ABSTRACT

Title of Dissertation: Information, Consumer Choice, and Firm Strategy in an Experience Good Market

Yan Chen, Doctor of Philosophy, 2008

Directed By: Professor Ginger Jin, Department of Economics

This paper models how consumers make brand choice when they have limited information. In an experience good market with frequent product entry and exit, consumers face two types of information problems: first, they have limited information about product existence; second, even if they know a product exists, they do not have full information about its quality until they purchase and consume the product. In this paper, I incorporate purchase experience and brand advertising as two sources of information, and examine how consumers utilize them in a dynamic process.

Specifically, to address the awareness problem, I model the consumer choice set as a function of experience and advertising, which varies across consumers and evolves over time. In terms of quality, I allow a first-time consumer to infer product quality from advertising. Once she buys the product, she learns the quality perfectly. To better capture the dynamics, I incorporate habit formation conditional on each consumer's purchase history.

The model is estimated using the AC Nielsen homescan data in Los Angeles, which records grocery shopping histories for 1,402 households over six years. Taking ready-to-eat cereal as an example, I find that consumers learn about new products quickly and form strong habits. More specifically, advertising has a significant effect informing consumers of product existence and signaling product quality. However, advertising's prestige effect is not significant. I also find that incorporating limited information about product existence leads to larger estimates of the price elasticity. Then I use instrument variables based on differentiated-products firm competition models to address the endogeneity problem of price and advertising with unobserved brand characteristics. Based on the IV estimates, I summarize the substitution pattern and simulate consumer choices under counterfactual experiments to evaluate a number of brand marketing strategies and a policy on banning children-oriented cereal advertising.

Information, Consumer Choice, and Firm Strategy in an Experience Good Market

By

Yan Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Ginger Jin, Chair
Professor John Rust
Professor Roger Betancourt
Professor Erkut Ozbay
Professor P.K.Kannan

Dedication

To my parents.

Acknowledgements

My first and greatest debt is to my advisor, Professor Ginger Jin. Ginger not only gives me guidance in choosing the topic, obtaining the data, and conducting the research, but also treats me like a sister in life.

I am grateful to Professor John Rust, Professor Roger Betancourt, Professor Erkut Ozbay, Professor Dan Vincent, and Professor P.K.Kannan for their comments and suggestions during the process this dissertation was written.

I have been blessed to meet many good friends at the University of Maryland and back at the University of British Columbia. Thanks for making the past few years of my life peaceful and offering many happy distractions from study and research.

As always, I have counted on the support and encouragement of my parents and my brother. Thanks for having confidence in me and showering me with love all these years. Although we are in different corners of the world, our hearts have always been together and will always be. 谢谢你们，我爱你们。

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction, Literature, and Data Summary

Section 1.1. Introduction

In an experience good market with frequent product entry and exit, consumers face at least two informational problems when they make product choices: on the one hand, they have limited information about product existence; on the other hand, even if consumers know a product exists, they are uncertain about its quality before consumption. Consumer choice in such an environment involves a dynamic process of information acquisition, about which a large body of literature has developed. However, empirical research typically deals with only one of the informational problems. To my knowledge, this paper is the first attempt to incorporate both kinds of informational problems in a structural framework and evaluate their relative importance using a rich panel of household purchase data. In this sense, the paper brings the literature closer to the reality and sheds new lights on both consumer demand and firm marketing strategies.

In this paper I focus on two mechanisms that alleviate the informational problems: consumption experience and brand advertising. First, consumers learn about product quality through experience. In particular, I assume that product quality can be fully learned after the first try. This type of one-period learning is a limiting case of Bayesian learning (the prior belief is updated to the true value after one experience) and is suitable for goods whose characteristics can be ascertained quickly after use. Second, advertising has two types of informative effects: signaling product quality (Nelson (1970, 1974), Kihlstrom & Riordan (1984), Milgrom & Roberts (1986), etc), and affecting the probability that a consumer is aware of the product (Butters (1977), Grossman & Shapiro

(1984)). In my empirical model, the former is captured in the utility function with advertising interacted with a new brand [1] dummy, while the latter is captured by advertising entering the choice set formulation.

Specifically, to address the awareness problem, I model the choice set as a function of brand advertising and purchase experience. As a result, choice sets vary across consumers and evolve over time. Previous research usually assumes a fixed choice set for all consumers, which is reasonable if the market is in steady state equilibrium with a few products. However, in many markets for experience goods, due to the large number of products available and the high rate of new product introduction, one is unlikely to be aware of all products for sale and has to restrict attention to a subset of products. This subset forms the consumer's choice set, which evolves over time as information accumulates through advertising and experience.

Advertising and experience can also directly affect a consumer's level of utility. On the one hand, advertising can directly provide utility and have a prestige effect on all consumers (Stigler & Becker (1977), Becker & Murphy (1993)). For example, a consumer may obtain a higher level of satisfaction from wearing a pair of Nike sneakers than from wearing an unknown brand of similar quality because of the image established by Nike commercials. On the other hand, experience may also directly change utility through habit formation if previous purchases of a product increase its current utility (and choice probability). Allowing for habit formation in the choice process has important implications for firm marketing strategy. For example, if consumers are habituated to a product, then the introductory price of a new product may need to be set lower than when there is only learning to warrant a product switch.

[1] New brand in this paper means new to a specific consumer, not newly introduced into the market.

The consumer choice model is estimated on a panel of AC Nielsen homescan data in the ready-to-eat (henceforth RTE) cereal market. A survey of 1,402 demographically balanced U.S. households in the Los Angeles market, the homescan data keeps track of on-going household purchasing of grocery products from December 1997 to December 2003. It records both the transaction information such as brand, price and quantity and the household demographic information. In the estimation, the homescan data is supplemented by two data sets on the supply side: manufacturer advertising data obtained from TNS Media Intelligence Company, and brand nutritional data collected from internet.

Empirical specifications with different information assumptions are estimated in the paper. In the benchmark specification, consumers are assumed to have full information about brand quality and existence. Then I introduce limited information about brand quality. Consumers are aware of all brands available in the market but are uncertain about the quality of unused brands. They form expectation about brand characteristics and infer quality from advertising. Lastly, consumers are assumed to have limited information about both brand quality and existence. Their choice sets are heterogeneous depending on consumption history and brand advertising.

In addition to tests of how different specifications fit the data, estimation results from different specifications are compared in terms of market share predictions. It is shown that the model specification with both limited information on brand existence and brand quality best captures consumer behavior. This model incorporates three sources of consumer heterogeneity: choice sets, tastes, and experiences. The results suggest that limited information about brand existence has a large impact on consumer behavior: price sensitivity is much higher among the set of brands that consumers are aware of, implying that demand is more elastic. I also distinguish different effects of advertising in the estimation and find that advertising has a significant effect in providing information about brand existence and signaling brand quality. However, the prestige effect of advertising is not significant. The results imply that advertising works for new consumers only. After the first experience with a brand, advertising will not change consumers' utility level and what matters to the consumers is the purchase history.

Instrumental variables selected on basis of differentiated-products competition models are applied to the limited information model to obtain consistent coefficient estimates. The IV estimation does not change the qualitative results of the model. A Hausman test of the IV coefficients and OLS coefficients also suggests that they are not systematically different. Finally, some policy experiments are conducted to evaluate a number of firm marketing strategies in pricing and advertising and to simulate the effect of banning children-oriented cereal advertising. The consumer-level limited information model offers new insights for pricing strategies and provides a micro foundation to evaluate the pulsing strategy in advertising in the RTE cereal market, which supplements the study of Dubé, Hitsch, and Manchanda (2005). The model can also be used to evaluate nutritional policy changes. For example, in recent years child obesity has become a serious problem in America and there have been renewed debates on whether food advertising targeted toward children should be ban. Using the model I can estimate the effects of such a ban on the nutritional intakes and expenditures of various demographic groups, which may provide some empirical evidence in the policy debate.

The rest of this paper is organized as follows. The remaining sections of chapter 1 review the literature and discuss the industry background and data sets used for this study. Chapter 2 specifies and estimates the demand system under different informational assumptions. Instrumental variables are selected in Chapter 3 and the the IV estimation algorithm is discussed and the results presented. Policy experiments are conducted in Chapter 4 on a number of brand marketing strategies.

Section 1.2. Related Literature

The paper is related to several lines of literature. The consumer learning literature addresses the problem of limited information about product quality. In their pioneer work, Erdem and Keane (1996) estimate how consumers learn about the cleaning power of laundry detergents. Both experience and advertising give consumers noisy signals about the detergent's quality and consumers update their beliefs about quality in a Bayesian way. Following this study, there have been many studies that model consumer learning in a Bayesian framework in various markets (Ackerberg (2003), Crawford and Shum (2005), Chintagunta, Jiang, and Jin (2007), to name a few). However, the consumer learning literature usually takes the consumer choice set as homogenous. It does not account for the fact that different consumers may be exposed to different sets of products due to limited awareness, which is the central research question in the literature on heterogeneous choice set (also called consideration set in the marketing literature).

There have been very few economic studies that consider heterogeneity in the choice set. Goeree (2008) presents a model in which advertising influences the set of products from which consumers choose to purchase. Specifically, the probability that a consumer is informed of a product is a function of the effectiveness of the product's advertising and the observed consumer characteristics. In the marketing literature, there are relatively more papers allowing for heterogeneity in consideration set. Brand choice is usually modeled as a two-stage process: at the first stage consumers identify a subset of brands which constitute their consideration set. They then choose the brand with the highest utility at the second stage. Roberts and Lattin (1997) review the theoretical and empirical marketing studies that develop an individual level model of consideration set and analyze how marketing mix affects consideration set and consumer choice including Andrews and Srinivasan (1995) and Allenby and Ginter (1995). They also point out some directions for future research including dynamics in consideration set, which is captured in this paper. Swait (2001) assumes that the probability a specific consideration set is formed is a function of the expected maximum utility from the alternatives in that set. Mehta, Rajiv, and Srinivasan (2003) formulate the process of consideration set formation as a trade-off between the expected benefit from including an additional brand and the additional search cost incurred. Eliaz and Spiegler (2007) study a market model in which firms use irrelevant alternatives to influence consumers' consideration set. All these studies, however, only model one-time purchase in a static setting and do not account for variation in choices and choice sets over time.

In terms of how to model advertising, this paper learns from both theoretical (as discussed above) and empirical literature on advertising (Ackerberg (2001, 2003), Anand and Sharchar (2005), etc). This paper also benefits from the insights of the literature on RTE cereal market that involves demand estimation (Hausman (1996), Nevo (2001), Shum (2004), Hitsch (2006), etc). Due to the richness of my data, compared to the previous studies, I am able to include more dynamics in consumer choice, identify consumer learning from habit formation based on the difference in choice dependence structure of new and old consumers as in Osbourne (2006), and distinguish different effects of advertising.

Last but not least, this paper is an extension of the literature on analyzing demand systems in differentiated product markets (Berry (1994), Berry, Levinsohn and Pakes (1995, 2004), etc). With household level data, the parameters that vary with individual households can be identified without any constraints on the distribution of unobserved brand characteristics. The parameters that do not vary with individuals such as the mean price coefficient need to be estimated with the market share data and instrumental variables. This paper applies the estimation method to a limited information environment.

Section 1.3. Industry Background and Data Description

1.3.1. Features of RTE Cereal Market

Readers can refer to Section 2 of Nevo (2001) for a more complete picture of the RTE cereal industry. For the purpose of this paper, there are several features of the RTE cereal market that make it a good case study for the empirical work. First, cereal is an experience good, the attributes of which are not completely known before consumption. Second, brand entry and exit happen frequently in the RTE cereal market and none of the national brands has a truly dominant hold on the market, which imposes a considerable informational burden on consumers.

As an example, Table 1 shows the entries and exits of RTE cereal brands[2] from December 1997 to December 2003 in the Los Angeles market. In the 6-year period, a total of 62 (46) brands enter (exit)[3] the market, which accounts for about 47.3% (35.1%) of the total number of brands existing at the end of 1997. Column 2 of Table 2 displays sales-based market shares of major brands from December 1997 to December 2003. Due to the existence of a large number of brands, in the table I select the top 50 brands (together they account for about 79% of the market) and combine the remaining ones into a composite brand, brand 51. The biggest brand (Brand 1[4]) has a market share of 6% while most brands take up less than 1% of the market.

Third, the RTE cereal market is heavily advertised. Advertising over sales ratio for RTE cereal was 13% in 2001. For well-established brands, the ratio was 18%[5]. In comparison, the average ad over sales ratio across 200 industries was 3.2%[6]. Heavy advertising shows that firms believe advertising is effective in promoting sales, thus it is important to analyze its effects in this market.

1.3. 2. Data Description

On the consumer side, I utilize AC Nielsen Homescan data on RTE cereal products from December 1997 to December 2003. Tracking 1,402 demographically balanced households in Los Angeles, the homescan data ties consumer purchase behavior with key demographic measures. Homescan panelists scan items at home from each shopping trip,

[2] Brand definition follows the classification on each manufacturer's website. Different box sizes are treated as the same brand, while extensions of a brand name are distinct brands. For example, Cheerios, Honey Nut Cheerios, and Berry Burst Cheerios are three different brands.
[3] Brand entry and exit are defined using the AC Nielsen homescan data. A brand entry is observed if the first transaction of the brand occurs after June 1998. A brand exit is observed if the last transaction of the brand occurs before June 2003.
[4] Brand names are not revealed due to a confidential agreement with the data provider.
[5] See Nevo (2001), page 311.
[6] See Advertising Age, March 1, 2006

recording price and quantity purchased as well as the age, income, and other demographic information of the shopper. The data set keeps track of on-going purchasing from the same household over time, hence offers insights into households' consumption habits and dynamics. On average a household stays in the homescan panel for 48 months. Once a household leaves the panel, a new one that is similar in all demographic measures is selected to take its place. Table 3 contains definitions and summary statistics of the key variables in the homescan data.

Using the homescan data I can then summarize the consumption pattern in the RTE cereal market. On average, a household makes 14 shopping trips with RTE cereal per year. The households usually have a couple of brands that they purchase repeatedly over time. Most brands are purchased once and never again (Figure 1 displays a histogram of the number of times a household purchases a brand). After a brand was first purchased by a household, the probability of the household repurchasing the brand is 14.1% on the next shopping trip, 12.9% on the second shopping trip, and about 11% on the following trips (Figure 2 shows the probability of repurchasing a brand on the shopping trips following the first try). Both Figure 1 and Figure 2 suggest that learning in the cereal market is mainly done after one shopping trip. Figure 2 also suggests that conditional on repurchase after the first experience with a new brand, a household exhibits loyalty to that brand.

On the product side, I obtain advertising data from TNS Media Intelligence, which tracks advertising expenditures of cereal manufacturers from January 1999 to December 2003. The advertising data covers 278 cereal brands across 11 different media types[7].

[7] The 11 media types include network TV, cable TV, sport TV, magazines, syndication, national sport radio, network radio, Sunday magazines, local newspaper, outdoor billboard, and national newspaper. In this paper advertising particularly refers to cereal manufacturers' advertising expenditures in these media types. Although retailer advertising such as retailer deal and store featuring is common in the RTE cereal market, it is not included in the estimation due to lack of data on retailers.

The brand advertising expenditures include both national advertising and local advertising. On average, national advertising accounts for 98.1% of the total advertising expenditure and is mainly on network TV and cable TV, while local advertising accounts for 1.9% of the total advertising and is mainly on local newspapers and outdoor billboards. Average monthly advertising expenditures of the top 50 cereal brands in Los Angeles are shown in Column 3 of Table 2.

