*:

$$
\begin{aligned}
(1+t)p^*_{\text{tax}} - p^* &= (1+t)\frac{a+(1+t)c}{2(1+t)} - \frac{a+c}{2} \\
&= \frac{a+(1+t)c}{2} - \frac{a+c}{2} = \frac{tc}{2} \\
&> 0
\end{aligned}
$$

We see the non-capacity constrained firm will not absorb all of the burden of the tax. A portion, $\frac{c}{2}$, will be passed onto the consumers. While we don't know the marginal cost, c, it is clear that the price difference is strictly positive.

Finally, consider a medium-sized firm that is capacity constrained without tax but not with tax (i.e. $\frac{a-(1+t)c}{2b} \leqslant k < \frac{a-c}{2b}$). In this case we get:

$$
\begin{aligned}
(1+t)p^*_{\text{tax}} - p^* &= (1+t)\frac{a+(1+t)c}{2(1+t)} - (a-bk) \\
&= \frac{(1+t)c-a}{2} + bk
\end{aligned}
$$

The last expression is greater than or equal to zero since $k \geqslant \frac{a-(1+t)c}{2b}$; it is also less than $\frac{ct}{2}$ since $k < \frac{a-c}{2b}$. Thus, the pass-through for medium-sized stores is between the one for large stores and the one for small stores.

These results are fairly intuitive. A store facing a binding capacity constraint cannot supply the optimal unconstrained quantity. In order to maximize profit it charges a price that allows it to get as close to the unconstrained optimum as possible, i.e. use its whole capacity. Since the capacity does not change when the tax is introduced the stores the store has to bear the whole burden of the tax in order to keep selling its whole capacity.

For large stores in this set up the capacity constraints do not bind. Therefore, all else equal, their profit maximizing quantities sold are larger than those of small stores. After a tax is introduced, the tax acts as an additional marginal cost. The large monopolist reacts to the increased marginal cost by increasing its price and hence passes some of the tax burden onto the consumers.

From this simple model, one point can be clearly shown: The pass-through level does not always and only depend on demand elasticities. We observe that under this fundamental setup, the capacity constrained firm passes zeros percent of the tax onto consumers and the non-capacity constrained firm passes a share $\frac{c}{2}$ of the tax burden on to consumers. The pass through level does not depend on demand parameters, a and b, in either case. Hence, it is not always clear that firms' pass-through decision depends purely on the demand elasticities.

We also find support for the predictions of the model in the data. We use the North Carolina and South Carolina data from 2003 to 2004. The two states have comparable demographical characteristics. In 2004, North Carolina changed its taxing scheme on soft drinks. In the two years before, soft drinks were not taxed. However, in 2004 the state instituted a 4.5% sales tax on all soda. In comparison, no tax changes were experienced by South Carolina, having a consistent 5% sales tax on all soda.

Since there are no pre-defined size of stores and no information of the square-footage of each supermarket are given, we have to construct measurements for the size of the store. A large store is defined here as a store that sells more than 300,000 liters of soft drinks per year. In this definition, the small stores account for the 25% of the sample. We are also conducting robustness checks on this definition as it is always somewhat arbitrary.

For straight-forward comparison of prices, we use only 2-liter bottles of Coke. Similar results seem to hold for other products. We employ the differences-in-differences technique in obtaining the pass-through estimate. As North Carolina is comparable to South Carolina, this allows us to tease out any time trends in the data. The regression results are presented below. Table 3.2 shows the price change as a reaction to the tax increase (coefficient on time and state interaction) of small stores and table 3.3 shows the pass-through of large stores.

Table 3.2: Small Store Diff-n-Diff Regression Results:

	coefficient	standard error	t-statistic	P-value
constant term	$1.24	0.0113	110.42	0.000
state dummy	$0.07	0.0154	4.62	0.000
time dummy	$0.07	0.0159	4.56	0.000
state-time interaction	-$0.07	0.0346	-2.00	0.047

Table 3.3: Large Store Diff-n-Diff Regression Results:

	coefficient	standard error	t-statistic	P-value
constant term	$1.17	0.0040	289.12	0.000
state dummy	$0.01	0.0051	2.03	0.042
time dummy	$0.73	0.0057	12.98	0.000
state-time interaction	-$0.01	0.0072	-1.64	0.100

From table 3.2, we see that the when a 4.5% sales tax is introduced in North Carolina, the price of soda in small stores decrease by seven cents (a pass-through of -1.33%). This estimate is statistically significant at a 95% confidence level. Note, the average price of this product is $1.24. According to our theory, in order to continue selling its original capacity, small stores would have to reduce its price such that the product costs consumers exactly the same as before the tax. In this case, the average price charge to consumers, before tax, would have to be around $1.18. This means a predicted price decrease of 6 cents. This is, surprisingly, very close to the observed price decrease (1 cent lower) in the data.

