Chapter 5
The Informative Role of Advertising and Experience in Dynamic Brand Choice: An Application to the Ready-to-Eat Cereal Market

Yan Chen and Ginger Zhe Jin

Abstract We study how consumers make brand choices when they have limited information. In a market of experience goods with frequent product entry and exit, consumers face two types of information problems: first, they have limited information about a product's existence; second, even if they know a product exists, they do not have full information about its quality until they purchase and consume it. In this chapter, we incorporate purchase experience and brand advertising as two sources of information and examine how consumers use them in a dynamic process. The model is estimated using the Nielsen Homescan data in Los Angeles, which consist of grocery shopping history for 1,402 households over 6 years. Taking ready-to-eat cereal as an example, we find that consumers learn about new products quickly and form strong habits. More specifically, advertising has a significant effect in informing consumers of a product's existence and signaling product quality. However, advertising's prestige effect is not significant. We also find that incorporating limited information about a product's existence leads to larger estimates of the price elasticity. Based on the structural estimates, we simulate consumer choices under three counterfactual experiments to evaluate brand marketing

This paper is based on Yan Chen's dissertation, "Information, Consumer Choice and Firm Strategy in an Experience Good Market" (August 2008). We are deeply grateful to Mark Denbaly and Ephraim Leibtag at the U.S. Department of Agriculture's Economic Research Service for providing access to the Nielsen Homescan data. We have also benefited from the insightful comments of John Rust, Roger Betancourt, Erkut Ozbay, Pallassana Kannan, and seminars participants at the University of Maryland and the Economic Research Service. All errors are ours.

Y. Chen
Bates White, Washington, DC, USA
e-mail: yan.chen@bateswhite.com

G.Z. Jin (✉)
Department of Economics, University of Maryland, College Park, MD, USA

National Bureau of Economic Research, Cambridge, MA, USA
e-mail: jin@econ.umd.edu

R. Laxminarayan and M.K. Macauley (eds.), *The Value of Information: Methodological Frontiers and New Applications in Environment and Health*,
DOI 10.1007/978-94-007-4839-2_5,

strategies and a policy on banning children-oriented cereal advertising. Simulation suggests that the advertising ban encourages consumers to consume less sugar and more fiber, but their expenditures are also higher because they switch to family and adult brands, which are more expensive.

Keywords Consumer choice • Experience goods • Informative and prestige advertising • Ready-to-eat cereal market • Child-oriented advertising • Childhood obesity

5.1 Introduction

We examine how advertising and experience influence consumers' choice of products in a dynamic setting. In a market with many brands, consumers may not recognize the existence of every brand, especially when there are frequent entries and exits. Even if a consumer knows a brand exists, she may not know all its attributes until she has consumed it. Using ready-to-eat (RTE) cereal as an example, we consider both types of information problems and find that they can be partially addressed by the manufacturer's advertising and the consumer's experience.

The importance of informative and prestige advertising has long been recognized in economics and marketing. Since Ackerberg (2001), a growing literature attempts to empirically distinguish these types of advertising, based on the assumption that informative advertising targets new customers but prestige advertising increases consumption utility for both new and experienced customers (Stigler and Becker 1977; Becker and Murphy 1993). We separate two types of informative advertising: one for indicating the existence of a product (Butters 1977; Grossman and Shapiro 1984), and the other for imparting information about the product's quality (Nelson 1970, 1974; Kihlstrom and Riordan 1984; Milgrom and Roberts 1986).

Both types of informative advertising focus on new consumers but have different implications in consumer choice. By definition, the information about a product's existence brings the product into the choice set, and this effect is the same for all brands, conditional on advertising intensity. In contrast, if consumers are already aware of a product's existence, the information about product quality affects the trade-off between product quality and other observable product attributes; hence its effect on consumer choice will differ across brands. Similar identification has been used by Goeree (2008), but to our knowledge, this paper is the first to distinguish the two types of informative advertising from prestige advertising in a dynamic setting using transaction-level panel data.

In a market of experience goods, dynamic considerations can be important for two reasons: first, experience allows consumers to acquire better knowledge of product attributes. For breakfast cereal, taste and freshness are difficult to ascertain beforehand, but a single instance of consumption could yield plenty of information. Second, past experience may influence the current choice through habit formation. This is different from the informative role of experience because most information about breakfast cereal can be learned by consuming it once, but habit formation may be gradual. Our model also incorporates the potential interaction between experience and advertising: for example, if advertising makes a consumer aware of a brand, the consumer may choose it and form a habit of consuming it. Not only do these effects influence a manufacturer's pricing and advertising strategies,[1] they could also have profound implications for public policy regarding advertising.

Breakfast cereal is heavily advertised toward children, and there has been a long standing debate on whether such advertising should be prohibited. In as early as 1978, the U.S. Federal Trade Commission (FTC) issued a staff report concluding 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, this statement generated strong opposition from broadcasters, ad agencies, and food and toy companies. In 1980 Congress passed the Federal Trade Commission Improvements Act of 1980 and barred FTC from issuing industry-wide regulations to stop unfair advertising practices.[2] However, as concern about childhood obesity grows, policymakers and 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 many food ads every day; of all food ads that target children or teens, 28% are for sugary cereal.[3] Kid brands have significantly more sugar and less fiber per serving than adult and family brands.[4] Based on our empirical estimates, we simulate what would happen to consumer expenditures and nutritional intake if cereal TV advertising directed at children were banned. Results suggest that, following the advertising ban, consumers would consume more fiber and less sugar; this effect is more pronounced for consumers who are younger, have lower income, and have children. Consumers also increase their expenditures after the policy change because they consume more adult and family cereals, which are more expensive than kid cereals.