Nutritional data on cereal brands is collected from www.nutritiondata.com[8], with nutrient information of 111 brands including calories, sugar, dietary fiber, protein, etc. The fiber and sugar contents per serving (30 gram) for the 50 top brands are displayed in Column 4 and Column 5 of Table 2. These two nutrients are selected because there is little variation in other nutrients across brands.

In all data sets, the characteristics of brand 51, the composite brand, are calculated as the average of all non top-50 brands.

[8] The nutritional information was collected on Sep 10, 2006, from the website. There is no variation of nutrients over time for the same brand.

Chapter 2: Demand Estimation without IV

Section 2.1. Demand

2.1.1. Demand Specification

Abstracting from quantity, the empirical model focuses on consumer brand choice conditional on purchasing RTE cereal. Multiple brand purchases on one shopping trip are treated as independent events[9]. For example, if on a shopping trip a consumer purchased brands A, B, and C, it is estimated as if she had made three separate transactions with A, B, and C within the same day. Suppose on the previous shopping trip the consumer purchased brand A. Then in the transaction with brand A, the past choice dummy would be set to 1. In the other two transactions, the past choice dummy would be set to 0. On the next shopping trip, if the consumer purchased any one of brands A, B, and C, her last-time choice dummy would be set to 1. Apart from not estimating the quantity choice, the model also does not consider the store choice or the brand choice conditional on visiting a store as store-level data are not available.

There are a number of consumers (index by i) choosing from a set of brands (indexed by j) on different shopping trips (indexed by t). The brand choice is a two-stage process.

[9] The model does not include non-purchase of RTE cereal as the outside good. There are two reasons for doing so: (i) consumers may choose not to purchase because they have cereal inventory at home, not because the utility of non-purchase is higher than all cereal brands. Treating non-purchase as the outside good, therefore, will bias the parameter estimates downward in the utility function. (ii) Consumers choose not to purchase RTE cereals on about 2/3 of all shopping trips. Including those shopping trips will further add to the already large computation burden. The model also restricts to brand choice but not quantity choice. Taking quantity into consideration requires tracking consumer's stockpiling and inventory, which will greatly complicate the model. About 52% of the purchases are associated with only one brand. Multiple-brand transactions are treated as independent transactions following Shum (2004). Shum (2004) fails to find across–brand synergies in demand patterns of RTE cereals that would require modeling the multiple-brand purchase decision. Readers interested in this can refer to Hendel (1999), Dubé(2004) for examples of multiple-discrete choice model which allows multiple-unit and multiple-brand purchases on one shopping trip, and Nevo and Hendel (2006) for an example of consumer inventory model.

At the first stage, based on previous purchase experience and brand advertising, the consumer is informed of a subset of brands that constitute her choice set on that shopping trip. At the second stage, the consumer chooses a brand from her choice set that maximizes her expected utility. On a specific shopping trip, the consumer's information set includes the quality and characteristics of brands she has purchased before, and prices and advertising intensities of all brands in her current choice set. Note that brands are differentiated both horizontally and vertically. The horizontal differentiation is on brand characteristics such as fiber and sugar contents. The vertical differentiation is on the brand quality. The quality of a cereal brand is determined by the quality of ingredients such as grain and rice, and the processing techniques.

Two assumptions are made about the choice set formulation. First, a brand purchased before would stay in the choice set. In other words, once a consumer tries a brand she never forgets about it, even though she may dislike it and choose not to purchase it again. Second, the probability of consumers being informed of a previously unused brand is a function of the brand's advertising stock. Formally, at time t, the probability that consumer i has choice set C_{it} is:

$$P(C_{it}) = \prod_{j \in C_{it}} q_{ijt} \prod_{k \notin C_{it}} (1 - q_{ikt}) \qquad (1)$$

where q_{ijt} is the probability of consumer i being informed of brand j at time t, and

$$q_{ijt} = \frac{\exp(\varphi_0 + \varphi_1\, adv_{jt} + \varphi_2 adv_{jt} \bullet inc_i + \varphi_3 adv_{jt} \bullet nokid_i + \varphi_4 adv_{jt}^2)}{1 + \exp(\varphi_0 + \varphi_1\, adv_{jt} + \varphi_2 adv_{jt} \bullet inc_i + \varphi_3 adv_{jt} \bullet nokid_i + \varphi_4 adv_{jt}^2)},\ \forall j \notin \mathrm{E}_{it}$$

$$= 1,\ \forall j \in \mathrm{E}_{it} \qquad (2)$$

where E_{it} is consumer i's experience set as of time t, i.e., the set of brands previously purchased by consumer i up to time t. In the estimation, transactions in the first year of a consumer's purchase history are used to initialize her experience set. The variable adv_{jt} is a depreciated stock of advertising expenditures for brand j at time t. Specifically,

$$adv_{jt} = \sum_{\tau=0}^{T} \delta^{\tau} a_{jt-\tau} \qquad (3)$$

where a_{jt} denotes brand j's advertising expenditure at time t,[10] and δ is the discount factor. Using stock instead of current flow of advertising allows advertising to have a lagged effect on consumer choice in the form of good will stock. Specifically, if a brand entered a consumer's choice set on the last shopping trip but was not purchased, the probability of it re-entering the consumer's current choice set may still be high even if the brand is not advertised in the current period due to the lagged effects of previous advertising. The term adv_{jt}^2 is included to account for the potential increasing or decreasing returns to scale of advertising. In equation (2), adv_{jt} is also interacted with household income and whether there is any kid in the household to reflect the heterogeneity in exposure to advertising for different types of households.

At the second stage, consumer i chooses brand j to maximize expected utility conditional on her choice set. As is now standard in the discrete choice literature, the expected utility consumer i obtains from brand j is a function of brand j's characteristics.

$$\begin{aligned} U_{ijt} = {} & E(X_j)\beta_i + \alpha_i price_{ijt} + \rho_i adv_{jt} + \kappa \cdot unused_{ijt} + \lambda_i \cdot unused_{ijt} \cdot adv_{jt} \\ & + pastchoice_{ijt} \bullet \gamma + \eta_{jt} + \varepsilon_{ijt} \end{aligned} \qquad (4)$$

where X_j =[*fiber sugar*]$_j$, β_i=[β_{1i} β_{2i}]$'$, $price_{ijt}$ is the price of brand j when consumer i it at time t. In the AC Nielsen Homescan data, the price of a brand is recorded as the weekly

[10] Advertising data is monthly while purchase data is daily. Therefore advertising expenditure at time t means advertising expenditure in the month that day t belongs to. In the empirical results reported in section 5, δ=0.95 and T=6. I also estimate the model with δ varying from 0.8 to 0.99 and T from 3 to 12. The robustness checks do not yield significant qualitative differences.

average price of that brand in the store where the brand was sold. In the estimation, I subtract the manufacturer coupon value and the retailer deal value from the price if a coupon or a deal is used[11].

Note that β_i, α_i, ρ_i, and λ_i are individual coefficients. Specifically,

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \alpha_i \\ \rho_i \\ \lambda_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \alpha \\ \rho \\ \lambda \end{bmatrix} + \Pi \bullet D_i + \Sigma \bullet v_i \qquad (5)$$

where D_i is a vector of observed household characteristics including household income, age of female household head, and presence of kids, and v_i represents a vector of unobserved household characteristics with standard normal distribution. The variable $unused_{ijt}$ is a dummy which equals 1 if brand j was never purchased by consumer i before time t. It interacts with adv_{jt}, implying that advertising may provide information about the quality of unused brands. For example, the fact that the cereal manufacturer is able to spend a huge amount on promoting a brand may signal to consumers that the manufacturer is in a good financial condition and can therefore product cereals with better ingredients and better technology. The vector $pastchoice_{ijt}$ = [$chosen_{ijt-1}$ $chosen_{ijt-2}$,…, $chosen_{ijt-\tau}$], where $chosen_{ijt-\tau}$ equals 1 if brand j was chosen τ shopping trips before t[12]. The term η_{jt} represents brand j's characteristics that are observed to the consumer but not to the researcher at time t. In the case of RTE cereals, η_{jt} encapsulates packaging,

[11] I am not able to control for coupons and deals systematically as in Nevo and Hendel (2006) as I do not have store level data and do not observe the availability of coupons and deals to consumers.

[12] In the empirical results I use T=6. Compared to previous studies where T is often equal to 1, my results show a more complete picture of time dependence of consumer choices. I also estimate the model with T=12 and the results are similar.

shelf space, etc. Lastly, ε_{ijt} is a mean 0 stochastic term independent across time, brands and consumers.

If brand j has not been purchased before, the consumer holds expectations of its fiber and sugar contents according to the following rule: E($fiber_j$)=mean($fiber_k$), and E($sugar_j$)=mean($sugar_k$) $\forall$ brand k tried by consumer i before and belonging to the same segment as brand j (Following Hausman (1996) and Shum (2004), I divide the brands into family, adult, and kid segments. The segment categorization is shown in column 7 of Table 2). If the brand has been purchased before, then the consumer knows its characteristics.

The utility maximization stage generates $P(j|C_{it})$, the conditional probability that brand j is chosen by consumer i at time t. By the law of conditional probability, multiplying $P(j|C_{it})$ and $P(C_{it})$ yields P_{ijt}, the unconditional probability of consumer i choosing brand j at time t.

$$P_{ijt} = \sum_{C_{it} \in S} \prod_{j \in C_{it}} q_{ijt} \prod_{k \notin C_{it}} (1 - q_{ikt}) P(j \mid C_{it}), \qquad (6)$$

where S is the set of all choice sets that include brand j. Matching the choice probabilities predicted by the model with the observed choices using maximum likelihood will then give us the parameter estimates.

2.1.2. Discussion

Key features of the demand model merits additional discussion. First, the choice set formation process addresses the informational problem about product existence. Even though the choice set is aggregated to contain the 50 biggest national brands and a composite brand, it is still unlikely that consumers would know and compare the utility of all 51 brands on each purchase occasion. Allowing the choice set to depend on consumption experience and brand advertising thus brings the model closer to real consumer behavior. Since the choice set is not observable in the data, I need to simulate them in the estimation. The details of simulation will be discussed in Section 4. Second, consumers learn about brand quality after their first experience with the brand, which captures learning in the RTE cereal market reasonably well as shown in Figure 2. Unlike some complicated products, consumers usually can attain precise knowledge about a cereal after consuming one box of it. Third, compared to most previous choice models where only the past choice is included, choices on the previous 6 shopping trips are included in the utility function. The coefficients on the set of past choice variables provide a better description of the temporal dependence of brand choices than when there is only the last time choice. For example, if a consumer's brand choice history consists of A, B, A, B, …, A, B, and only the last time choice dummy is included, then we would wrongly infer that the consumer only seeks variety and is not subject to habituation. If we extend the model to have additional past choices, then it is possible to better capture the potential for habit formation. The distinction is important because if variety-seeking is dominant, then temporary promotions' impact on demand would be short-lived. On the other hand, if consumers are susceptible to habit formation, temporary promotions may affect sales well into the future. Thus adding more past choice variables not only better describes time dependence, but also helps managers optimize decisions on marketing strategies. Fourth, advertising has three roles in the model: (I) affecting consumer choice set, which represents advertising's informative effect on brand existence and is captured by the φ parameters; (II) signaling quality of an unused brand, which represents advertising's informative effect on brand quality and is captured by the parameter λ; (III) directly providing utility, which represents advertising's prestige effect and is captured by the parameter ρ. Identification of the different effects will be discussed below.

In the demand setup, I assume that the consumer is myopic and maximizes her current period utility. When state dependence (habit formation) is present, a forward-looking consumer would consider the future effects of her current choice. Forward-looking is important in many cases, especially in situations where the experimentation cost is high, examples including choice of durable goods such as computer and digital camera and decision about whether to accept a job offer or to continue job search. It is less critical in this situation where consumers choose a frequently purchased product and the cost of trying a new product is low as consumers can easily switch back to the brands they have been using. Moreover, previous marketing research shows that consumers spend an average of 13 seconds in selecting a brand out of the shelf[13]. This is a very short time for a consumer to plan her future choices. Therefore, I tend to believe that the myopic assumption is reasonable in this application and in the choice process of many other nondurable goods, for example, beverages and cosmetics.

2.1. 3. Identification

Intuitive identification strategies are discussed in this subsection. The parameters to be estimated θ include φ_0, φ_1, φ_2, φ_3, φ_4. β, α, ρ, $\prod$, Σ, κ, λ, and γ. Variation of brand choices corresponding to observed brand characteristics, price, and advertising for all

[13] See Cesar Costantino, Ph.D. Dissertation, Chapter 4, "Gone in Thirteen Seconds: Advertising and Search in the Supermarket", 2004.

consumers is used to identify β, α, and ρ. A RTE cereal may also have attributes that are favored by a subgroup of consumers. For example, older consumers may prefer higher fiber content while kids may prefer higher sugar content. Substitution pattern of consumers with different demographics when brand characteristics vary helps identify $\prod$. And heterogeneity in substitution pattern of consumers with the same demographics helps identify Σ. Comparing the average probability of choosing a used brand with the average probability of choosing an unused brand on each purchase occasion identifies κ. Comparing the repurchase probability after purchase of a new brand with the repurchase probability of a previously purchased brand identifies learning from habit formation, and variation in brand choices over time pins down γ.

In terms of advertising's different effects, the main identification assumption of the prestige effect and informative effects is that the prestige effect does not vary by consumption experiences. As in Ackerberg (2001, 2003), the prestige effect impacts both experienced and inexperienced consumers in the same way, while the informative effects only works on consumers who have never tried the brand before. Therefore, variation in the ratio of choice probability between experienced and inexperienced consumers as advertising intensity changes can be used to identify the informative effect from the prestige effect. The two types of informative effect (coefficients φ versus coefficient λ) both affect the choice probability of inexperienced consumers. An inexperienced consumer may choose to try a brand because advertising alerts her attention to the existence of the brand or because advertising raises the expected quality of this brand. Ignoring advertising's prestige effect for the moment, if advertising's only provides information about brand existence, consumers will include the brand in their choice set with a higher probability if the brand's advertising increases. In this case, advertising does not enter the consumer utility function and hence the marginal effect of advertising on brand choice probability is independent of the observed brand characteristics. If two brands with different characteristics increase advertising by the same percentage, their choice probability will go up by the same percentage. If, furthermore, advertising provides a signal about brand quality, then consumers have two information channels to evaluate a brand --- the advertising signal and the other brand characteristics. They would trade off the information inferred from advertising with the information observed from the brand characteristics. If the quality perception of the brand is already high based on the brand characteristics, the marginal effect of advertising on brand choice probability would be small as there are fewer consumers on the margin who would switch to the brand due to more exposure to advertising. If, on the other hand, the quality perception of the brand is relatively low from the brand characteristics, then the marginal effect of a surge in advertising would be big as more consumers would be convinced to switch. Therefore, the two types of informative effect can be distinguished by whether the marginal effect of advertising on brand choice probability depends on the brand characteristics, as advertising only enters the utility function and interacts with the brand characteristics if the informative effect about brand quality exists.