From table 3.3, we see that the when a 4.5% sales tax is introduced in North Carolina, the price of soda in large stores decreases by one cent (a pass-through of 87.84%). This estimate, however, is not statistically significant at a 95% confidence level. According to our theoretical model, large store will let consumer bear some burden of the tax. However, since we don't observe their marginal cost, we cannot determine the level of predicted pass-through exactly. From our empirical result, it seems that in large stores, most of the burden of the tax will fall on consumers.

We conclude from this piece that capacity-constrained stores will bear all of the burden of the tax while non-capacity constrained stores will pass through a strictly positive portion of the tax to consumers. This prediction is supported by the data. Moreover, we also show using this simple setup that demand elasticities are not the sole determinant of pass through levels. In the following subsection, we study how the geographical location influences store's tax pass through level.

3.3.2 Rural vs. Urban Locations

A second possible determinant of tax pass-through levels could be the geographical location of a store. It is conceivable that due to the higher concentration of consumers in more urban locations, supermarkets too are built within closer vicinity of each other. As supermarkets become more concentrated competition then becomes fiercer. In more rural areas, consumers live in more spread out communities. This increases travel cost between stores and lessens competition. Assuming this intuition is correct, we develop a theoretical model with consumer travel cost. As previously, we then compare the predictions of this model to results found in our empirical analysis.

We now develop a model that allows us to capture differences in density of stores and consumers. Consider a population of consumers with unit demand. The consumers live in a linear Hotelling-type city. There are two retailers and two manufacturers (Coke and Pepsi). Both retailers sell both goods. They compete in a Bertrand-type game, i.e. they compete in prices. Retailer r pays price p^m_{br} for one unit of item b. This price is determined by Nash Bargaining between the store and the manufacturer.

Consumers face the decision which of the goods to buy (if any) and at which retailer. They make this decision not only based on prices but also on travel costs:

A consumer might buy at a more expensive store if that one is closer. Consumer i chooses his purchase such that he maximizes his utility function:

$$U_{ibr} = \alpha_b - \beta p_{br} + \epsilon_{ibr} - \gamma |l_i - l_r|$$

α_b is a brand fixed effect; β is consumers' marginal utility of income and p_{br} is the price of brand b in store r. γ is the travel cost, and l_i and l_r respectively are the locations of consumer i and retailer r. Finally, ϵ_{ibr} is an idiosyncratic shock which is distributed normal with mean zero and standard deviation σ.

Given locations and manufacturer prices, retailer r maximizes his profit with respect to retail prices p_{br}. The profit is given by:

$$\pi_r = \sum_b (p_{br} - c_{br} - p^m_{br})\, s_{br} M$$

c_{br} is the marginal cost to retailer r associated with selling an item of brand b; we assume constant marginal costs so that c_{br} does not depend on sale quantitities. Thus, $(p_{br} - c_{br} - p^m_{br})$ is the surplus r earns per unit of product b. M is the size of the market and s_{br} is the market share of a given retailer-brand combination, so that $s_{br}M$ captures retailer r's total sales of brand b.

Producer b's profit is:

$$\pi_b = \sum_r (p^m_{br} - c^m_b)\, s_{br} M$$

Again, $s_{br}M$ are the total sales of brand b at store $r.c^m_b$ is producer b's marginal cost of production; again we assume constant marginal cost, so that $(p^m_{br} - c^m_b)$ is the producer's surplus per piece.

The producer's price, p^m_{br}, is determined by Nash bargaining between store r and manufacturer b. Here we are assuming that each bargaining process is independent from the other ones; i.e. retailers and manufacturers accurately predict their profits if bargaining succeeds as well as if it fails.

This model is extremely hard to solve analytically since there are so many layers of optimization involved. Therefore we solve the game numerically in order to compare the theoretical predictions with the empirical results discussed below. For our simulations we use a setting with two retailers and the following parameters:

Table 3.4: Simulation Parameters

	Coke		Pepsi
Retailer's marginal cost			
c_{b1}	0.2		0.2
c_{b2}	0.1		0.1
Manufacturer's marginal cost			
c_b^m	0.3		0.1
Demand parameters			
α_b	2.0		1.7
β		0.9	
σ		1	
γ		0.5	
M		1000	

Since we want to compare rural and urban settings, we calculate results for city lengths of 1 and 2, respectively. In the former case we assume that the stores are located at $l_1 = 0.25$ and $l_2 = 0.75$, in the latter case we assume $l_1 = 0.5$ and $l_2 = 1.5$. Consumers are in both cases distributed uniformly.