Although the simulation highlights various roles of advertising and experience, it is worth noting that we do not model the potentially "misleading" effects of advertising. Hence, in our model, the ban of advertising is welfare reducing from the consumer's point of view because the ban leads to a smaller choice set and a less informative choice within the choice set. Since we find little evidence in support of the prestige advertising, our findings rule out the psychological gain from

[1] For instance, 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.

[2] See the article "Limiting Food Marketing to Children," at www.cspinet.org/nutritionpolicy.

[3] See "Food for Thought: Television Food Advertising to Children in the United States," released by The Kaiser Family Foundation, March 28, 2007.

[4] In particular, the average sugar content of kid brands is 10.98 g per serving, compared with 5.88 in adult brands and 7.68 in family brands. The average fiber content of kid brands is 5.41 g per serving compared with 9.92 in adult brands and 7.38 in family brands.

consuming highly advertised brands. However, if advertising misleads consumers into choosing sugary cereals—either because they are unaware of the “unhealthfulness” of the advertised food or because they like the sugary taste without much health consideration into the future—limiting their choice set could be beneficial to them.

The rest of this chapter is organized as follows. Section 5.2 describes the industry background of RTE cereal and summarizes the transactional-level panel data from Nielsen Homescan, manufacturers’ advertising data from TNS Media Intelligence Company, and the brand nutritional data collected from the Internet. Section 5.3 reviews the literature. Section 5.4 lays out the dynamic model of consumer choice, with an emphasis on empirical identification. Section 5.5 reports the estimation results. Section 5.6 describes three counterfactual experiments, two on manufacturers’ pricing strategy and one on a ban of advertising for kids’ brands. A brief conclusion is offered in Sect. 5.7.

5.2 Background and Data

Several features of the RTE cereal market make it suitable for our study.[5] 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 have a truly dominant hold on the market, which imposes a considerable informational burden on consumers.

Using Nielson Homescan data (to be described below), Table 5.1 shows the entries and exits of RTE cereal brands[6] from December 1997 to December 2003 in the Los Angeles market. In the 6-year period, a total of 62 (46) brands enter (exit)[7] 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 5.2 displays sales-based market shares of major brands from December 1997 to December 2003. Because there are so many brands, we select the top 50 (which together account for about 79% of the market) and combine the rest into a composite brand, Brand 51. The biggest brand (Brand 1[8]) has a market share of 6%; most brands take up $<1\%$ of the market.

The third reason the RTE cereal market makes an interesting case is that it is heavily advertised. The advertising-to-sales ratio for RTE cereal was 13% in 2001.

---

[5] Readers can refer to Section 2 of Nevo (2001) for a more complete picture of the RTE cereal industry.

[6] Brand definition follows the classification on each manufacturer’s website. Different box sizes are treated as the same brand, but extensions of a brand name are distinct brands. For example, Cheerios, Honey Nut Cheerios, and Berry Burst Cheerios are three different brands.

[7] Brand entry and exit are defined using the 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.

[8] Brand names are not revealed because of a confidential agreement with the data provider.

Table 5.1 Brand entry and exit

	Exit year						
Enter year	1998	1999	2000	2001	2002	Remaining	Total
1998 and 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, December 1997 to December 2003, Los Angeles market
A brand entry is observed if the first transaction with the brand occurred after June 1998. A brand exit is observed if the last transaction with the brand occurred before July 2003

Table 5.2 Brand summary statistics

Brand number	Sample market share[a](percentage)	Average transaction price (cents/oz.)	Average monthly advertising ($k)	Fiber content (percentage daily value per 30g)	Sugar content (percentage daily value per 30g)	Segment[b]
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

(continued)

Table 5.2 (continued)

Brand number	Sample market share[a] (percentage)	Average transaction price (cents/oz.)	Average monthly advertising ($k)	Fiber content (percentage daily value per 30g)	Sugar content (percentage daily value per 30g)	Segment[b]
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
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[c]	21.41	14.90	47.44	8.83	8.00	Family

Data source: Columns II, III from Neilsen Homescan data, December 1997 to December 2003; column IV from TNS Media Intelligence data, January 1999 to December 2003; columns V, VI from www.nutritiondata.com

[a]Sample market is the Los Angeles market from December 1997 to December 2003

[b]Brand segment categorization is based on each brand's description on the manufacturer's website

[c]Characteristics of the 51st brand are computed as the average of the nontop 50 brands

For well-established brands, the ratio was 18%.[9] In comparison, the average ad-to-sales ratio across 200 industries was 3.2%.[10] Heavy advertising indicates that firms believe advertising is effective in promoting sales.

Our data consist of four parts. On the consumer side, we use the 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 tie consumer purchasing behavior with demographic measures. Homescan panelists scan items at home from each shopping trip, recording price and quantity purchased as

[9] See Nevo (2001, 311).

[10] See *Advertising Age,* March 1, 2006.

Table 5.3 Summary statistics of Homescan data

Variable	Definition	NumObs	Mean	Std. Dev.	Min	Max
size	household size	1,402	3.25	1.53	1	9
inc	household income ($K)	1,402	57.11	29.58	2.5	125
age	age of female household head	1,402	48.86	12.99	20	70
nokid	=1 if no kid in the household	1,402	0.55	0.50	0	1
price	transaction price (cent/oz)	69,134	17.84	4.73	0	797.44

Data source: Neilsen Homescan data, December 1997 to December 2003, Los Angeles market

well as the age, income, and other demographic information of the shopper. When available, Nielson uses store-average price instead of the consumer's self-recorded transaction price. Einav et al. (2010) document the measurement error in this data set and conclude that the magnitude of errors in the Homescan data is comparable to that of commonly used, government-collected economic data sets.