To see this mathematically, let us consider a simple example: there are two brands in the market, one (brand 1) has been established for a long time and the other (brand 2) was newly introduced. Consumers all know about brand 1, and brand 2 launches an advertising campaign. Ignoring in this example the returns to scale of advertising in choice set formation and heterogeneity in coefficients across households, if advertising's only effect is informing consumers of the existence of brand 2, then the probability that consumers choose brand 2 is:

$$P = \frac{\exp(\varphi_0 + \varphi_1 adv_2)}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} * \frac{\exp(E(X_2)\beta + \alpha * price_2 + \Psi)}{1 + \exp(E(X_2)\beta + \alpha * price_2 + \Psi)} \qquad (7)$$

where ψ denotes the sum of variables in utility function other than price and observed brand characteristics. The marginal effect of advertising on the change in choice probability is:

$$\frac{\partial \ln(P)}{\partial (adv_2)} = \frac{\varphi_1}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} \qquad (8)$$

Note that equation (8) is independent of brand 2's characteristics.

If advertising also provides information about quality, the choice probability of brand 2 is:

$$P = \frac{\exp(\varphi_0 + \varphi_1 adv_2)}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} * \frac{\exp(E(X_2)\beta + \alpha * price_2 + \rho * adv_2 + \Psi)}{1 + \exp(E(X_2)\beta + \alpha * price_2 + \rho * adv_2 + \Psi)} \qquad (9)$$

The marginal effect of advertising on the change in choice probability is:

$$\frac{\partial \ln(P)}{\partial (adv_2)} = \frac{\varphi_1}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} + \frac{\rho}{1 + \exp(E(X_2)\beta + \alpha * price_2 + \rho * adv_2 + \Psi)} \qquad (10)$$

The higher the utility consumers infer from the brand characteristics, the less the need to rely on the information in advertising. Comparing equation (8) and equation (10), we can see that whether marginal effect of advertising on choice probability is dependent on the brand characteristics identifies the informative effect about brand quality from the informative effect about brand existence. To illustrate this point, figure 3 depicts the marginal effect of advertising on choice probability. The non-stochastic part of the utility function other than advertising is denoted by Q. When only the informative effect about existence exists, the marginal change in choice probability is a declining function of advertising expenditure and is independent of Q. When advertising also signals quality, the marginal change in choice probability is not only a declining function of advertising, but also is a function of Q. As Q increases, the marginal change in choice probability decreases.

2.1.4. Estimation Issues

Before I present and discuss the empirical results of demand estimation, it is appropriate to discuss the issues that are encountered in the estimation and how I attempt to deal with them.

(1) Endogeneity

If the manufacturers set up prices and advertising levels according to the consumers' willingness to pay, then the endogeneity problem may arise as price and advertising levels could be correlated with unobserved brand characteristics in the utility function. For example, if the brand manager coordinates media advertising and store promotion activities, then the unobserved brand characteristics such as shelf space or store featuring can be correlated to the price and advertising expenditures of the brand. As a result the coefficients on price and advertising can be overestimated. The best way to deal with the endogeneity problem is to use instrumental variables. However, the notion of valid instruments involves supply side specifications and IV estimation in this non-linear consumer choice model is rather complicated. The focus in this section is on how demand estimation varies with informational assumptions, thus the IV estimation is deferred to a later section. In this part, I only include brand fixed effects to control for unobserved brand characteristics invariant over time. For example, if the government dietary policies promote the health effects of whole grain foods, then the price and advertising levels of the whole grain cereals may be increased. Whether a cereal is a whole grain is invariant over time and is absorbed by brand dummies and therefore this type of endogeneity is controlled for.

(2) Unobserved Consumer Heterogeneity

In a model with lagged dependent variables, state dependence (habit formation) is observationally equivalent to consumer heterogeneity as individual specific effects can lead to persistence in choices. State dependence can be exaggerated if unobserved consumer preferences are mistakenly assumed to be homogeneous. For example, an overweight consumer can have a high preference for a low sugar cereal and repeatedly purchase it. If consumer specific preference is not controlled for, repeated purchases will be captured by the past choice variables and regarded as strong habit. Therefore, it is important to disentangle the true state dependence from consumer heterogeneity. In the estimation I use consumer-brand random effects to control for unobserved consumer heterogeneity. The details of the implementation are provided in Appendix 1.

(3) Choice Set Simulation

To address the informational problem about brand existence I allow for heterogeneity in consumer choice sets. The underlying choice sets over which consumers make utility comparison are unobservable to researchers. Moreover, the number of potential choice sets can be very large — with 51 brands in the market, the number of possible choice sets is 2^{51}. Hence instead of attempting to exhaust all possibilities, I simulate the choice sets. In the simulation, the probability of a brand being included in a consumer's choice set is a function of brand advertising and purchase experience according to equation (2). The details of the choice set simulation process are provided in Appendix 2.

(4) Properties of Estimator

After simulating the choice sets, I can calculate $\hat{P}_{ijt}$, the simulated choice probability of each brand for each household on every purchase occasion, and conduct a maximum simulated likelihood (MSL) estimation. The joint simulated likelihood function is:

$$SL(\theta) = \prod_i \prod_t \hat{p}_{ijt}(\theta)^{Yijt} \qquad (11)$$

where $Y_{ijt}=1$ if consumer i purchases brand j at time t, and $Y_{ijt}=0$ otherwise. The joint simulated log likelihood is:

$$SLL(\theta) = \sum_i \sum_t Y_{ijt} \log(\hat{p}_{ijt}(\theta)) \qquad (12)$$

The MSL estimator $\hat{\theta}$ is a vector of parameters that maximize equation (12). Train (2003) shows that if the number of simulation draws rises faster than square root of sample size, then the MSL estimator is not only consistent but also efficient asymptotically equivalent to the maximum likelihood estimator[14]. Specifically, the MSL estimator is distributed

$$\hat{\theta} \overset{a}{\sim} N(\theta^*, -H^{-1}/N) \qquad (13)$$

where θ^* is the true parameter value, N is sample size, and $-H = -E(\frac{\partial^2 LL(\theta^*)}{\partial\theta\partial\theta'})$ is the information matrix. In practice, I use $\hat{H} = \frac{\partial^2 SLL(\hat{\theta})}{\partial\theta\partial\theta'}$ to approximate the value of H and calculate the estimated variance.

[14] Monte-Carlo studies done by Keane (1994) and Geweke et al. (1994) also suggest that MSL has excellent small sample properties if reasonably good simulators are used.

Section 2.2: Results of Demand Estimation without IV

The demand estimation is validated on the panel data in the Los Angeles RTE cereal market[15]. There are 1,402 households with 69,134 cereal purchases in the LA market from December 1997 to December 2003. The first 12 months of each household's purchase history are used to construct its experience set. As a result, households staying in the homescan panel for less than 12 months are dropped out of the estimation. The unit of observation in the estimation is a transaction, i.e., a household-purchase date-brand combination. Observations with missing values on key estimation variables are dropped out[16]. The regressions start from July 1999 since the earliest advertising data is available in January 1999 and to calculate the advertising stock we need advertising data six months ago. The estimation sample consists of 844 households and 37,858 transactions, which remains unchanged in all specifications. Values of the key variables in the estimation sample are summarized in Table 4. In all specifications 50 brand dummies are used.

To guide the choice of variables, I first run a preliminary regression --- multinomial logit regression with full information. Consumers are assumed to know all brands for sale and also their quality. The coefficient estimates are shown in Table 5. The interaction terms of household demographics and brand characteristics that are not significant are excluded in later regressions. Since the multinomial logit model is subject to independence of irrelevant alternatives and does not capture the realistic substation

[15] The modeling technique and estimation method in this paper are not specific to a particular geographical market or a particular experience good. I can apply the model to environments where consumers face the two types of informational problems, for example, consumer choice of cosmetics, credit cards, health care plans, etc.
[16] The missing values do no happen systematically so I am not concerned with a selection bias.

patterns, the random coefficient logit model is used instead where a random component is added in the coefficients of price, advertising, fiber content, and sugar content, respectively. After the variable selection guided by the multinomial logit, I run three random coefficient logit models with different informational assumptions. First, I assume that consumers have full information about both brand quality and brand existence. The choice set is the same over time and across consumers. The second specification is a regression with learning about quality information, where consumers are assumed to know all brands for sale in the market but not the quality of unused brands. In the third specification, consumers are assumed to have limited information about both the quality of unused brands and the brand existence. A random coefficient logit with quality learning and heterogeneous choice sets is estimated. Note that the three specifications are non-nested. To compare them, ideally I would like to construct a test statistic with a limiting distribution, so that I can select one specification over the other with some confidence level. However, our panel data do not satisfy the distributional assumptions of tests for non-nested models, for example, Vuong (1989) and Chen & Kuan (2002). Therefore, to assess the goodness of fit I use two methods. First I compare the different specifications using the Akaike Information Criterion (AIC) and using a measure of predictive performance developed by Betancourt and Clague (1981). Then I construct a variable which measures market share prediction errors of the three specifications to see how well they predict consumer choices.

2.2.1. Estimation with Full Information

The benchmark specification is a random coefficient logit regression where consumer choice sets include all 51 brands and characteristics of all brands are known. The benchmark model allows me to examine in a simple way how price and advertising affect demand and have a sense of the temporal dependence of consumer choices. The results also serve as a baseline for later comparisons.

The parameter estimates of the benchmark specification are reported in column I of Table 6. Price is negative and significant. The price sensitivity decreases as household income increases and if the household has no kids. On average, advertising's prestige effect is negative and marginally significant. But the prestige effect increases as income grows and when there is no kid in the household. *unused* is negative and significant. If we calculate the odds ratio, we can see that the fact that a brand was never purchased before decreases the brand choice probability by 75%. However, *unused*adv* is positive and significant, suggesting that more advertising signals better quality to inexperienced consumers. The signaling effect diminishes with income and when the household has no kid. All six past choice variables are positive and significant. The coefficient of *chosen_2* is slightly higher than that of *chosen_1*, consistent with the fact that consumers usually switch away from the brand last purchased if the brand was a new try. Both *fiber* and *sugar* are negative and significant. Older consumers without kids prefer more fiber and less sugar.

2.2.2. Estimation with Limited Information about Brand Quality

In the second specification, I run a random coefficient logit regression where all consumers face the same choice set of 51 brands but they do not know the quality of brands not bought before. Consumers form expectation of brand characteristics based on their previous experience with brands in the same segment. They also infer brand quality from advertising and brand quality can be ascertained after one purchase. The signs of many coefficients (column II of Table 5) are the same as those of the benchmark regression, and for most coefficients the magnitudes are comparable. The coefficient on *adv* is still negative but no longer significant. The only coefficient that changes sign is the *fiber* coefficient, but it is not significant.

The similarity of the coefficients (and the log likelihood) to the benchmark suggests that limited information about brand quality does not significantly affect consumer behavior. This is probably due to the nature of the RTE cereal market — the cost of experimenting with an unused brand is low, thus uncertainty about brand quality may not be an important factor when consumers decide which brand to buy.

2.2.3. Estimation with Limited Information about Both Brand Quality and Brand Existence

In the third specification consumers have limited information on both brand quality and brand existence. They still infer quality of unused brands from experience and advertising, but their choice sets are now heterogeneous and vary over time. The probability of having a particular choice set for each consumer on each purchase occasion follows equations (1) and (2), and the choice set is simulated as described in Appendix 2.

The price coefficients (column III of Table 5) suggest that allowing for heterogeneous choice sets increases price sensitivity. The coefficient on price is significantly bigger than in the first two scenarios. To get a sense of how the price coefficient translates into price elasticity, I increase each brand's price by 1% separately and simulate the consumer choices based on the parameter estimates. Consumer choices are then aggregated to calculate the % change in brand market shares resulting from the 1% price change. The values of own price elasticity for the top 10 brands are reported in Table 7. Compared to the previous two specifications, the price elasticity in the current one is much larger. The estimated price elasticities in the third specification are more plausible since their absolute values are all bigger than 1, which is consistent with the fact that profit-maximizing firms should be operating at the elastic part of the demand curve.

When consumers have limited information about brand existence, they are not aware of brands outside their choice set and therefore cannot respond to the price changes of those brands. If we estimate the model as if consumers had full information about brand existence, we are in essence imposing that consumers know the price changes of all brands but choose not to respond to some of them. As a result, the price elasticity is lower in the "full information" case. The price estimate in the third specification suggests that consumers are actually much more sensitive to price changes of the brands that they are aware of. Should the consumers have lower information search costs and know more brands for sale, they would switch more frequently when price cuts are available. Therefore, if the information problem about product existence is alleviated, the market should be more competitive as consumers would be more responsive to price variations.

In the utility function, the coefficient on *adv* is negative but not significant, implying that advertising's prestige effect is not important. The coefficient on *unused*adv* is positive and insignificant, suggesting that advertising's informative effect on brand quality is not significant. In the choice set formation, φ_1 (coefficients on *adv* in equation (2)) is positive and significant while φ_4 (coefficients on adv^2 in equation (2)) is negative and significant. Advertising raises the probability that consumers are informed of the brand, but this effect exhibits decreasing returns to scale. The coefficient on *adv*inc*, φ_2, is negative and significant, suggesting that the informative effect of advertising on brand existence decreases with household income. In contrast, the coefficient on *unused*adv*inc* in the utility function is positive and significant, suggesting the informative effect of advertising on brand quality increases with household income. This makes sense if richer consumers have higher opportunity cost of time and watch less TV commercials, but once they are alerted to the availability of an unused brand, they rely more on advertising to obtain the quality information than other methods of searching. The coefficient on adv*nokid in choic set formation, φ_3, is positive but not significant, implying that the effect of advertising does not vary with the presence of kids. Figure 4 plots the probability of a brand entering a consumer's choice set against the brand's advertising expenditure evaluated at the mean level of household income and presence of kids. At the mean of advertising stock ($3.22 million), the probability of a brand being included in the choice set is 88%. Increasing advertising stock by $1 million from the mean will result in a probability of 99% that the brand is included in the choice set. What is consistent over the three specifications is that advertising plays a significant role in providing information to consumers but it does not have a significant prestige effect.

The past choice variables are still positive and significant, suggesting that consumers form persistent habit in cereal purchases. Compared to the results obtained without heterogeneous choice sets, the dependence on the past choice variables falls. The smaller coefficients on past choices are consistent with the larger (in absolute value) coefficient on price: consumers are more likely to switch brands in response to price changes when they rely less on previous experience.

2.2.4. Comparison of Goodness of Fit

To compare the goodness of fit of the three specifications, two measures are computed. The first measure is the Akaike Information Criterion (AIC), which equals 2*k-2lnL, where k is the number of parameters and lnL is the log likelihood. The AIC imposes a penalization on more parameters and the smaller the value of AIC, the better the model fit. The AIC for the first specification is 212930, for the second one is 201386, and for the third one is 164516. Hence according to the AIC the third specification fit the data best.