We report our simulation results in table 3.5. The table shows that pass-through levels are higher in rural settings than in urban ones. This is a result that gets confirmed in the data as is shown below.

Table 3.5: Simulation Results for Different Cases

		No Tax	5% Tax	Pass-Through
Urban	Retail 1 - Coke	\$2.59	\$2.56	75.68%
	Retail 1 - Pepsi	\$2.92	\$2.87	64.04%
	Retail 2 - Coke	\$2.86	\$2.80	55.94%
	Retail 2 - Pepsi	\$3.13	\$3.08	66.45%
		No Tax	5% Tax	Pass-Through
Rural	Retail 1 - Coke	\$2.33	\$2.34	104.51%
	Retail 1 - Pepsi	\$2.92	\$2.91	92.81%
	Retail 2 - Coke	\$2.86	\$2.83	77.97%
	Retail 2 - Pepsi	\$3.24	\$3.22	87.04%

To compare any tax pass through differences between an urban and a rural location, we need not only two such places but also 1) each location experiences significant tax changes and 2) each location has a demographically comparable place that has no tax changes. The first requirement is obvious. The unit prices of soft drinks are small. Therefore, if we observe only minor tax changes, less than 2%, we would not be likely to observe any significant pass through. The second requirement comes from the fact that we use differences-in-differences method to control for any unobservable time-varying trends.

The set of locations that satisfy the first criterion is listed in table 3.1. Out of those seven possible states, we choose DC as the representative of the urban location and North Carolina as the representative of the rural location. DC experienced a dramatic change in soda tax in 2002, when the sales tax on soda went from 5.75% to 0%. In the same period, however, places on the DC-Maryland border experienced no tax changes. All soda in these locations were being taxed at the usual 5%. The population in these two places are very comparible and thus making DC is good candidate for analysis. The same holds for North Carolina. As the previous subsection points out, after two years of no taxation on soda North Carolina instituted a 4.5% sales tax in 2004. In comparison, no tax changes were experienced by South Carolina, which has had a consistent 5% sales tax on all soda. The two states share comparable demographical attributes.

The following table 3.6 documents diff-n-diff regression results of each of the locations. Similarly to before, to ease the comparison of prices, we use 2-liter Coke products in these regressions.

As we can see, in DC area, after the 5.75% soda tax is removed, prices increased by four cents. Since this estimate is statistically significant at a 95% confidence level, it shows that there is some indication that the pass through level is not 100% and stores are absorbing some of the burden of the tax. On the other hand, in North

Table 3.6: Diff-n-Diff Regression Results for Different States:

		coefficient	standard error	t-statistic	P-value
	constant term	$1.26	0.0107	117.62	0.000
DC:	state dummy	$0.01	0.0126	0.73	0.466
	time dummy	-$0.01	0.0117	-0.58	0.561
	state-time interaction	$0.04	0.0143	2.47	0.014
		coefficient	standard error	t-statistic	P-value
	constant term	$1.18	0.0038	305.86	0.000
NC:	state dummy	$0.00	0.0049	0.53	0.593
	time dummy	$0.07	0.0054	13.31	0.000
	state-time interaction	-$0.01	0.0069	-1.88	0.060

Carolina, after the introduction of a 4.5% soda tax, no significant changes in prices can be found. This implies that consumers bear most of the burden of the tax by paying the original prices in addition to the tax.

This analysis indicates that there is some evidence to suggest the competition in urban areas drive stores to absorb some of the burdens of the tax and hence have lower pass through rates.

3.4 Conclusion

In this paper we study the pass-through of taxation levied on soft drinks. To do so we empirically analyse price changes when major changes in the sales tax on soda occur. We find that small stores exhibit a smaller pass-through than their larger counterparts. This is consistent with a simple theoretical model in which capacity constrained retailers shift less of the tax burden to consumers than those stores for which the capacity constraint is not binding. We also observe different pass-through levels between stores in Washington, DC and in North Carolina. We think this might be due to the differences of an urban versus a rural setting; results from a theoretical framework support this interpretation, but more work has to be done to exclude alternative explanations. Overall we find that pass-through levels are highly dependent on the specifics of the setting. In future work we will investigate the major influences in more detail.

Bibliography

Ackerberg, D. (2009), "A new use of importance sampling to reduce computational burden in simulation estimation," *Quantitative Marketing and Economics*, 7, 343–376.