Homescan keeps track of on-going purchasing from the same household over time and thus 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 5.3 contains definitions and summary statistics of the major variables in the Homescan data.

Using the Homescan data, we can summarize the consumption pattern in the RTE cereal market. On average, a household makes 14 shopping trips for RTE cereal per year. The households usually have two or three brands that they purchase repeatedly over time. Most brands are purchased once and never again (Fig. 5.1). After a brand is first purchased by a household, the probability of the household's repurchasing the brand is 14.1% on the next shopping trip, 12.9% on the second trip, and about 11% on the following trips (Fig. 5.2). This suggests that learning in the cereal market is mainly done after one shopping trip. Figure 5.2 also suggests that a household that repurchases a brand after the first experience then exhibits loyalty to that brand.

On the product side, we obtain advertising data from TNS Media Intelligence, which tracks advertising expenditures of cereal manufacturers from January 1999 to December 2003. The advertising data cover 278 cereal brands across 11 media types.[11] The brand advertising expenditures include both national 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, whereas 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 5.2, above.

[11] The 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 because of a lack of data on retailers.

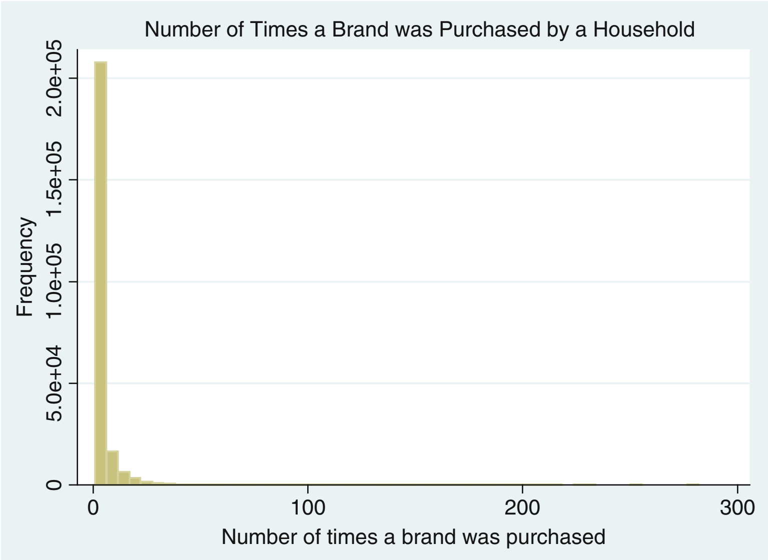

Fig. 5.1 Purchase frequency per brand

The third part of the data is nutritional information on 111 cereal brands, collected from www.nutritiondata.com[12]; it includes calories, sugar, dietary fiber, protein, and so forth. The fiber and sugar content per 30-g serving for the 50 top brands is displayed in columns 4 and 5 of Table 5.2. 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.

The fourth part of the data involves cost factors that could serve as instruments to address the potential endogeneity of price and advertising. From the website of the Bureau of Labor Statistics, we collect 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 metropolitan statistical area from 1999 to 2003. Corn and wheat prices are obtained from the Farmdoc project of the University of Illinois. Gasoline and electricity prices are collected from the website of the Energy Information Administration, an office in the Department of Energy. As detailed in Sect. 5.4.4, these cost factors are likely to correlate with a firm's decisions about price and advertising but are uncorrelated with unobserved demand for RTE cereal.

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

5.3 Literature

Several lines of literature are relevant to our inquiry. 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 a detergent's quality, and consumers update their beliefs about quality in a Bayesian way. Following this research are many studies that model consumer learning in a Bayesian framework in various markets (e.g., Ackerberg 2003; Crawford and Shum 2005; Chintagunta et al. 2009). However, the consumer learning literature usually takes the consumer choice set as homogeneous. It does not account for the fact that different consumers may be exposed to different sets of products because of 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 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, and at the second stage they choose the brand with the highest utility. 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 chapter. 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 et al. (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 (2011) study a market model in which firms use irrelevant alternatives to influence consumers' consideration set. All these studies, however, model one-time purchases 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 chapter learns from both theoretical (as cited in the introduction) and empirical literature on advertising (e.g., Ackerberg 2001, 2003; Anand and Shachar 2011). This chapter also benefits from the insights of the literature on RTE cereal market that involves demand estimation (Hausman 1996; Nevo 2001; Shum 2004; Hitsch 2006) and simulation of counterfactual pricing and advertising strategies (Dubé et al. 2005). Compared with the previous studies, the richness of the data allows us 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 Osborne 2006), and distinguish different effects of advertising.

Last but not the least, this chapter is an extension of the literature on analyzing demand systems in differentiated product markets (Berry 1994; Berry et al. 1995, 2004a, b). 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 chapter applies the estimation method to a limited information environment.

5.4 Model

5.4.1 Setup

Consider 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. 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 he 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 taste, fiber, and sugar content. The vertical differentiation is on brand quality, including the quality of ingredients such as types of grains and rice, the processing techniques, and freshness.

Before going into the details of the model, we mention two simplifications implicit in the above framework. First, we focus on consumer brand choice conditional on purchasing RTE cereal. There are two reasons for not including nonpurchase of RTE cereal as the outside good. Consumers may choose not to purchase because they have cereal at home, not because the utility of nonpurchase is higher than all cereal brands. Treating nonpurchase as the outside good, therefore, would bias the parameter estimates downward in the utility function. In addition, consumers choose not to purchase RTE cereals on about two-thirds of all shopping trips. Including those shopping trips will further add to the already large computation burden.