Second, I compute a measure of predictive performance for discrete choice models developed by Betancourt and Clague (1981). The measure is based on the idea of information entropy. It rewards correct predictions when predicted choices are the same as observed choices and penalizes wrong predictions when predicted choices are different from observed choices. Moreover, the summary measure scores each choice prediction by giving it points not only in accordance with whether the prediction is correct but also in a way that reflects the degree of certainty of the prediction[17]. To obtain the measure, we first need to calculate the entropy for an observation in terms of predicted probabilities, $E_{it} = -(\sum_{j=1}^{51} P_{ijt} \log P_{ijt})$. Then the amount of information contained in the predicted probabilities P_{ijt} is defined as $I_{it} = 1 - E_{it} / E_{\max}$, where $E_{\max} = -\frac{1}{51}\log(\frac{1}{51})$ [18] and represents the maximum amount of uncertainty associated with the data distribution. Defining a correct prediction as $P_{ijt} > 1/51$ when brand j is chosen at time t and $P_{ijt} < 1/51$ when it is not chosen, we can calculate the amount of information contained in the sample set of predictions as $\overline{I} = (I_1 - I_2)/N$, where I_1 is the sum of information for all correct predictions, I_2 is the sum of misinformation for all incorrect predictions, and N is the number of observations. The specification with the highest value of $\overline{I}$ predicts the data best[19]. Applying the formula to our data, I obtain that the $\overline{I}$ for the first specification is -11.5, for the second one is -13.2, and for the third one is 0.8[20]. Again the third specification represents the best fit.

Next I construct a variable to check how well the three specifications predict aggregate consumer behavior. Using the parameter estimates, I first predict consumer brand choice on each shopping occasion, which is the brand that generates the highest

[17] For a more detailed discussion of the measure, please refer to Section 4.6 of "Capital Utilization: A Theoretical and Empirical Analysis", Betancourt R. and C. Clague, Cambridge University Press, 1981. The original measure is defined for cross-section data but can be easily extended to panel data. When choice sets are simulated, the probabilities used in the calculation are the mean of simulated probabilities.

[18] The formula is $E_{\max} = -\frac{1}{J}\log(\frac{1}{J})$, where J is the number of alternatives. In our case J=51.

[19] Betancourt and Clague (1981) continue to develop several measures that capture the amount of information provided by the introduction of the theoretical model relative to the information contained in the sample. Since my goal is only to compare the three specifications, I do not calculate the other measures. Interesting readers should refer to their book for more information.

[20] A negative value of $\overline{I}$ suggests that the misinformation contained in wrong predictions exceed the information contained in correct predictions. It can arise for two reasons: (1) there more wrong predictions than correct predictions, (2) the wrong predictions generate probabilities farther away from 1/51 relative to the correct predictions.

utility for the consumer on that shopping trip. Assuming that the consumer would purchase the same quantity of cereal as in the data, I can then calculate the consumer expenditure on that shopping trip. Summing up the consumer expenditures for each brand in the sample period, I get the predicted brand sales and brand market shares. Then I square the difference of predicted market share and observed market share for each brand, sum up the squared differences for all brands, and take the squared root of it to obtain the measure of market share prediction error. As shown in Table 8, the third specification generates the smallest market share prediction error compared to the first two.

In summary, introducing limited information about brand existence into the model improves the data fit and better captures consumer behavior. Therefore, I will base the following estimation on the limited information specification where consumer choice sets are heterogeneous.

The estimated parameters have important implications for brand pricing and advertising strategies. A brand's pricing decision depends on the price elasticity of demand. Advertising provides product information and impacts the composition of consumer choice sets, which can also affect consumer substitution. Therefore, a brand's advertising level also depends on the consumers' sensitivity to changes in advertising.

Given the parameter estimates in Column III of Table 6, I calculate the own and cross price elasticities for the top 25 brands[21], which are reported in Table 9. The formula for computing the price elasticities are included in Appendix 4. The price elasticities are evaluated at the median of each brand's price and the sample market shares.

[21] The remaining 25 brands have market shares less than 1% and relatively few observations, therefore the simulation errors might be big.

As is previously mentioned, the researchers do not observe some brand characteristics but market participants do. This leads to inconsistency in price and advertising coefficients because firm pricing and advertising decisions are most likely related to the unobserved brand characteristics. A solution to this endogeneity problem involves instrument variables. BLP show that variables that shift markups are valid instruments for price in a market of differentiated products. A similar argument also applies to advertising. Therefore, in Chapter 3, I choose instruments for price and advertising and re-estimate the demand function.

Chapter 3: Instrumental Variable Estimation

Section 3.1: IV Estimation Algorithm

A variety of differentiated-products pricing models predict that price is a function of marginal cost and a mark-up term. If advertising is also included as a decision variable, then the models also predict that advertising is a function of marginal cost and characteristics of other brands.

The firm side competition suggests that the optimal price and advertising levels depend on the characteristics, prices, and advertising levels of all brands offered. Brands facing more competition (due to existence of close substitutes in the characteristic space) will tend to have lower markups relative to brands facing less competition. If brand characteristics are exogeneous, then the characteristics of other brands are valid instruments for price and advertising. In the RTE cereal market, characteristics of a brand will not change once the brand is introduced into the market. Therefore the exogeneity of brand characteristics is a reasonable assumption. However, the price and advertising levels of other brands are not valid instruments, since they are correlated with unobserved brand characteristics through consumer utility maximization. On the other hand, variables that shift production costs, for example, ingredient prices, wages of manufacturer workers, are candidates for instruments too. The following section will present the results of demand estimation using these instruments. Note that the estimation results do not depend on any assumption on firm competition except that the firms are competing in a differentiated product market by choosing both price and advertising.

In the nonlinear discrete choice model, IV estimation cannot be directly implemented on the consumer level data. One way to make use of instrument variables involves aggregating consumer choices and matching predicted brand market shares with observed brand market shares, then inversing the market shares to get the component of utility function that does not vary with individuals. This component is a linear function of price, advertising, and other brand characteristics, and one can estimate this function with IV for price and advertising.

Formally, let $\chi_{jt} = [fiber_{jt}\ sugar_{jt}\ price_{jt}\ adv_{jt}]$[22],

$\beta_i = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \alpha_i \\ \rho_i \end{bmatrix} = \bar{\beta} + \Pi \bullet D_i + \Sigma \bullet v_i$, where $\bar{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \alpha \\ \rho \end{bmatrix}$, $D_i = [income_i\ age_i\ nokid_i]'$, and $v_i = [v_{1i}\ v_{2i}\ v_{3i}\ v_{4i}]'$.

Then we can write the utility function as

$$U_{ijt} = \chi_{jt}\beta_i + \kappa \cdot unused_{ijt} + \lambda_i \cdot unused_{ijt} \cdot adv_{jt} + pastchoice_{ijt} \bullet \gamma + \eta_{jt} + \varepsilon_{ijt} \qquad (20)$$

Let $\delta_{jt} = x_{jt}\beta + \eta_{jt}$ (21)

Note that although Π, Σ, κ, λ, and γ can be estimated with micro data, we cannot estimate $\bar{\beta}$ without a further assumption to separate out the effect of η from the effect of χ on δ. To provide consistent estimates of $\bar{\beta}$ we will use IV for price and advertising.

All parameters are estimated simultaneously. The estimation involves three sets of moment conditions:

[22] Although the true fiber and sugar contents of brands do not vary over time, the expected fiber and sugar contents do.

(1)The consumer brand choices, which match the model's predicted brand choice probabilities to observed brand choices,

(2)Brand market shares, which match the model's prediction for brand j's market share in year t to its observed market share in year t,

(3)Firms pricing and advertising decisions, which express an orthogonality between the unobserved characteristics and the instruments.

To be more specific, the estimation algorithm consists of four steps. (i) Given an initial guess of Π, Σ, κ, λ, and γ, I first find the values of δ_{jt} that equates the predicted market shares ($\sigma_{jt}(\delta, \Pi, \Sigma, \kappa, \lambda, \gamma)$) and the observed market shares (S_j) using the iteration $\delta_{jt}^{h+1} = \delta_{jt}^{h} + \ln(S_{jt}) - \ln(\sigma_{jt}(\delta^{h}))$. The details of calculating $\sigma_{jt}(\delta, \Pi, \Sigma, \kappa, \lambda, \gamma)$ and the proof that the above iteration is a contraction mapping are provided in Appendix 3. (ii) Given δ_{jt}, I provide random draws for unobserved consumer heterogeneity and for choice set formation, then use maximum simulated likelihood to obtain estimates of Π, Σ, κ, λ, and γ by matching the observed choices with the predicted choice probabilities. Note that these estimates do not depend on any distributional assumptions of η. The probability of a household with observed characteristics D_i choosing brand j given δ, Π, Σ, κ, λ, and γ b is given by

$$\Pr(j \mid D_i, \delta, \Pi, \Sigma, \kappa, \lambda, \gamma) = \int_v \frac{\exp(\delta_{jt} + \chi_{jt} \bullet \Pi \bullet D_i + \chi_{jt} \bullet \Sigma \bullet v + \kappa \cdot unused_{ijt} + \lambda_i \cdot unused_{ijt} \cdot adv_{jt} + pastchoice_{ijt} \bullet \gamma)}{\sum_{k=1}^{51} \exp(\delta_{kt} + \chi_{kt} \bullet \Pi \bullet D_i + \chi_{kt} \bullet \Sigma \bullet v \kappa \cdot unused_{ijt} + \lambda_i \cdot unused_{ijt} \cdot adv_j + pastchoice_{ijt} \bullet \gamma)} f(v) d(v)$$

(22)

The integrals are computed by simulation. (iii) Given the new values of Π, Σ, κ, λ, and γ, repeat the first two steps until δ, Π, Σ, κ, λ, and γ converge. (iv) Using the δ_{jt}

obtained in step (iii), construct the moment condition $E(\eta_{jt} \mid z) = 0$, where $\eta_{jt} = \delta_{jt} - \chi_{jt}\overline{\beta}$ and z represents instrument variables, and estimate $\overline{\beta}$ by minimizing the sample moments $G(\overline{\beta}) = \sum_j \eta_j \bullet z_j$ [23].

Section 3.2: IV Estimation Results

I implement the estimation algorithm with two sets of instruments. The first set of instruments includes the fiber and sugar contents of all other brands. The second set of instruments are the cost shifters, including wage of food workers, wage of advertising managers, corn price, wheat price, gasoline price, and electricity price. From the website of Bureau of Labor Statistics, I collect the hourly wage data for food workers (under the category Food and Tobacco Roasting, Baking, and Drying Machine Operators and Tenders) and for advertising mangers (under the category Advertising , and Public Relations Managers) in the Los Angeles – Long Beach MSA from year 1999 to year 2003. Corn price and wheat price are obtained from the Farmdoc project of University of Illinois. Gasoline and electricity prices are collected from the website of Energy Information Administration of Department of Energy. These cost measures are interacted with brand dummies to serve as instruments in the estimation. The $\overline{\beta}$ estimates combining both sets of instruments are reported in Column IV of Table 6.

To test the endogeneity of price and advertising, I run an OLS regression of equation (21) after I obtain δ_{jt} in step (iii) and compare the coefficients with the IV estimates. The Hausman test of the two sets of estimates yields a P value of 0.55, therefore the OLS

[23] Berry, Linton and Pakes (2004) show that in this type of BLP model with two sources of errors, the sampling error and the simulation error, both the number of observations and the number of random draws for simulation need to grow at rate J^2 for the parameter vector to have an asymptotically normal distribution.

estimates are not significantly different from with IV estimates. Hence the endogeneity of price and advertising does not affect the coefficient estimates much in this application. Since the price and advertising coefficient estimates without IV are much more precise than the IV estimates (in the IV estimation only 255 observations (δ by brand and by year) can be used while in the estimations without IV 37,858 transactions are used), I will conduct policy experiments using the estimates without IV in the following sections.

Chapter 4. Policy Experiments

In this chapter I conduct some counterfactual experiments to evaluate some of the brand marketing strategies and a hypothetical food policy change. In the first two experiments, I choose brand 28 as an example because it was newly introduced into the market in January 2003 (Figure 5 summarizes brand 28' average monthly prices, sales, and advertising in the estimation sample). Marketing managers are usually concerned with what price to charge and how to schedule advertising expenditures when a new product is launched. Therefore looking into the data of brand 28 offers us an opportunity to evaluate the marketing strategies of a product at the beginning of its life cycle. In the third experiment, I try to explore the effect of a hypothetical policy change — banning cereal advertising targeted to children — on consumer choices. A caveat should be born in mind when we interpret the results of the experiments: I keep the strategies of other firms unchanged when I simulate the results, thus the optimal responses of rival firms are not taken into account[24].

Section 4.1: Pricing Strategy for Brand 28

I first vary brand 28's price from its observed price by +1%, +5%, -1%, and -5%, separately. Each time under the new pricing scheme, I calculate every household's simulated choices and aggregate them to get brand market shares and sales. The resulting

[24] To derive the optimal responses, we need to solve a competitive equilibrium. However, the static Bertrand equilibrium is not realistic and the dynamic equilibrium is very hard to solve.

changes in market share and sales of brand 28 are reported in Table 10. We can see that if the price is cut down by 5%, the sales improve by 2.3% compared to the sales figure before the price cut. The market share expands by 6.5%, which more than compensates for the reduction in price. Therefore, brand 28's price was too high in general.

To see how the price cut affects different types of consumers, I calculate the changes in expenditures for different demographic groups after the price drops by 5%. I divide consumers by household income (high if household income>=$55,000, low otherwise), by age of female household head (old if age>=32, young otherwise), and by the presence of kid in the household. The results by demographic groups are shown in Table 11. Consumers with lower income and having kids respond more to the price cut than their counterparts, while the response does not vary with age groups.

Next I look at the average (weighted by volume) daily transaction prices of brand 28 at its introductory stage (the first 3 months of year 2003) and see if its sales can be increased by altering the depth and frequency of the price discounts. The observed daily transaction price series for brand 28 from Jan 2003 to March 2003 is shown in Figure 6. The initial price was very high, followed by a period of medium price level. Deep price discounts happened twice when the price was about 60% of the average level. I consider an alternative pricing strategy, whereby price is set to be 70% of the average price in this period for the first week of each of the 3 months and 100% of the average price in the remaining weeks. The observed prices and the counterfactual prices in this period are plotted in Figure 4. With the new pricing strategy, I find that brand 28's market share goes up by 1.5% and sales go up by 1.2%. High introductory price is not desirable in this case because consumers are loyal to brands they are already using. To warrant a switch, the utility associated with the new brand needs to be sufficiently high, which could be achieved by lower introductory price. Consumers who are lured into purchase by the low introductory prices will then form brand loyalty, thus the brand manager can profit by setting price low initially and increasing it later.

There may be two reasons why the brand manager would set a high initial price as observed in the data. On the one hand, higher prices may be used by the brand manager as a signal for better quality in a market with limited information and hence attracts consumers with higher willingness to pay. However, in the cereal market, many private label products have been introduced at low prices and many consumers have come to realize that lower price does not necessarily affect the quality or taste[25]. Therefore, a high initial price would limit the consumer demand. On the other hand, the brand manager may have underestimated the price elasticity. As shown in section 5, if demand is estimated ignoring that consumers have limited information about product brand existence, price elasticities would be understated, which could lead the manager to set a higher than optimal price.

Section 4.2: Advertising Strategy for Brand 28

A major consideration of a brand manager is to determine the best schedule of advertising expenditures for a certain budget. Conceptually, the manager could choose to do continuous advertising (i.e., schedule ad expenditure smoothly over all times) or follow a strategy of pulsing (i.e., advertise in some weeks of the year and not at other times). We observe in Figure 5 that brand 28's advertising was relatively smooth over time. In

[25] See the article "Eating Well", New York Times, September 22, 1993.

contrast, many advertisers of consumer packaged goods use pulsing strategies. For example, Dubé, Hitsch, and Manchanda (2005) find that pulsing is the optimal advertising strategy in the frozen entrée market. Naik , Mantrala , and Sawyer (1998) develops a model of dynamic advertising that shows that pulsing strategies can generate greater total awareness than the continuous advertising when the effectiveness of advertisement varies over time. Specifically, ad effectiveness declines because of advertising wears out during periods of continuous advertising and it restores during periods of no advertising. Such dynamics make it worthwhile for advertisers to stop advertising when ad effectiveness becomes very low and wait for ad quality to restore before starting the next "burst" again. They also show that the form of the best advertising spending strategy is pulsing for a major cereal brand.