Arcidiacono, P. and Miller, R. (2007), "CCP estimation of dynamic discrete choice models with unobserved heterogeneity," *mimeo*.

Bajari, P. and Benkard, C. (2005), "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach," *Journal of Political Economy*, 113.

Benítez-Silva, H., Hall, G., Hitsch, G., Pauletto, G., Brook, S., and Rust, J. (2000), "A comparison of discrete and parametric approximation methods for continuous-state dynamic programming problems," *mimeo*.

Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Brownell, K. and Frieden, T. (2009), "Ounces of preventionthe public policy case for taxes on sugared beverages," *New England Journal of Medicine*, 360, 1805–1808.

Brownell, K., Farley, T., Willett, W., Popkin, B., Chaloupka, F., Thompson, J., and Ludwig, D. (2009), "The public health and economic benefits of taxing sugar-sweetened beverages," *New England Journal of Medicine*, 361, 1599–1605.

Byrd, S. (2004), "Civil Rights and the Twinkie Tax: The 900-Pound Gorilla in the War on Obesity," *Louisiana Law Rev.*, 65, 303.

Conlon, C. (2010), "A Dynamic Model of Costs and Margins in the LCD TV Industry," *mimeo*.

Erdem, T., Imai, S., and Keane, M. (2003), "Consumer Price and Promotion Expectations: Capturing Consumer Brand and Quality Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5–64.

Finkelstein, E., Trogdon, J., Cohen, J., and Dietz, W. (2009), "Annual medical spending attributable to obesity: payer-and service-specific estimates," *Health Affairs*, 28, w822.

Goettler, R. and Gordon, B. (2008), "Durable goods oligopoly with innovation: theory and empirics," *mimeo*.

Gostin, L. (2007), "Law as a tool to facilitate healthier lifestyles and prevent obesity," *JAMA: the journal of the American Medical Association*, 297, 87.

Gowrisankaran, G. and Rysman, M. (2009), "Dynamics of consumer demand for new durable goods," .

Hartmann, W. (2006), "Intertemporal effects of consumption and their implications for demand elasticity estimates," *Quantitative Marketing and Economics*, 4, 325–349.

Hartmann, W. and Nair, H. (2010), "Retail Competition and the Dynamics of Demand for Tied Goods," *Marketing Science*, 29, 366–386.

Hendel, I. and Nevo, A. (2006a), "Measuring the implications of sales and consumer inventory behavior," *Econometrica*, 74, 1637–1673.

Hendel, I. and Nevo, A. (2006b), "Sales and Consumer Inventory," *The RAND Journal of Economics*, 37, 543–561.

Jacobson, M. and Brownell, K. (2000), "Small taxes on soft drinks and snack foods to promote health," *American Journal of Public Health*, 90, 854.

Johnson, R., Appel, L., Brands, M., Howard, B., Lefevre, M., Lustig, R., Sacks, F., Steffen, L., Wylie-Rosett, J., et al. (2009), "Dietary sugars intake and cardiovascular health: a scientific statement from the American Heart Association," *Circulation*, 120, 1011.

Kaplan, M. (2010), "Taxing sugar-sweetened beverages." *The New England journal of medicine*, 362, 368.

Mello, M., Studdert, D., and Brennan, T. (2006), "Obesitythe new frontier of public health law," *New England Journal of Medicine*, 354, 2601–2610.

Nevo, A. (2000), "Mergers with differentiated products: The case of the ready-to-eat cereal industry," *The RAND Journal of Economics*, pp. 395–421.

Powell, L., Slater, S., Mirtcheva, D., Bao, Y., and Chaloupka, F. (2007), "Food store availability and neighborhood characteristics in the United States," *Preventive Medicine*, 44, 189–195.

Rust, J. (1987), "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, pp. 999–1033.

Biography

Yucai Wang, more often known as Emily, was born in Changchun, China on February 1, 1983. She moved to the United State in 1998 and started to attend school in Tucson, Arizona. She earned a double degree in mathematics and economics, magna cum laude, from the University of Arizona in 2006. Her undergraduate studies were made possible with funding from the UA International Student Scholarship, which covered all her school expenses throughout the four years of college.

Emily later attended Duke University where she earned an M.A. in economics in 2008 and a Ph.D. in economics in 2010. She completed her Ph.D. with funding from the Economics Department Doctoral Fellowship from fall 2006 to spring 2007 and James B. Duke Summer Research Fellowship in the summers of 2009 and 2010. Emily will marry Christoph Bauner in the summer of 2011 and then will begin her appointment as Assistant Professor of Economics at the University of Massachusetts, Amherst, Isenberg School of Management.