The second simplification of the model is the absence of 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).[13] For example, if on a shopping trip a consumer purchased brands A, B, and C, it is estimated as if he 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, since store-level data are not available.

Now we return to the first stage of the model. 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, he never forgets about it, even though he may dislike it and choose not to purchase it again. Second, the probability of consumers' being informed of a previously untried 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

$$P(C_{it}) = \prod_{j \in C_{it}} q_{ijt} \prod_{k \notin C_{it}} (1 - q_{ikt}) \tag{5.1}$$

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

$$q_{ijt} = \begin{cases} \dfrac{\exp(\varphi_0 + \varphi_1 adv_{jt} + \varphi_2 adv_{jt} inc_i + \varphi_3 adv_{jt} nokid_i + \varphi_4 adv_{it}^2)}{1 + \exp(\varphi_0 + \varphi_1 adv_{jt} + \varphi_2 adv_{jt} inc_i + \varphi_3 adv_{jt} nokid_i + \varphi_4 adv_{it}^2)}, \forall j \notin E_{it} \\ = 1, \forall j \notin E_{it} \end{cases} \tag{5.2}$$

where E_{it} is consumer i's experience set as of time t—that is, 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,

$$adv_{jt} = \sum_{\tau=0}^{T} \delta^{\tau} a_{jt-\tau} \tag{5.3}$$

[13] Shum (2004) fails to find across-brand synergies in demand patterns of RTE cereals that would require modeling the multiple-brand purchase decision. See Hendel (1999) and Dubé (2004) for examples of a multiple-discrete choice model that allows multiple-unit and multiple-brand purchases on one shopping trip; and see Hendel and Nevo (2006) for an example of a consumer inventory model. Multiple brand purchases on one shopping trip are treated as independent events.

where a_{jt} denotes brand j's advertising expenditure at time t,[14] 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 goodwill stock. Specifically, if a brand entered a consumer's choice set on the previous shopping trip but was not purchased, the probability of its reentering the consumer's current choice set may still be high even if the brand is not advertised in the current period, because of the lagged effects of previous advertising. The termadv_{jt}^2 is included to account for the potential increasing or decreasing returns to scale of advertising. In Eq. (5.2), adv_{jt} is also interacted with household income and whether there are any children 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.

$$\begin{aligned} U_{ijt} = {} & E(X_j)\beta_i + \alpha_i price_{ijt} + \rho_i adv_{jt} + \kappa \cdot unused_{ijt} + \lambda_i \cdot unused_{ijt} \cdot adv_{jt} \\ & + pastchoice_{ijt} \bullet \gamma + \eta_{jt} + \varepsilon_{ijt} \end{aligned} \tag{5.4}$$

where $X_j = [fiber\ sugar]_j$, $\beta_i = [\beta_{1i}\ \beta_{2i}]'$, $price_{ijt}$ is the price of brand j when consumer i it at time t. In the Nielsen Homescan data, the price of a brand is recorded as the weekly average price of that brand in the store where the brand was sold. In the estimation, we subtract the manufacturer's coupon value and the retailer's deal value from the price if a coupon or a deal is used.[15]

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

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \alpha_i \\ \rho_i \\ \lambda_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \alpha \\ \rho \\ \lambda \end{bmatrix} + \Pi \bullet D_i + \Sigma \bullet v_i \tag{5.5}$$

where D_i is a vector of observed household characteristics, including household income, age of female household head, and presence of children; v_i represents a vector of unobserved household characteristics with standard normal distribution.

[14] Advertising data are monthly; purchase data are daily. Therefore advertising expenditure at time t means advertising expenditure in the month that day t belongs to. In the empirical results, reported in Sect. 5.5, $\delta = 0.95$ and $T = 6$. We 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.

[15] We are not able to control for coupons and deals systematically, as in Nevo and Hendel (2006), because we do not have store-level data and do not observe the availability of coupons and deals to consumers.

The variable $unused_{ijt}$ is a dummy equal to 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 produce cereals with better ingredients and better technology. The vector $pastchoice_{ijt} = [chosen_{ijt-1}\ chosen_{ijt-2}, \ldots, chosen_{ijt-\tau}]$, where $chosen_{ijt-\tau}$ equals 1 if brand j was chosen τ shopping trips before t.[16] The term η_{jt} represents brand j's characteristics that are observable to the consumer but not to the researcher at time t. In the case of RTE cereals, η_{jt} encapsulates packaging, shelf space, etc. Lastly, ε_{ijt} is a mean-zero 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 content according to the following rule: $E(fiber_j) = \text{mean}(fiber_k)$, and $E(sugar_j) = \text{mean}(sugar_k)$ $\forall$ brand k tried by consumer i before and belonging to the same segment as brand j . Following Hausman (1996) and Shum (2004), we divide the brands into family, adult, and kid segments. The segment categorization is shown in column 7 of Table 5.2. 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.

$$P_{ijt} = \sum_{C_{it} \in S} \prod_{j \in C_{it}} q_{ijt} \prod_{k \notin C_{it}} (1 - q_{ikt}) P(j|C_{it}), \tag{5.6}$$

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 by maximum likelihood yields the parameter estimates.

5.4.2 Discussion

Several features of the demand model merit additional discussion. First, the choice set formation process addresses the informational problem about a product's 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 brings the model closer to real

[16] In the empirical results I use T = 6. Compared with 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.