To mimic the pulsing strategy, I reschedule brand 28's advertising by equally dividing the 2003 total ad expenditure into the six odd months, while in the six even months advertising is set to zero (Figure 7 plots the observed advertising v.s. the counterfactual pulsing advertising). Then I recalculate consumer choices under the new advertising strategy. The results show that brand 28's market share and sales both increase by 1.9%. The pulsing strategy works better because it can increase the probability of brand 28 entering the consumer choice set in the first two months after its introduction. In the observed data, the advertising expenditure for brand 28 in January is 0 while in the pulsing strategy it is 3.3 million. The increase in the advertising expenditure in January raises the probability of an average consumer (with mean income, mean age, and mean presence of kid) being aware of brand 28 from almost 0 to 89.7%. In February the pulsing strategy increases the advertising expenditure of brand 28 from 2.99

to 3.14 million and the probability of an average consumer being aware of brand 28 from 78.3% to 84.5%. In the following months an average consumer will be aware of the brand with probability close to 1 in both strategies. Therefore, under the pulsing strategy, more consumers are aware of the brand since the beginning and have a higher probability of choosing it. Some of these consumers become habituated to the brand and hence the pulsing strategy can increase the overall market share of this brand. I also examine how different consumer groups respond to the pulsing strategy. Results in Table 12 suggest that consumers with higher income and kids are more sensitive to the change in advertising strategy, while age does not matter.

In the advertising data, 98.9% of advertising expenditure is spent on national media such as network TV, national sport radio, and national newspaper. Therefore, the pulsing strategy could also increase sales in other local markets without changing the advertising budget, and could potentially be very profitable.

Section 4.3: Effects of Banning Children-Oriented Cereal Advertising

There have been many public debates on food marketing aimed at children. As early as in 1978, the FTC issued a staff report that concluded that "television advertising for any product directed to children who are too young to appreciate the selling purpose of, or otherwise comprehend or evaluate, the advertising is inherently unfair and deceptive", and that "it is hard to envision any remedy short of a ban adequate to cure this inherent unfairness and deceptiveness." Naturally FTC faced strong opposition from broadcasters, ad agencies, and food and toy companies. And in 1980 Congress passed the Federal Trade Commission Improvements Act of 1980 that barred the FTC from issuing industrywide regulations to stop unfair advertising practices[26]. However, as the childhood obesity problem becomes more and more of a concern in recent years[27], policymakers in Congress, FTC, and some consumer advocates are calling for restrictions on advertising to children about candy, sugary cereal, and other junk food.

A study by The Kaiser Family Foundation indicates that children of all age groups are exposed to a large amount of food ads every day (Table 13). Of all genres on TV, shows specifically designed for children under 12 have the highest proportion of food advertising (50% of all ad time). And of all food ads that target children or teens, 28% are for sugary cereal[28]. Table 14 summarizes the sugar and fiber contents of different types of cereal brands. Not surprisingly the kid brands are highest in sugar and lowest in fiber. It is, therefore, interesting to see if the cereal TV advertising toward children were to be banned, what would happen to consumer expenditures and nutritional intakes.

In our data, I do not directly observe how much ad dollars are for marketing toward children. In order to measure the effects of the policy change, I approximate the ban of children oriented cereal advertising by eliminating the advertising expenditures for kid cereal brands while holding other factors unchanged. In the experiment, I replace the ad stock of these brands with 0, and calculate how the brand market shares change. The total changes for each brand segment (Family, Adult, Kid) are summarized in Table 15. After the hypothetical policy change, the total market share of kid brands goes down by about 6%, of which 2% goes to the adult brands and 4% goes to the family brands.

[26] See the article "Limiting Food Marketing to Children" on www.cspinet.org/nutritionpolicy.
[27] On Wikipedia.com, it is stated that the rate of overweight and obese children in the United Staes is 32% in 2008.
[28] See "Food for Thought: Television Food Advertising to Children in the United States", released by The Kaiser Family Foundation, March 28, 2007.

Then I look at how the policy change affects the nutrition intakes and expenditures of different consumer groups. The results are summarized in Table 16. Overall after the ban of children oriented cereal advertising, consumers will consume more fiber and less sugar, which is better for their health. Consumers who are younger, with lower income and with kids reduce their sugar intake and increase their fiber intake more than their counterparts. Therefore, the policy change seems to have more effect on the "right" group of consumers. In the meantime, after the ban consumers of all demographic groups will have to increase their expenditures after the policy change as they consume more adult and family cereals, which are more expensive than kid cereals.

Section 4.4: Concluding Remarks

This paper considers limited information on both product existence and product quality in a dynamic model of consumer choice in an experience good market. On each purchase occasion, a consumer first forms a choice set depending on her purchase experience and brand advertising. Conditional on the choice set, she then chooses the brand that maximizes her expected utility. The empirical model is estimated on a rich panel of household purchase data in the RTE cereal market. The results suggest that limited information about product existence leads to larger estimates on price elasticity. Advertising has a significant effect informing consumers of brand existence and signaling product quality, while its prestige effect is not significant. I also find that conditional on repurchase after the first experience with a brand, consumers form strong habits. The counterfactual experiments show that the observed brand marketing strategies are not always optimal: managers can increase sales by resetting price, rescheduling price discounts, or altering their advertising strategies. An evaluation of a hypothetical policy change – ban on cereal advertising toward children – suggests that the policy could direct consumers to healthier diet.

The results of the experiments should be taken with caution though. In all of the experiments I do not consider the competitive responses of other firms to the change in brand strategies. Neither do I account for the fact that firms may change the way they promote kid brands once the government regulation comes into play. To control for these responses, I will need to solve the firm profit maximization problem. In a model with brand loyalty on the consumer side, the firm side problem should involve dynamic optimization: firms not only consider the effect of pricing and advertising on current consumer choices, but also the effect on future demand and future profits. However, the dynamic optimization problem with multiple firms each with multiple brands is extremely hard to solve and thus left for future research. In addition, many brand marketing strategies are decided by manufacturers and retailers together. This paper only focuses on the role of manufacturers. A vertical competition model will be needed to analyze the role of retailers.

Appendices

Appendix 1. Controlling for Unobserved Consumer Heterogeneity

I use introduce consumer-brand random effects to capture the unobserved consumer heterogeneity in brand preferences. Specifically, the utility function can be written as:

$$U_{ijt} = Z_{ijt} \bullet \Phi + \nu_{ij} + \varepsilon_{ijt}$$

where Z_{ijt} represents the vector of explanatory variables, Φ represents the vector of coefficients corresponding to Z_{ijt} , and ν_{ij} represents consumer i's unobserved preference for brand j, which is independent from Z_{ijt} and ε_{ijt} .

Let $\nu_{ij} = \mu_{ij} + \omega_j$. $\mu_{ij} \sim N(0, \varsigma_{ij}^2)$, and $\omega_j = E(\nu_{ij})$ is a constant. Assuming ε_{ijt} has a Generalized extreme value distribution, then we can write the probability of consumer i choosing j conditional on μ_{i1} , μ_{i2} , …, μ_{i51} , and choice set C_{it} as:

$$P(j \mid \mu_{i1}, \mu_{i2}, ..., \mu_{i51}, C_{it}) = \frac{\exp((Z_{ijt} - Z_{i51t}) \bullet \Phi + \mu_{ij} + \omega_j - \omega_{51})}{\sum_{l=1}^{51} \exp((Z_{ilt} - Z_{i51t}) \bullet \Phi + \mu_{il} + \omega_l - \omega_{51})}$$

$$= \frac{\exp(z_{ijt} \bullet \Phi + \mu_{ij} + \xi_j)}{\sum_{l=1}^{51} \exp(z_{ilt} \bullet \Phi + \mu_{il} + \xi_l)}$$

where for the second equal sign we use $z_{ijt} = Z_{ijt} - Z_{i51t}$ and $\xi_j = \omega_j - \omega_{51}$.

$p(j \mid C_{it})$ is equal to $P(j \mid \mu_{i1}, \mu_{i2}, ..., \mu_{i51}, C_{it})$ integrated over the marginal distribution of the μ_{ij} 's. Specifically, it is equal to

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \frac{\exp(z_{ijt} \bullet \Phi + \mu_{ij} + \xi_j)}{\sum_{l=1}^{51} \exp(z_{ilt} \bullet \Phi + \mu_{il} + \xi_l)} f(\mu_{i1}) f(\mu_{i2}) ... f(\mu_{i51}) d\mu_{i1} d\mu_{i2} ... d\mu_{i51}$$

It is hard to compute $p(j \mid C_{it})$ analytically and I simulate it by taking S draws from the distribution of μ_{ij}, for all j. The simulator for $p(j \mid C_{it})$ is:

$$\hat{p}(j \mid C_{it}) = \frac{1}{S}\sum_{s=1}^{S} \frac{\exp(z_{ijt} \bullet \Phi + \mu_{ij}^{s} + \xi_j)}{\sum_{l=1}^{51} \exp(z_{ilt} \bullet \Phi + \mu_{il}^{s} + \xi_l)}$$

To reduce the number of parameters to be estimated, I allow ω_j to vary across brand segment, and ς_{ij}^2 to vary across both brand segment and whether the household has kids or not. There are a total of eight parameters to estimate for unobserved consumer-brand preferences, among which six are scale parameters: $\varsigma_{FK}^2, \varsigma_{FN}^2, \varsigma_{AK}^2, \varsigma_{AN}^2, \varsigma_{KK}^2, \varsigma_{KN}^2$, where the first subscript denotes whether the brand belongs to family, adult or kid segment, and the second subscript denotes whether there is any kid in the household; two are location parameters: ω_A and ω_K, where the subscript denotes whether the brand belongs to adult or kid segment. ω_F is normalized to zero.

Appendix 2. Choice Set Simulation Details

In the simulation, I assume that choice set is a function of brand advertising and purchase experience as shown in equation (1) and equation (2). The specific choice set simulation process is outlined as follows.

Step 1. Calculate q_{ijt} (φ) for each consumer, each brand, and each time, where $\varphi=(\varphi_0, \varphi_1, \varphi_2)$.

Step 2. For each consumer-time-brand combination, draw a random number u_{ijt}^{r} from the uniform distribution between 0 and 1.

Step 3. If $u^r_{ijt} < q_{ijt}$, then brand j is included in consumer i's choice set at time t. Otherwise it is not. This defines the choice set in the rth simulation C^r_{it}.

After simulating the choice set, I can calculate simulated brand choice probabilities for each consumer.

Step 4. Calculate $P^r(j|C_{it})$, consumer i's probability of choosing brand j conditional on C^r_{it}. (The formula for calculating $P^r(j|C_{it})$ depends on the distributional assumption on the error term in the utility function).

Step 5. Calculate $p^r_{ijt} = \prod_{j \in C^r_{it}} q_{ijt} \prod_{k \notin C^r_{it}} (1-q_{ikt}) \times P^r(j|C_{it})$, consumer i's unconditional probability of choosing brand j at time t in the rth simulation.

Step 6. Draw the random numbers u^r_{ijt} repeatedly for R times, each time repeat steps 2-5.

Step 7. Calculate the simulated choice probability $\hat{p}_{ijt} = \frac{1}{R}\sum_{r=1}^{R} p^r_{ijt}$.

Appendix 3. Contraction Mapping Details

In the instrumental variable estimation, we need to find the δ that make predicted market shares based on the model equal to the observed market shares. Given an initial guess of δ, Π, and Σ, the predicted market share for brand j, $\sigma_j(\delta^h, \Pi, \Sigma, \kappa, \lambda, \gamma)$, is calculates as follows.

First, based on advertising data and household characteristics, simulate choice sets for each consumer on each shopping occasion.

Second, given δ, Π, Σ, κ, λ, and γ, a consumer compares the utility levels of all brands in her choice set on the shopping occasion and chooses the one that yields the highest utility.

Third, sum up consumer brand choices in a year to get predicted brand market shares.

To obtain the values of δ that solves $\sigma_j(\delta^h,\Pi,\Sigma,\kappa,\lambda,\gamma)=S_j$, we use the iteration

$\delta_j^{h+1}=\delta_j^h+\ln(S_j)-\ln(\sigma_j(\delta^h,\Pi,\Sigma,\kappa,\lambda,\gamma))$. The proof that the iteration is a contraction mapping follows Goeree (2007).

Define $f(\delta_j)=\delta_j+\ln(S_j)-\ln(\sigma_j(\delta^h,\Pi,\Sigma,\kappa,\lambda,\gamma))$. To show that f is a contraction mapping, we need to show that $\forall$ j and m, $\partial f(\delta_j)/\partial\delta_m\geq 0$, and $\sum_{m=1}^{J}\partial f(\delta_j)/\partial\delta_m<1$.

We can write $\sigma_j=\int\sum_{C_i\in\Omega_j}\prod_{l\in C_i}q_{ilt}\prod_{k\notin C_i}(1-q_{ikt})P(j\mid C_i)f(v)dv$,

where , and Ω_j denotes the set of choice sets that include j.

$$\partial f(\delta_j)/\partial\delta_m=\frac{1}{\sigma_j}\int\sum_{C_i\in\Omega_j}\prod_{l\in C_i}q_{ilt}\prod_{k\notin C_i}(1-q_{ikt})P(j\mid C_i)Q_j^m f(v)dv,$$

where

$$p(j\mid C_i)=\int_v\frac{\exp(\delta_j+\chi_j\bullet\Pi\bullet D_i+\chi_j\bullet\Sigma\bullet v+\kappa\cdot unused_{ij}+\lambda_i\cdot unused_{ij}\cdot adv_j+pastchoice_{ij}\bullet\gamma)}{\sum_{k=1}^{51}\exp(\delta_k+\chi_k\bullet\Pi\bullet D_i+\chi_k\bullet\Sigma\bullet v\kappa\cdot unused_{ij}+\lambda_i\cdot unused_{ij}\cdot adv_j+pastchoice_{ij}\bullet\gamma)}f(v)d(v)$$

$$Q_j^m=\frac{\exp(\delta_m+\chi_m\bullet\Pi\bullet D_i+\chi_m\bullet\Sigma\bullet v+\kappa\cdot unused_{im}+\lambda_i\cdot unused_{im}\cdot adv_m+pastchoice_{im}\bullet\gamma)}{\sum_{l\in C_i}\exp(\delta_l+\chi_l\bullet\Pi\bullet D_i+\chi_l\bullet\Sigma\bullet v+\kappa\cdot unused_{il}+\lambda_i\cdot unused_{il}\cdot adv_l+pastchoice_{il}\bullet\gamma)},if\ m\in\Omega_j$$
$$=0,if\ m\notin\Omega_j$$

Note that for m=j, $Q_j^m=P(j\mid C_i)$

Since all elements in the integral are non-negative, we have $\partial f(\delta_j)/\partial\delta_m\geq 0$.

Moreover, $\sum_{m\in\Omega_j,m\neq 51}Q_j^m<1$, therefore $\sum_{m\in\Omega_j,m\neq 51}\partial f(\delta_j)/\partial\delta_m<1$ is satisfied.