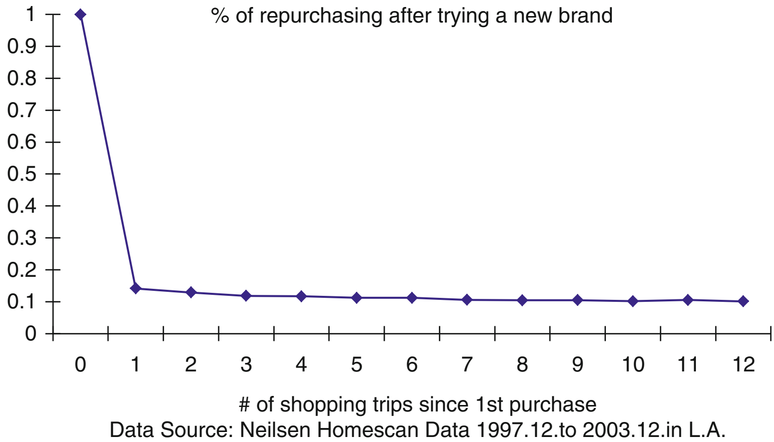

Fig. 5.2 Repurchase pattern

consumer behavior. Since the choice set is not observable in the data, we simulate them in the estimation. The details of simulation will be discussed in Sect. 5.4.4.

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 Fig. 5.2. Unlike some complicated products, consumers usually attain precise knowledge about a cereal after consuming one box of it.

Third, compared with most previous choice models, where only the past choice is included, choices on the previous six 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 most recent 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 she 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' effect 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: (1) affecting consumer choice set, which represents advertising's informative effect on brand existence and is captured by the φ parameters; (2) signaling quality of an unused brand, which represents advertising's informative effect on brand quality and is captured by the parameter λ; (3) 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, we assume that the consumer is myopic and maximizes her current utility. When state dependence (habit formation) is present, a forward-looking consumer considers the future effects of her current choice. Forward-looking behavior is important in many cases, especially in situations where the experimentation cost is high, as in the choice of durable goods (computer, digital camera) and decisions about whether to accept a job offer or continue searching. It is less critical in this situation, where consumers choose a frequently purchased product and the cost of trying a new product is low because they can easily switch back to previous brands. Marketing research shows that consumers spend an average of 13 s in selecting a brand out of the shelf[17]—a very short time for a consumer to make choices. Therefore, we tend to believe that the myopic assumption is reasonable in this application and in the choice process of many other nondurable goods, such as beverages and cosmetics.

5.4.3 Identification

The parameters to be estimated (denoted as θ) include φ_0, φ_1, φ_2, φ_3, φ_4. β, α, ρ, $\prod$, $\sum$, κ, λ, and γ. Variation of brand choices corresponding to observed brand characteristics, price, and advertising for all consumers is used to identify β, α, and ρ. A 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 $\sum$. 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.

The main identification assumption of the prestige effect is that it does not vary by consumption experience. As in Ackerberg (2001, 2003), the prestige effect affects both experienced and inexperienced consumers in the same way, but the informative effects works only on consumers who have never tried the brand before. Therefore, variation in the ratio of the choice probability between experienced and inexperienced consumers as advertising intensity changes can be used to distinguish the informative effect from the prestige effect. The two types of informative effect (coefficient φ versus coefficient λ) both affect the choice probability of inexperienced consumers. An inexperienced consumer may choose to try a brand because advertising alerts him 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 provides information only about brand existence, consumers will include the brand in their

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

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: there are fewer consumers on the margin who would switch to the brand because of 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 because more consumers would be persuaded 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, since advertising enters the utility function and interacts with the brand characteristics only if the informative effect about brand quality exists.

To see this mathematically, let us consider a simple example with two brands in the market. Brand 1 has been established for a long time, and 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 it is

$$P = \frac{\exp(\varphi_0 + \varphi_1 adv_2)}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} * \frac{\exp(E(X_2)\beta + \alpha^* price_2 + \Psi)}{1 + \exp(E(X_2)\beta + \alpha^* price_2 + \Psi)} \tag{5.7}$$

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

$$\frac{\partial \ln(P)}{\partial(adv_2)} = \frac{\varphi_1}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} \tag{5.8}$$

Note that Eq. (5.8) is independent of Brand 2's characteristics. If advertising also provides information about quality, the choice probability of Brand 2 is

$$P = \frac{\exp(\varphi_0 + \varphi_1 adv_2)}{1 + \exp(\varphi_0 + \varphi_1 adv_2)} * \frac{\exp(E(X_2)\beta + \alpha^* price_2 + \rho^* adv_2 + \Psi)}{1 + \exp(E(X_2)\beta + \alpha^* price_2 + \rho^* adv_2 + \Psi)} \tag{5.9}$$

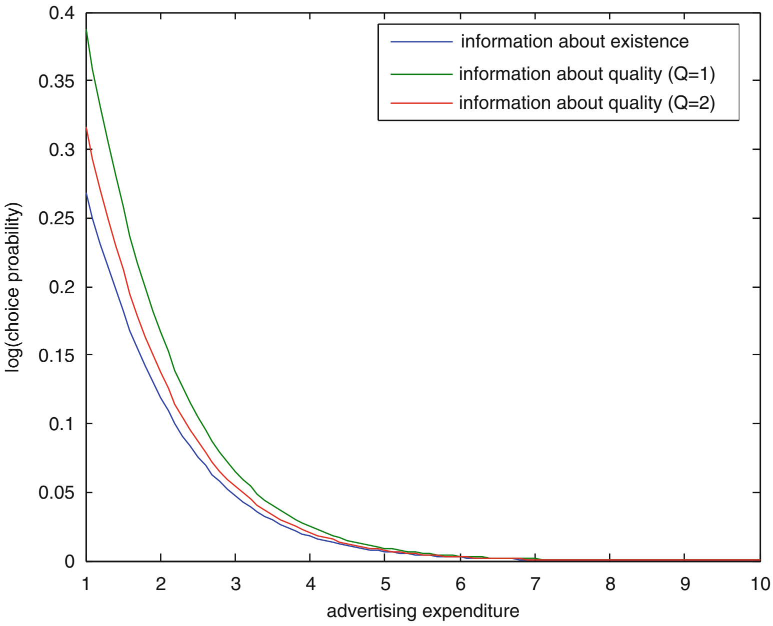

Fig. 5.3 Effect of advertising on marginal change in choice probability

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

$$\frac{\partial \ln(P)}{\partial(adv_2)} = \frac{\varphi_1}{1+\exp(\varphi_0+\varphi_1 adv_2)} + \frac{\rho}{1+\exp(E(X_2)\beta+\alpha^* price_2+\rho^* adv_2+\Psi)} \tag{5.10}$$

The higher the utility consumers infer from the brand characteristics, the less the need to rely on the information in advertising. Comparing Eqs. (5.8) and (5.10), we can see that whether the marginal effect of advertising on choice probability depends on the brand characteristics distinguishes the informative effect about brand quality from the informative effect about brand existence. To illustrate this point, Fig. 5.3 depicts the marginal effect of advertising on choice probability. The nonstochastic 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 a function of Q. As Q increases, the marginal change in choice probability decreases.

5.4.4 Estimation Issues

Here we discuss four estimation issues: unobserved consumer heterogeneity, choice set simulation, property of simulators, and the potential endogeneity of price and advertising.

5.4.4.1 Unobserved Consumer Heterogeneity

In a model with lagged dependent variables, state dependence (habit formation) is observationally equivalent to consumer heterogeneity because 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 the consumer's 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 we use consumer-brand random effects to control for unobserved consumer heterogeneity. The details of the implementation are provided in Appendix 5.1.

5.4.4.2 Choice Set Simulation

To address the informational problem about brand existence, we allow for heterogeneity in consumer choice sets. The underlying choice sets over which consumers make utility comparisons 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, we simulate the choice sets. In the simulation, the probability of a brand's being included in a consumer's choice set is a function of brand advertising and purchase experience according to Eq. (5.2). The details of the choice set simulation process are provided in Appendix 5.2.

5.4.4.3 Estimation Procedure Without Instruments

After simulating the choice sets, we 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

$$SL(\theta) = \prod_{i} \prod_{t} \hat{p}_{ijt}(\theta)^{Y_{ijt}} \tag{5.11}$$

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

$$SLL(\theta) = \sum_i \sum_t Y_{ijt} \log(\hat{p}_{ijt}(\theta)) \tag{5.12}$$

The MSL estimator $\widehat{\theta}$ is a vector of parameters that maximize Eq. (5.12). Train (2003) shows that if the number of simulation draws rises faster than the square root of sample size, then the MSL estimator is not only consistent but also asymptotically equivalent to the maximum likelihood estimator.[18] Specifically, the MSL estimator is distributed

$$\hat{\theta} \overset{a}{\sim} N(\theta^*, -H^{-1}/N) \tag{5.13}$$

where $\theta^*$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, we use $\hat{H} = \frac{\partial^2 SLL(\hat{\theta})}{\partial\theta\partial\theta'}$ to approximate the value of H and calculate the estimated variance.

5.4.4.4 Endogenous Price and Advertising

If the manufacturer sets up prices and advertising levels according to consumers' willingness to pay, then an endogeneity problem may arise, since 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. It is worth noting that we include brand fixed effects to control for unobserved brand characteristics invariant over time. For example, if 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 made with whole grains is invariant over time and is absorbed by brand dummies. However, unobserved time-varying brand characteristics, such as shelf space, are not absorbed by brand dummies and could create endogeneity.

One way to deal with the endogeneity problem is using instrumental variables (IV). Competition among differentiated products 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 exogenous, then the

[18] 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.

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 (ingredient prices, wages of food workers) are candidates for instruments, too.

In the nonlinear discrete choice model, IV estimation cannot be directly implemented on the consumer-level data. Following Berry et al. (1995, 2004), we first aggregate individual consumer choices into market shares and then match predicted and observed brand market shares to recover the component of utility 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 $\chi_{jt} = \left[fiber_{jt} sugar_{jt} price_{jt} adv_{jt}\right]$,[19]

$$\beta_{\mathrm{i}} = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \alpha_i \\ \rho_i \end{bmatrix} = \bar{\beta} + \Pi \bullet D_i + \Sigma \bullet v_i, \quad \text{where} \quad \bar{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \alpha \\ \rho \end{bmatrix},$$

$\mathrm{D_i} = [income_i\ age_i\ nokid_i]'$, and $\mathrm{v_i} = [\mathrm{v_{1i}}\ \mathrm{v_{2i}}\ \mathrm{v_{3i}}\ \mathrm{v_{4i}}]'$.

Then we can write the utility as

$$U_{ijt} = \chi_{jt}\beta_i + \kappa \cdot unused_{ijt} + \lambda_i \cdot unused_{ijt} \cdot adv_{jt} + pastchoice_{ijt} \bullet \gamma + \eta_{jt} + \varepsilon_{ijt}\,. \tag{5.20}$$

Let

$$\delta_{jt} = x_{jt}\bar{\beta} + \eta_{jt}. \tag{5.21}$$

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 the effect of η from the effect of χ on δ. To provide consistent estimates of $\bar{\beta}$, we need IV for price and advertising.