Appendix 4. Price Elasticity Calculation

Suppressing the time subscript, we can write the consumer utility function as
$U_{ij} = \alpha_i p_j + \Upsilon_j \bullet \beta_{\Upsilon i} + \varepsilon_{ij}$
where $\alpha_i = \alpha + \Pi_3 \bullet D_i + \Sigma_{33} \cdot v_3$, Υ_j represents the vector of variables other than price, and $\beta_{\Upsilon i}$ the vector of coefficients for Υ_j.

The formula for price elasticity is given by

$$\rho_{jk} = \frac{\partial s_j}{\partial p_k} \cdot \frac{p_k}{s_j} = \begin{cases} \dfrac{p_j}{s_j} \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \int \alpha_i \hat{p}_{ij} (1 - \hat{p}_{ij}) f(v) dv, \; j = k \\ -\dfrac{p_k}{s_j} \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \int \alpha_i \hat{p}_{ij} \hat{p}_{ik} f(v) dv, \; j \neq k \end{cases}$$

where p_{ij} represents the probability of consumer i choosing brand j.

In the estimation, I take NR random draws of v from f(v) to get α_i and compute ρ_{jk} using the formula

$$\hat{\rho}_{jk} = \begin{cases} \dfrac{p_j}{s_j} \dfrac{1}{N * NR} \displaystyle\sum_{i=1}^{N} \sum_{nr=1}^{NR} \alpha_i^{nr} \hat{p}_{ij} (1 - \hat{p}_{ij}), \; j = k \\ -\dfrac{p_k}{s_j} \dfrac{1}{N * NR} \displaystyle\sum_{i=1}^{N} \sum_{nr=1}^{NR} \alpha_i^{nr} \hat{p}_{ij} \hat{p}_{ik}, \; j \neq k \end{cases}$$

Figure 1.

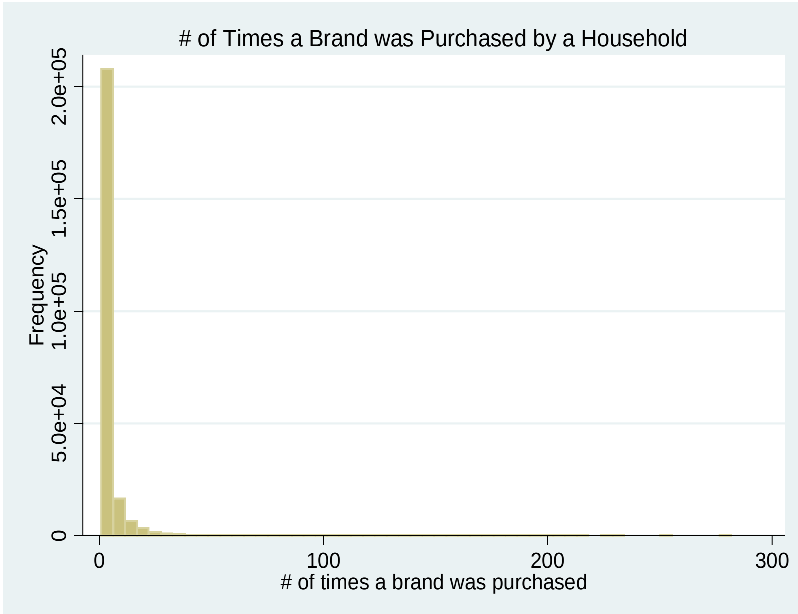

Figure 2.

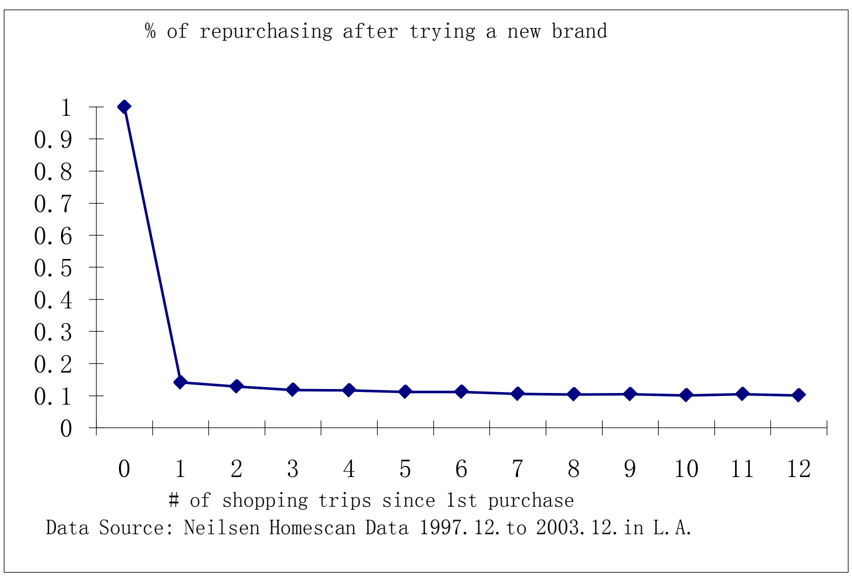

Figure 3.

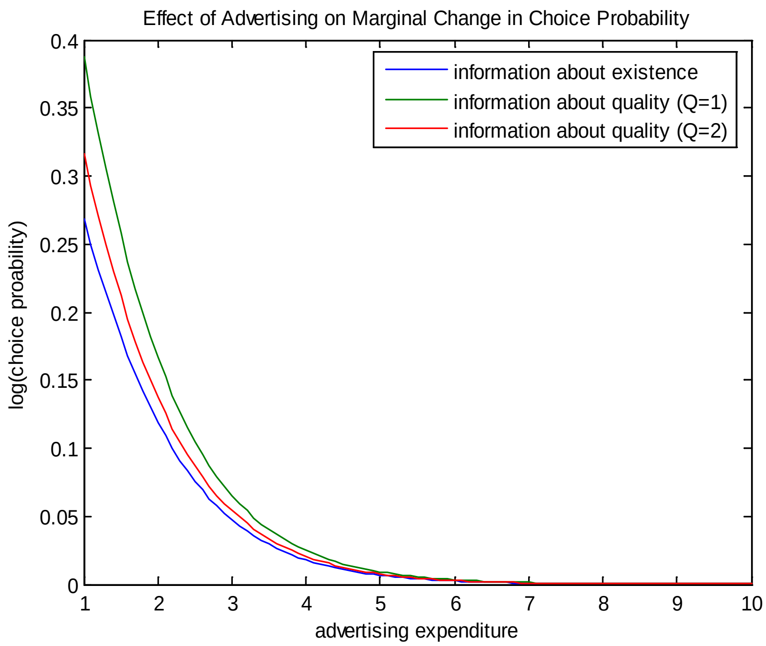

Figure 4.

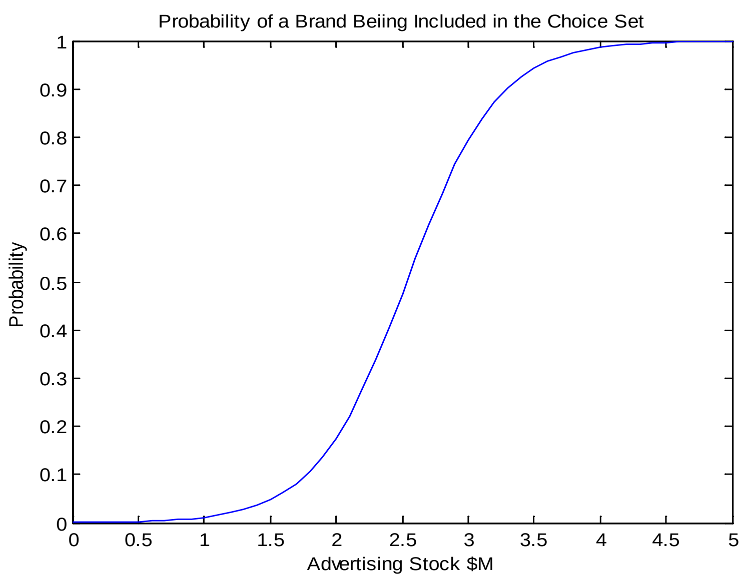

Figure 5.

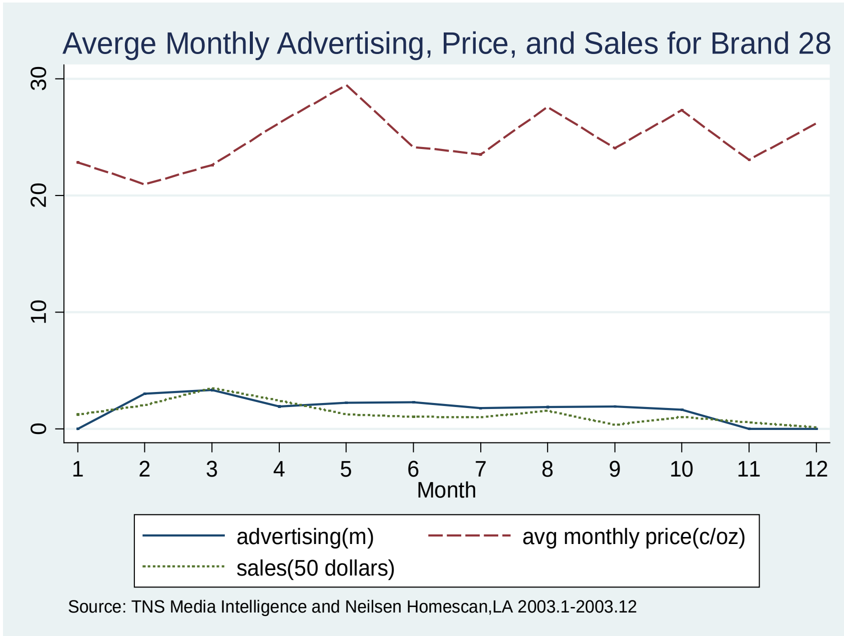

Figure 6.

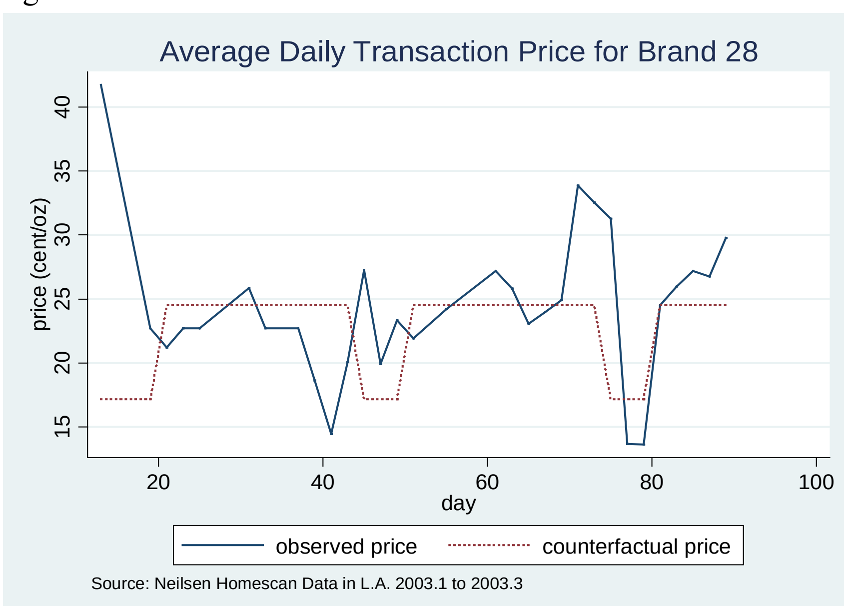

Figure 7.

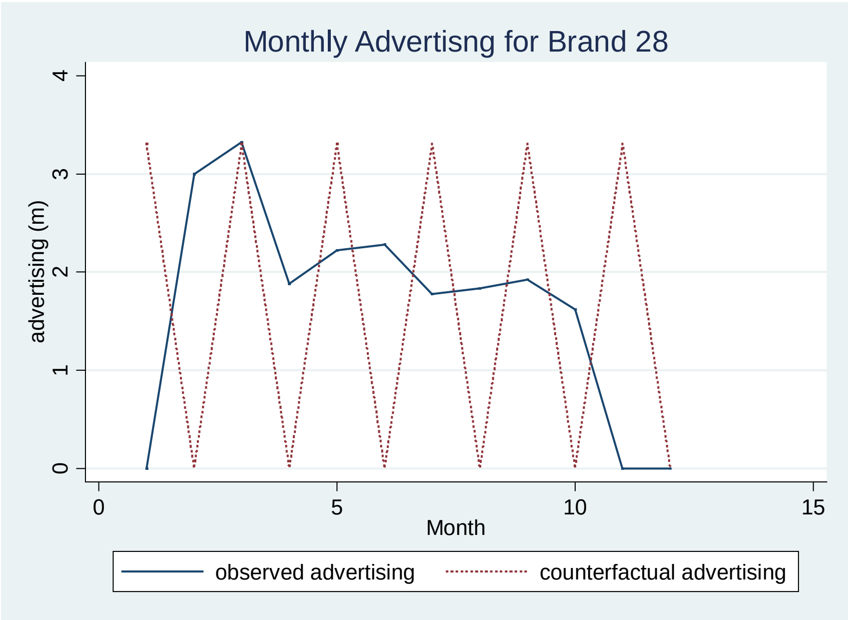
Monthly Advertisng for Brand 28
advertising (m)
0
1
2
3
4
0
5
10
15
Month
observed advertising
counterfactual advertising

Table 1. Brand Entry and Exit

	Exit Year						
Enter Year	1998	1999	2000	2001	2002	Remaining	Total
1998 & before	5	3	3	10	7	103	131
1999	0	1	4	1	2	10	18
2000	0	0	0	1	3	3	7
2001	0	0	0	1	3	8	12
2002	0	0	0	0	10	30	12
2003	0	0	0	0	0	13	13
Total	5	4	7	13	17	147	193

Data Source: Neilsen Homescan Data 1997.12. to 2003.12. in Los Angeles Market.
A brand entry is observed if the first transaction with the brand occurred after 1998.6. A brand exit is observed if the last transaction with the brand occurred before 2003.7.