We use two sets of instruments. The first set includes the fiber and sugar content of all other brands. The second set of instruments comprises the cost shifters, including wage of food workers, wage of advertising managers, corn price, wheat

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

price, gasoline price, and electricity price. The data sources of these cost factors are described in Sect. 5.2.

The IV estimation involves three sets of moment conditions: (1) the model's predicted brand choice probabilities are matched to observed individual brand choices; (2) the model's prediction for brand j's market share in year t is matched to its observed market share in year t; and (3) the unobserved time-varying brand characteristics are assumed to be orthogonal to all the observed brand attributes and the instruments.

More specifically, the estimation algorithm consists of four steps.

Step 1. Given an initial guess of Π, Σ, κ, λ, and γ, we first find the values of δ_{jt} that equate 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 5.3.

Step 2. Given δ_{jt}, we 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 $\gamma$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 that a household with observed characteristics D_i will choose brand j given δ,Π,Σ,κ, λ, and γb is given by

$$\Pr(j|D_i, \delta, \Pi, \Sigma, \kappa, \lambda, \gamma) = \int_v \frac{\begin{array}{l}\exp(\delta_{jt} + \chi_{jt} g \Pi g D_i + \chi_{jt} g \Sigma g v + \kappa \cdot unused_{ijt} \\ + \lambda_i \cdot unused_{ijt} \cdot adv_{jt} + pastchoice_{ijt} \bullet \gamma)\end{array}}{\begin{array}{l}\sum_{k=1}^{51} \exp(\delta_{kt} + \chi_{kt} g \Pi g D_i + \chi_{kt} g \Sigma g v \kappa \cdot unused_{ijt} \\ + \lambda_i \cdot unused_{ijt} \cdot adv_j + pastchoice_{ijt} \bullet \gamma)\end{array}} f(v) d(v) \tag{5.22}$$

The integrals are computed by simulation.

Step 3. 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.

Step 4. Using the δ_{jt} obtained in step 3,construct the moment condition $E(\eta_{jt}|z) = 0$, where $\eta_{jt} = \delta_{jt} - \chi_{jt}\bar{\beta}$ and z represents instrument variables, and estimate $\bar{\beta}$ by minimizing the sample moments $G(\bar{\beta}) = \sum_j \eta_j Z_j$[20]

[20] Both Berry, Levinsohn and Pakes (2004) and 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.

5.5 Results

This section presents the demand estimates with and without instrumental variables. We carry out the demand estimation based on the panel data in the Los Angeles RTE cereal market.[21] 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 is used to construct its experience set; households staying in the Homescan panel for <12 months are dropped. The unit of observation in the estimation is a transaction—that is, a household-purchase date-brand combination. Observations with missing values on key estimation variables are dropped.[22] The regressions begin with July 1999, since the earliest advertising data are available in January 1999, and to calculate the advertising stock, we need advertising data for the previous 6 months. The estimation sample consists of 844 households and 37,858 transactions and remains unchanged in all specifications. Values of the key variables in the estimation sample are summarized in Table 5.4. In all specifications 50 brand dummies are used.

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

5.5.1 *Estimation Without Instrumental Variables*

After the variable selection guided by the conditional logit, we run three random coefficient logit models with different informational assumptions. First, we 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 untried brands. In the third specification, consumers are assumed to have limited information about both the quality of untried brands and the brand existence. A random coefficient logit with quality learning and heterogeneous choice sets is estimated.

[21] The modeling technique and estimation method in this paper are not specific to a particular geographical market or a particular experience good. We can apply the model to environments where consumers face the two types of informational problems—for example, consumer choice of cosmetics, credit cards, and health care plans.

[22] The missing values do no happen systematically, so we are not concerned with a selection bias.

Table 5.4 Summary of variables in estimation sample[a]

Variable	Definition	Mean	Std. Dev.	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($K)	1031.84	629.38	0	13,000
price*nokid	transcation price(cent/oz)*1{household has nokid}	892.50	348.03	0	7,400
adv	stock of advertising expenditure ($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	4,060
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

[a]Estimation sample consists of 890 households with 42,396 transactions from January 1999 to December 2003 in Los Angeles market

Note that the three specifications are nonnested. To compare them, ideally we would like to construct a test statistic with a limiting distribution. However, our panel data do not satisfy the distributional assumptions of tests for nonnested models (e.g., Vuong 1989 and Chen and Kuan 2002). Therefore, to assess the goodness of fit, we use two methods. First we compare the different specifications using the Akaike information criterion (AIC) and using a measure of predictive performance developed by Betancourt and Clague (1981). Then we construct a variable that measures market share prediction errors of the three specifications to see how well they predict consumer choices.

5.5.1.1 Estimation with Full Information

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

The parameter estimates of the benchmark specification are reported in column I of Table 5.5. Price is negative and significant. The price sensitivity decreases as household income increases and if the household has no children. On average, advertising's prestige effect is negative and marginally significant. But the prestige effect increases as income grows and when there are no children in the household. The *unused* (untried) variable 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 children. 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 they were trying the brand for the first time. Both *fiber* and *sugar* are negative and significant. Older consumers without children prefer more fiber and less sugar.