Table 2. Brand Summary Statistics

Brand Number	Sample Market Share[1] (%)	Average Transaction Price (cent/oz)	Average Monthly Advertising ($K)	Fiber Content (% Daily Value per 30g)	Sugar Content (% Daily Value per 30g)	Segment[2]
1	5.73	17.22	1718.89	14.00	1.00	Family
2	4.51	18.44	1977.76	6.45	4.59	Family
6	4.45	12.26	2036.62	11.25	6.23	Adult
11	4.07	14.01	1045.85	5.90	6.39	Adult
8	3.99	11.95	1667.88	2.42	9.68	Family
7	3.69	11.96	1445.58	11.49	10.37	Family
3	3.56	15.85	1701.03	7.00	11.00	Family
12	2.88	14.98	406.50	4.12	12.39	Kid
4	2.71	17.74	785.01	5.00	10.00	Kid
16	2.50	13.71	319.99	3.56	6.71	Family
20	2.49	11.43	1623.61	10.34	4.40	Adult
9	2.38	18.22	878.28	0.03	7.91	Family
10	2.36	22.81	2143.30	9.15	5.21	Adult
15	2.35	18.66	437.65	6.00	13.00	Kid
13	2.09	13.80	1377.95	13.91	1.86	Adult
14	1.92	16.38	634.15	2.92	12.46	Kid
17	1.62	14.07	1293.78	7.50	7.03	Kid
23	1.62	9.60	5.65	14.75	8.64	Family
21	1.56	16.77	604.60	0.97	14.52	Family
18	1.55	15.61	698.44	8.30	8.61	Family
19	1.55	16.80	1243.06	1.59	13.39	Kid
38	1.48	17.48	435.98	4.00	13.00	Kid
5	1.39	20.91	1611.56	7.48	6.98	Adult
24	1.29	15.77	459.00	4.00	15.00	Kid
22	1.26	21.80	56.11	1.00	6.00	Adult
42	1.09	16.88	739.14	7.94	8.02	Adult
30	1.06	18.64	379.89	3.00	14.00	Kid
26	0.76	19.15	72.82	5.00	13.00	Family
25	0.72	19.11	423.71	4.00	11.00	Kid
47	0.72	15.08	746.76	3.10	11.38	Kid
46	0.71	17.01	87.48	8.69	5.88	Family
39	0.66	18.01	3.26	49.00	4.33	Adult
43	0.63	18.19	108.85	27.00	5.00	Adult
50	0.59	15.38	157.45	8.57	4.82	Adult
48	0.57	18.91	303.66	6.67	12.22	Kid
49	0.56	19.91	280.83	6.32	7.89	Family
27	0.54	19.55	0.00	7.09	7.64	Adult
40	0.52	15.64	102.95	1.94	10.65	Kid
45	0.46	21.43	177.74	4.36	6.00	Family
29	0.46	20.01	0.00	6.00	9.27	Family
31	0.44	23.30	208.99	2.00	13.00	Kid
37	0.43	21.49	381.66	0.00	12.00	Family
28	0.41	24.39	1653.81	12.00	10.00	Family
33	0.41	16.35	13.43	4.00	13.00	Family
44	0.35	16.95	229.33	8.13	6.10	Family
32	0.35	25.79	6.58	58.00	0.00	Adult
34	0.33	24.19	1.68	11.00	6.00	Family

Brand Number	Sample Market Share[1] (%)	Average Transaction Price (cent/oz)	Average Monthly Advertising ($K)	Fiber Content (% Daily Value per 30g)	Sugar Content (% Daily Value per 30g)	Segment[2]
41	0.31	14.64	0.00	4.44	16.67	Kid
35	0.29	17.64	62.32	9.00	9.27	Adult
36	0.25	17.39	0.00	10.91	8.73	Adult
51	21.41	14.90	47.44	8.83	8.00	Family

Data Source: Columns II&III-AC Neilsen Homescan Data 1997.12 to 2003.12, Column IV-TNS Media Intelligen Data Jan 1999 to Dec 2003, Columns V & VI - www.nutritiondata.com

1. Sample market is the Los Angeles Market from Dec 1997 to Dec 2003.

2. Brand segment categorization is based on each brand's description on manufacturer's website.

3. Characteristics of the 51th brand are computed as the average of the non top 50 brands.

Table 3: Summary Statistics of Homescan Data

Variable	Definition	NumObs	Mean	Std. Dev.	Min	Max
size	household size	1402	3.25	1.53	1	9
inc	household income ($K)	1402	57.11	29.58	2.5	125
age	age of female household head	1402	48.86	12.99	20	70
nokid	=1 if no kid in the household	1402	0.55	0.50	0	1
price	transaction price (cent/oz)	69134	17.84	4.73	0	797.44

Data Source:Neilsen Homescan Data 1997.12. to 2003.12 in Los Angeles Market

Table 4. Summary of Variables in Estimation Sample[1]

Variable	Definition	Mean	Stdev	Min	Max
chosen	1{brand is chosen in current transaction}	0.02	0.14	0	1
price	transaction price (cent/oz)	17.84	4.75	0	797.44
*price*inc*	transcation price(cent/oz)*household income($K)	1031.84	629.38	0	13000
*price*nokid*	transcation price(cent/oz)*1{household has nokid}	892.50	348.03	0	7400
adv	stock of advertising expenditure ($M)	3.22	4.02	0	22.30
*adv*inc*	advertising stock($M)*household income($K)	188.47	281.56	0	2787.46
*adv*nokid*	advertising stock($M)*1{household has nokid}	1.95	3.51	0	22.30
unused	1{brand not purchased previously}	0.11	0.32	0	1
*unused*adv*	1{brand not purchased previously}*adv	1.82	3.29	0	22.30
*unused*adv*inc*	1{brand not purchased previously}*adv*household income($K)	106.57	223.03	0	2787.46
*unused*adv*nokid*	1{brand not purchased previously}*adv*1{household has nokid}	1.21	2.79	0	22.30
chosen_1	1{brand chosen on last shopping trip}	0.03	0.17	0	1
chosen_2	1{brand chosen 2 shopping trips ago}	0.03	0.17	0	1
chosen_3	1{brand chosen 3 shopping trips ago}	0.03	0.17	0	1
chosen_4	1{brand chosen 4 shopping trips ago}	0.03	0.17	0	1
chosen_5	1{brand chosen 5 shopping trips ago}	0.03	0.17	0	1
chosen_6	1{brand chosen 6 shopping trips ago}	0.03	0.17	0	1
sugar	sugar content(% daily value per 30g)	8.81	3.73	0	16.67
*sugar*age*	sugar content(% daily value per 30g)*age of female head	439.51	227.30	0	1166.67
*sugar*nokid*	sugar content(% daily value per 30g)* 1{household has no kid}	5.22	5.19	0	16.67
fiber	fiber content(% daily value per 30g)	8.58	10.28	0	58
*fiber*age*	fiber content(% daily value per 30g)*age of female head	428.18	544.41	0	4060
*fiber*nokid*	fiber content(% daily value per 30g)* 1{household has no kid}	5.08	8.96	0	58
*adult*size*	1{brand is adult brand}*household size	0.98	1.72	0	9
*adult*inc*	1{brand is adult brand}*household income($K)	17.25	30.98	0	125
*adult*age*	1{brand is adult brand}*age of female head	14.69	23.51	0	70
*adult*nokid*	1{brand is adult brand}*1{household has no kid}	0.16	0.37	0	1
*kid*size*	1{brand is kid brand}*household size	0.98	1.72	0	9
*kid*inc*	1{brand is kid brand}*household income($K)	17.22	30.95	0	125
*kid*age*	1{brand is kid brand}*age of female head	14.67	23.50	0	70
*kid*nokid*	1{brand is kid brand}*1{household has no kid}	0.16	0.37	0	1

1.Estimation sample consists of 890 households with 42396 transactions from 1999.1 to 2003.12 in LA market

Table 5. Preliminary Regression Results

Variable	*Coefficient*
price	0.014**
	(0.006)
*price*inc*	0.000***
	(0.000)
*price*age*	-0.000
	(0.000)
*price*nokid*	0.012***
	(0.003)
adv	-0.004
	(0.005)
*adv*inc*	0.000**
	(0.000)
*adv*age*	-0.000
	(0.000)
*adv*nokid*	-0.000
	(0.003)
unused	-1.802***
	(0.018)
*unused*adv*	0.698***
	(0.030)
*unused*adv*inc*	-0.001**
	(0.000)
*unused*adv*age*	-0.001
	(0.001)
*unused*adv*nokid*	0.037**
	(0.015)
chosen_1	0.563***
	(0.015)
chosen_2	0.589***
	(0.015)
chosen_3	0.565***
	(0.015)
chosen_4	0.503***
	(0.015)
chosen_5	0.488***
	(0.015)
chosen_6	0.491***
	(0.015)
fiber	-0.132***
	(0.006)
*fiber*age*	0.000***
	(0.000)
*fiber*nokid*	0.026***
	(0.003)

sugar	-0.136***
	(0.013)
*sugar*age*	-0.002***
	(0.000)
*sugar*nokid*	-0.033**
	(0.006)

Table 6: Estimation Results

	I RCL	II RCL+Learning	III RCL+Learning+HCS[1]	IV IV Estimation
price	-0.164*** (0.003)	-0.187*** (0.054)	-0.233*** (0.009)	-0.368*** (0.051)
*price*inc*	0.002*** (0.000)	0.001** (0.001)	0.001*** (0.000)	0.001*** (0.000)
*price*nokid*	0.127*** (0.003)	0.129*** (0.007)	0.118*** (0.007)	0.014*** (0.005)
adv	-0.007* (0.004)	-0.018 (0.065)	-0.023 (0.050)	0.274 (1.731)
*adv*inc*	0.001** (0.000)	0.000 (0.001)	0.001*** (0.000)	0.000*** (0.000)
*adv*nokid*	0.073*** (0.003)	0.009 (0.006)	0.073*** (0.003)	-0.003 (0.003)
unused	-1.874*** (0.021)	-2.071*** (0.017)	-1.800** (0.043)	-1.801*** (0.091)
*unused*adv*	0.335*** (0.006)	0.083*** (0.010)	0.286*** (0.023)	0.696*** (0.041)
*unused*adv*inc*	-0.004** (0.000)	-0.000 (0.001)	0.027*** (0.001)	0.028*** (0.001)
*unused*adv*nokid*	-0.155*** (0.005)	-0.041 (0.030)	-0.019** (0.009)	-0.035 (0.083)
chosen 1	0.614*** (0.017)	0.649*** (0.018)	0.578*** (0.018)	0.563*** (0.021)
chosen 2	0.638*** (0.017)	0.629*** (0.018)	0.603*** (0.018)	0.590*** (0.021)
chosen 3	0.612*** (0.017)	0.574*** (0.018)	0.577*** (0.019)	0.565*** (0.019)
chosen 4	0.548*** (0.017)	0.504*** (0.018)	0.514*** (0.019)	0.503*** (0.019)
chosen 5	0.531*** (0.017)	0.466*** (0.019)	0.497*** (0.019)	0.488*** (0.019)
chosen 6	0.534*** (0.017)	0.468*** (0.019)	0.501*** (0.019)	0.491*** (0.019)
fiber	-0.112*** (0.005)	0.014 (0.046)	-0.068*** (0.011)	4.466*** (0.853)
*fiber*age*	0.001*** (0.000)	0.001 (0.001)	0.000 (0.000)	0.001*** (0.000)
*fiber*nokid*	0.018*** (0.003)	0.037*** (0.011)	0.047*** (0.005)	0.026*** (0.004)
sugar	-0.083*** (0.009)	-0.060 (0.059)	-0.099*** (0.020)	1.996 (2.293)

	I RCL	II RCL+Learning	III RCL+Learning+HCS[1]	IV IV Estimation
*sugar*age*	0.001	-0.001	0.000	-0.001***
	(0.001)	(0.003)	(0.000)	(0.000)
*sugar*nokid*	-0.062**	-0.007	-0.066***	-0.032***
	(0.005)	(0.006)	(0.007)	(0.005)
Φ_0			-7.350***	-7.350***
			(0.005)	(0.001)
Φ_1			2.996***	2.996***
			(0.001)	(0.001)
Φ_2			-0.002***	-0.002***
			(0.000)	(0.000)
Φ_3			0.001	0.001***
			(0.001)	(0.000)
Φ_4			-0.001***	-0.001***
			(0.000)	(0.000)
Σ_{11}	0.187***	0.081***	0.512***	0.512***
	(0.001)	(0.023)	(0.007)	(0.002)
Σ_{22}	0.000	0.000	0.005	0.005
	(0.003)	(0.094)	(0.014)	(0.004)
Σ_{33}	0.036***	0.018	0.004	0.006**
	(0.001)	(0.033)	(0.005)	(0.003)
Σ_{44}	0.000	0.000	0.002	0.010***
	(0.004)	(0.090)	(0.009)	(0.003)
log likelihood	-106389	-100617	-82177	

Standard errors in parentheses, brand dummies not reported.
* significant at 10%; ** significant at 5%; ***significant at 1%
1. RCL denotes random coefficient logit. HCS denotes heterogeneous choice set. Number of draws for choice set simulation = 20

Table 7: Predicted Market Shares

Brand Number	Sample Market Share (%)	RCL	RCL+Quality Learning	RCL+Quality Learning+HCS
1	6.03	8.04	7.42	6.64
2	5.07	3.96	3.28	3.72
3	3.63	1.62	1.94	2.38
4	2.84	2.04	2.29	2.57
5	1.56	0.76	0.75	0.92
6	4.67	3.36	4.80	4.18
7	4.04	2.19	3.53	3.44
8	4.11	6.21	3.42	3.52
9	2.32	3.85	1.00	1.13
10	2.75	3.35	2.18	2.25
11	4.56	7.25	5.20	4.26
12	2.61	1.47	2.12	2.91
13	2.12	0.97	0.94	1.64
14	1.82	1.03	1.27	2.42
15	2.47	1.27	2.34	2.97
16	2.32	3.28	1.42	1.98
17	1.49	0.42	0.39	1.62
18	1.78	1.02	1.10	1.12
19	1.47	0.87	1.18	1.43
20	2.84	1.25	1.62	2.03
21	1.50	0.31	0.43	0.56
22	1.19	0.84	0.19	0.54
23	1.72	0.23	0.48	0.65
24	1.25	0.18	0.60	0.99
25	0.76	0.14	0.15	0.25
26	0.93	0.20	0.31	0.42
27	0.53	0.18	0.17	0.18
28	0.51	0.23	0.71	0.41
29	0.46	0.08	0.09	0.13
Brand Number	Sample Market Share (%)	RCL	RCL+Quality Learning	RCL+Quality Learning+HCS
30	0.97	0.14	0.20	0.83
31	0.44	0.04	0.03	0.14
32	0.36	0.00	0.27	0.29
33	0.37	0.00	0.02	0.00
34	0.28	0.01	0.11	0.01
35	0.26	0.00	0.00	0.02
36	0.23	0.00	0.01	0.00
37	0.44	0.08	0.06	0.05
38	1.44	0.31	0.45	0.98
39	0.83	0.00	0.59	0.62
40	0.38	0.02	0.02	0.04
41	0.30	0.01	0.07	0.08
42	1.09	0.21	0.35	0.88
43	0.62	0.07	0.27	0.34
44	0.26	0.01	0.01	0.01
45	0.50	0.23	0.09	0.24

46	0.75	0.18	0.23	0.36
47	0.61	0.10	0.08	0.19
48	0.53	0.04	0.14	0.17
49	0.49	0.12	0.10	0.24
50	0.66	0.04	0.03	0.05
Prediction Error	0	7.26	5.29	3.81

Data: estimation sample for all regressions

Prediction Error = Squareroot of sum of squared deviations of predicted market share to sample markte share

Table 8. Own Price Elasticity for Top 10 Brands

Brand	RCL	RCL & Learning	RCL & Learning & HCS
1	−1.01	−1.27	−2.32
2	−1.27	−0.61	−2.63
6	−0.82	−0.68	−1.71
11	−1.27	−0.65	−1.39
8	−0.98	−0.74	−2.23
7	−0.68	−0.79	−1.66
3	−1.24	−1.04	−1.46
12	−0.98	−1.48	−2.82
4	−1.13	−1.63	−2.42
16	−1.42	−1.59	−2.57

Table 9. Estimated Price Elasticities for Top 25 Brands Based on IV Estimation

Brand #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-2.428	0.367	0.146	0.010	0.099	0.338	0.220	0.315	0.003	0.673	0.034	0.002	0.100	0.001
2	0.136	-2.768	0.265	0.018	0.178	0.612	0.398	0.569	0.005	1.217	0.062	0.003	0.180	0.003
3	0.120	0.594	-1.545	0.015	0.157	0.540	0.352	0.503	0.005	1.075	0.054	0.002	0.159	0.002
4	0.092	0.455	0.179	-3.703	0.120	0.414	0.269	0.385	0.004	0.823	0.042	0.002	0.122	0.002
5	0.070	0.349	0.137	0.009	-2.762	0.317	0.206	0.295	0.003	0.630	0.032	0.001	0.093	0.001
6	0.292	1.446	0.569	0.038	0.383	-1.994	0.856	1.223	0.012	0.815	0.133	0.006	0.387	0.005
7	0.212	1.051	0.414	0.027	0.278	0.955	-1.561	0.889	0.008	1.901	0.096	0.004	0.281	0.004
8	0.249	1.234	0.485	0.032	0.326	1.121	0.730	-2.209	0.010	1.331	0.113	0.005	0.330	0.005
9	0.047	0.232	0.091	0.006	0.061	0.211	0.137	0.196	-1.679	0.420	0.021	0.001	0.062	0.001
10	0.141	0.698	0.275	0.018	0.185	0.635	0.413	0.591	0.006	-4.991	0.064	0.003	0.187	0.003
11	0.045	0.222	0.087	0.006	0.059	0.202	0.131	0.188	0.002	0.401	-1.561	0.001	0.059	0.001
12	0.022	0.109	0.043	0.003	0.029	0.099	0.064	0.092	0.001	0.197	0.010	-2.798	0.029	0.000
13	0.153	0.756	0.298	0.020	0.200	0.688	0.448	0.640	0.006	1.368	0.069	0.003	-3.629	0.003
14	0.022	0.108	0.043	0.003	0.029	0.099	0.064	0.092	0.001	0.196	0.010	0.000	0.029	-1.094
15	0.033	0.165	0.065	0.004	0.044	0.150	0.098	0.140	0.001	0.299	0.015	0.001	0.044	0.001
16	0.044	0.217	0.085	0.006	0.057	0.197	0.128	0.183	0.002	0.392	0.020	0.001	0.058	0.001
17	0.116	0.573	0.225	0.015	0.152	0.520	0.339	0.484	0.005	1.035	0.052	0.002	0.153	0.002
18	0.034	0.168	0.066	0.004	0.045	0.153	0.100	0.142	0.001	0.305	0.015	0.001	0.045	0.001
19	0.078	0.385	0.152	0.010	0.102	0.350	0.228	0.326	0.003	0.697	0.035	0.002	0.103	0.001
20	0.168	0.831	0.327	0.022	0.220	0.755	0.492	0.703	0.007	1.503	0.076	0.003	0.222	0.003
21	0.027	0.136	0.053	0.004	0.036	0.123	0.080	0.115	0.001	0.245	0.012	0.001	0.036	0.001
22	0.014	0.070	0.028	0.002	0.019	0.064	0.042	0.059	0.001	0.127	0.006	0.000	0.019	0.000
23	0.116	0.573	0.225	0.015	0.152	0.521	0.339	0.485	0.005	1.036	0.053	0.002	0.153	0.002
24	0.025	0.126	0.049	0.003	0.033	0.114	0.074	0.106	0.001	0.227	0.012	0.001	0.034	0.000
25	0.031	0.155	0.061	0.004	0.041	0.141	0.092	0.131	0.001	0.280	0.014	0.001	0.041	0.001

Table 9. Estimated Price Elasticities for Top 25 Brands Based on IV Estimation

Brand #	15	16	17	18	19	20	21	22	23	24	25
1	0.001	0.065	0.261	0.027	0.115	0.469	0.004	0.004	0.011	0.003	0.001
2	0.001	0.077	0.306	0.032	0.135	0.551	0.005	0.004	0.013	0.003	0.003
3	0.001	0.014	0.058	0.006	0.025	0.104	0.001	0.001	0.002	0.001	0.004
4	0.001	0.043	0.173	0.018	0.076	0.312	0.003	0.002	0.008	0.002	0.004
5	0.001	0.014	0.055	0.006	0.024	0.099	0.001	0.001	0.002	0.001	0.001
6	0.003	0.007	0.027	0.003	0.012	0.049	0.000	0.000	0.001	0.000	0.002

Brand #	15	16	17	18	19	20	21	22	23	24	25
7	0.002	0.047	0.188	0.020	0.083	0.338	0.003	0.003	0.008	0.002	0.006
8	0.003	0.007	0.027	0.003	0.012	0.048	0.000	0.000	0.001	0.000	0.008
9	0.001	0.010	0.041	0.004	0.018	0.074	0.001	0.001	0.002	0.000	0.001
10	0.002	0.013	0.054	0.006	0.024	0.097	0.001	0.001	0.002	0.001	0.004
11	0.000	0.036	0.142	0.015	0.063	0.256	0.002	0.002	0.006	0.002	0.001
12	0.000	0.010	0.042	0.004	0.018	0.075	0.001	0.001	0.002	0.000	0.001
13	0.002	0.024	0.096	0.010	0.042	0.172	0.002	0.001	0.004	0.001	0.005
14	0.000	0.052	0.206	0.022	0.091	0.371	0.003	0.003	0.009	0.002	0.001
15	−1.351	0.008	0.034	0.004	0.015	0.061	0.001	0.000	0.001	0.000	0.001
16	0.000	−2.409	0.017	0.002	0.008	0.031	0.000	0.000	0.001	0.000	0.001
17	0.001	0.036	−3.788	0.015	0.063	0.256	0.002	0.002	0.006	0.002	0.003
18	0.000	0.008	0.031	−1.305	0.014	0.056	0.000	0.000	0.001	0.000	0.001
19	0.001	0.010	0.038	0.004	−2.384	0.069	0.001	0.001	0.002	0.000	0.002
20	0.002	0.092	0.010	0.041	0.166	−3.958	0.001	0.004	0.001	0.002	0.005
21	0.000	0.167	0.017	0.074	0.300	0.003	−1.434	0.007	0.002	0.004	0.001
22	0.000	0.148	0.015	0.065	0.265	0.002	0.002	−0.944	0.002	0.004	0.000
23	0.001	0.113	0.012	0.050	0.203	0.002	0.002	0.005	−3.178	0.003	0.003
24	0.000	0.087	0.009	0.038	0.156	0.001	0.001	0.004	0.001	−1.171	0.001
25	0.000	0.359	0.037	0.158	0.645	0.006	0.005	0.016	0.004	0.009	−1.409

Table 10. Change in Sales under Alternative Pricing Strategies

Δprice	+1%	+5%	-1%	-5%
brand 28				
Δmarket share(%)	-1.23	-7.09	0.14	6.48
Δsales (%)	-0.78	-2.41	0.01	2.32

Δmarket share = market share in the experiment - market share observed in data

Δsales = sales in the experiment - sales obsered in data

Table 11. Change in Expenditure by Demographic Group under 5% Price Cut

	Δ in expenditure
Highinc	1.21%
Lowinc	3.59%
Old	2.33%
Young	2.32%
Nokid	3.02%
WithKid	1.09%

Table 12. Change in Expenditure by Demographic Group under Pulsing Strate

	Δ in expenditure
Highinc	2.64%
Lowinc	1.59%
Old	1.91%
Young	1.91%
Nokid	1.81%
WithKid	2.15%

Table 13. Food Ads Seen by Children of Different Ages

	Age 2-7	Age 8-12	Age 13-17
Food ads seen per	12	21	17
Food ads seen per year	4,400	6,000	7,600
Percentage of ads seen where food was the main product advertised	32%	25%	22%

Source: "Food for Thought: Television Food Advertising to Children in the United States", The Kaiser Family Foundation, March 28, 2007

Table 14. Sugar and Fiber Contents of Brands by Segment

	Sugar (g per serving)	Fiber (g per serving)
Kid Brands	10.98	5.41
Adult Brands	5.88	9.92
Family Brands	7.68	7.38

Table 15. Change in Segment Share After the Ban

	Δ in mktshare
kid	-5.98%
adult	2.01%
family	3.96%

Table 16. Effects of the Ban across Consumer Groups

	Δ in sugar	Δ in fiber	Δ in expenditure
Highinc	-3.41%	0.46%	6.43%
Lowinc	-5.27%	2.67%	4.36%
Old	-4.22%	0.95%	4.87%
Young	-5.91%	4.24%	8.67%
Nokid	-2.69%	-1.24%	5.15%
Withkid	-6.92%	7.10%	5.37%

Bibliography

Ackerberg D. (2001), "Empirically Distinguishing Informative and Prestige Effects of Advertising", *The Rand Journal of Economics*, Vol. 32, No.2, 316-333.

Ackerberg D. (2003), "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination", *International Economic Review* Vol. 44, No. 3, 1007-1040.

Ackerberg, D., C. Benkard, S.Berry, and A. Pakes (2007), "Econometric Tools for Analyzing Market Outcomes", *Handbook of Econometrics*, Vol. 6A, Chapter 63.

Allenby, G. and J. Ginter (1995), "The Effects of In-Store Display and Feature Advertising on Consideration Sets", *International Journal of Research in Marketing*, Vol. 12, 67-80.

Anand, B and R. Shachar (2005), "Advertising, the Matchmaker", working paper, Harvard University and Tel-Aviv University.

Andrews, R and T. Srinivasan (1995), "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data", Journal of Marketing Research, Vol 32 (Feburary), 30-41.

Bajari, P., C.L. Benkard (2005), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach", *Journal of Political Economy*, Vol. 113, Issue 6, 1239-1276.

Becker, G. and K. Murphy (1993), "A Simple Theory of Advertising as a Good or Bad", *Quarterly Journal of Economics*, Vol.108, 942-64.

Benaissa Chidmi & Rigoberto A. Lopez, 2007. "Brand-Supermarket Demand for Breakfast Cereals and Retail Competition", *American Journal of Agricultural Economics*, Vol.89, Issue 2, 324-37

Berry, S. (1994), "Estimating Discrete Choice Models of Product Differentiation", *Rand Journal of Economics*, Vol. 25, Issue 2, 242-62.

Berry, S., J. Levinsohn, and A. Pakes, (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, Vol. 63, Issue 4, 841-90.

Berry, S., J. Levinsohn, and A. Pakes, (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market", *Journal of Political Economy*, Vol. 112, Issue 1, 68-105.

Berry, S., O. Linton, and A. Pakes, (2004), "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems", *Review of Economics Studies*, Vol. 71, Issue 3, 613-54.

Betancout, R. and C. Clague (1981), *Capital Utilization: A Theoretical and Empirical Analysis*, Cambridge University Press.

Butters, G. (1977), “Equilibrium Distribution of Prices and Advertising”, *Review of Economic Studies*, Vol.44, 465-492.

Costantino, C. (2004), Ph.D. Dissertation, “Three Essays on Vertical Product Differentiation: Exclusivity, Non-Exclusivity and Advertising”, University of Maryland.

Chen, Y. and C. Kuan (2002), “The Pseudo-true Score Encompassing Test for Non-Nested Hypotheses”, *Journal of Econometrics*, Vol. 106, 271-295.

Chintagunta, P., R. Jiang, and G. Jin, (2007) “Information, Learning, and Drug Diffusion: the Case of Cox-2 Inhibitors”, working paper, University of Chicago and University of Maryland.

Crawford, G. and M. Shum, (2005) “Uncertainty and Learning in Pharmaceutical Demand”, *Econometrica*, Vol. 73, 1137-1174

Dubé, J. (2004), "Multiple Discreteness and Product Differentiation: Demand of Carbonated Soft Drinks", *Marketing Science*, Vol.23, Issue1, 66-81.

Erdem, T. and M. Keane (1996), “Decision-Making Under Uncertainty: Capturing Dynamic Brand Choices in Turbulent Consumer Goods Markets”, *Marketing Science*, Vol. 15, 1-20.

Eliaz, K. and R. Spiegler (2007), “Consideration Sets and Competitive Marketing”, working paper, New York University, and University College London.

Geweke, J., M. Keane., and D. Runkle (1994), “Alternative Computational Approaches to Inferences in the Multinomial Probit Model”, *Review of Economics and Statistics*, Vol.76, No.4, 609-32.

Goeree, M. (2008), “Limited Information and Advertising in the US Personal Computer Industry”, *Econometrica*, forthcoming.

Grossman, G. and C. Shapiro, (1984), “Informative Advertising with Differentiated Products”, *The Review of Economic Studies*, Vol.51, 63-81.

Hausman, J. (1996), “Valuation of New Goods under Perfect and Imperfect Competition”, *The Economics of New Goods*, University of Chicago Press, 209-237.

Hendel, I. (1999), "Estimating Multiple-Discrete Choice Models", *Review of Economic Studies*, Vol.66, No.2, 423-446.

— and Nevo, A. (2006), "Measuring the Implications of Sales and Consumer Inventory Behavior", Econometrica, Vol.74, No.6, 1637-1636.

Hitsch, G. (2006), “An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty”, *Marketing Science*, Vol.25, No.1, 25-50.

Hitsch, G., J. Dubé, and P. Manchanda (2005), “An Empirical Model of Advertising Dynamics”, *Quantitative Marketing and Economics,* Vol. 3, No.2, 107-144.

Keane, M. (1994), “A Computationally Practical Simulation Estimator for Panel Data”, *Econometrica*, Vol.61, No.1, 95-116.

Kihlstrom, R. and M. Riordan (1984), “Advertising as a Signal”, *Journal of Political Economy*, Vol. 92, 427-50.

Mehta, N., S. Rajiv and K. Srinivasan (2003), “Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation”, *Marketing Science*, Vol.22, No.1, 58-84.

Milgrom, P. and J. Roberts (1986), “Price and Advertising Signals of Product Quality”, *Journal of Political Economy*, Vol.94, 796-21.

Naik, P., Mantrala M., and A. Sawyer (1998), “Planning Media Schedules in the Presence of Dynamic Advertising Quality”, Marketing Science, Vol.17, No.3, 214-35.

Nelson, P. (1970), “Information and Consumer Behavior”, *Journal of Political Economy*, Vol. 78, No. 2 , 311-329

—, (1974), “Advertising as Information”, *Journal of Political Economy*, Vol. 82, 729-53.

Nevo, A. (2001), “Measuring Market Power In the Ready-to-Eat Cereal Industry”, *Econometrica*, Vol.69, No.2, 307-342.

Osborne, M. (2006), “Consumer Learning, Habit Formation, and Heterogeneity: A Structural Estimation”, Mimeo, Stanford University.

Roberts, J. and J. Lattin (1997), “Consideration: Review of Research and Prospects for Future Insights”, Journal of Marketing Research, Vol.34, Issue 3, 406-410.

Shum, M. (2004), “Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market”, *Journal of Economics & Management Strategy*, Vol.13, No.2, 241-272.

Stern, S. (1997), “Simulation-Based Estimation”, *Journal of Economic Literature*, Vol.35 2006-2039.

Stigler, G. and G. Becker, (1977), “De Gustibus Non Est Disputandum”, *The American Economic Review*, Vol. 67, No. 2, 76-90

Swait, J. (2001), “Choice Set Generation within the Generalized Extreme Value Family of Discrete Choice Models”, *Transportation Research*, Part B 35: 643-666

Swait, J., and T. Erdem, (2007), “Brand Effects on Choice and Choice Set Formation Under Uncertainty”, *Marketing Science*, Vol. 26, Issue 5, 679-697.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press

Vuong, Q. (1989), “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses”, *Econometrica*, Vol. 57, 307-333.