5.5.1.2 Estimation with Limited Information About Brand Quality

In the second specification, we run a random coefficient logit regression where all consumers face the same choice set of 51 brands but do not know the quality of brands not bought before. Consumers form expectations 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 5.5) are the same as those of the benchmark regression, and for most coefficients the magnitudes are

Table 5.5 Estimation results

	I	II	III	IV
	RCL	RCL + Learning	RCL + Learning + HCS	IV Estimation
price	−0.164***	−0.187***	−0.233***	−0.368***
	(0.003)	(0.054)	(0.009)	(0.051)
price*inc	0.002***	0.001**	0.001***	0.001***
	0.000	(0.001)	0.000	0.000
price*nokid	0.127***	0.129***	0.118***	0.014***
	(0.003)	(0.007)	(0.007)	(0.005)
adv	−0.007*	−0.018	−0.023	0.274
	(0.004)	(0.065)	(0.050)	(1.731)
adv*inc	0.001**	0	0.001***	0.000***
	0.000	(0.001)	0.000	0.000
adv*nokid	0.073***	0.009	0.073***	−0.003
	(0.003)	(0.006)	(0.003)	(0.003)
unused	−1.874***	−2.071***	−1.800**	−1.801***
	(0.021)	(0.017)	(0.043)	(0.091)
unused*adv	0.335***	0.083***	0.286***	0.696***
	(0.006)	(0.010)	(0.023)	(0.041)
unused*adv*inc	−0.004**	0	0.027***	0.028***
	0.000	(0.001)	(0.001)	(0.001)
unused*adv*nokid	−0.155***	−0.041	−0.019**	−0.035
	(0.005)	(0.030)	(0.009)	(0.083)
chosen1	0.614***	0.649***	0.578***	0.563***
	(0.017)	(0.018)	(0.018)	(0.021)
chosen2	0.638***	0.629***	0.603***	0.590***
	(0.017)	(0.018)	(0.018)	(0.021)
chosen3	0.612***	0.574***	0.577***	0.565***
	(0.017)	(0.018)	(0.019)	(0.019)
chosen4	0.548***	0.504***	0.514***	0.503***
	(0.017)	(0.018)	(0.019)	(0.019)
chosen5	0.531***	0.466***	0.497***	0.488***
	(0.017)	(0.019)	(0.019)	(0.019)
chosen6	0.534***	0.468***	0.501***	0.491***
	(0.017)	(0.019)	(0.019)	(0.019)
fiber	−0.112***	0.014	−0.068***	4.466***
	(0.005)	(0.046)	(0.011)	(0.853)
fiber*age	0.001***	0.001	0	0.001***
	0.000	(0.001)	0.000	0.000
fiber*nokid	0.018***	0.037***	0.047***	0.026***
	(0.003)	(0.011)	(0.005)	(0.004)
sugar	−0.083***	−0.06	−0.099***	1.996
	(0.009)	(0.059)	(0.020)	(2.293)
sugar*age	0.001	−0.001	0	−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)
φ0			−7.350***	−7.350***

(continued)

Table 5.5 (continued)

	I	II	III	IV
	RCL	RCL + Learning	RCL + Learning + HCS	IV Estimation
			(0.005)	(0.001)
φ1			2.996***	2.996***
			(0.001)	(0.001)
φ2			−0.002***	−0.002***
			0.000	0.000
φ3			0.001	0.001***
			(0.001)	0.000
φ4			−0.001***	−0.001***
			0.000	0.000
Σ11	0.187***	0.081***	0.512***	0.512***
	(0.001)	(0.023)	(0.007)	(0.002)
Σ22	0	0	0.005	0.005
	(0.003)	(0.094)	(0.014)	(0.004)
Σ33	0.036***	0.018	0.004	0.006**
	(0.001)	(0.033)	(0.005)	(0.003)
Σ44	0	0	0.002	0.010***
	(0.004)	(0.090)	(0.009)	(0.003)
log likelihood	−106,389	−100,617	−82,177	

RCL represents random-coefficient logit model. Learning uses expected product attributes for untried (“unused”) brands in the utility function. *HCS* stands for heterogeneous choice set. ***p<0.01, **p<0.05, *p<0.1

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 untried brand is low, thus uncertainty about brand quality may not be an important factor when consumers decide which brand to buy.

5.5.1.3 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 untried 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 Eqs. (5.1) and (5.2), and the choice set is simulated as described in Appendix 5.2.

The price coefficients (column III of Table 5.5) 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, we 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 percentage 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 5.6. Compared with 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 the idea 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 case of full information. 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 the price is reduced. Therefore, if the information problem about a product's existence is alleviated, the market should be more competitive because 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 Eq. (5.2)) is positive and significant, whereas φ_4 (coefficients on adv^2 in Eq. (5.2)) 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 fewer TV commercials, but once they are alerted to the availability of an untried brand, they rely more on advertising to obtain the quality information than other methods of searching. The coefficient on adv*nokid in choice set formation, φ_3, is positive but not significant, implying that the effect of advertising does not vary with the presence of children. Figure 5.4 plots the probability of a brand's entering a consumer's choice set against the brand's advertising expenditure evaluated at the mean level of household income and presence of children. At the mean of advertising stock ($3.22 million), the probability of that a brand is included in the choice set is 88%. Increasing advertising stock by $1 million from the mean will result in a 99% probability 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.

Table 5.6 Predicted market shares

Brand number	Sample market share (percentage)	RCL	RCL + QualityLearning	RCL + QualityLearning + 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	4.8	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	1.13
10	2.75	3.35	2.18	2.25
11	4.56	7.25	5.2	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	1.1	1.12
19	1.47	0.87	1.18	1.43
20	2.84	1.25	1.62	2.03
21	1.5	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	0.6	0.99
25	0.76	0.14	0.15	0.25
26	0.93	0.2	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
30	0.97	0.14	0.2	0.83
31	0.44	0.04	0.03	0.14
32	0.36	0	0.27	0.29
33	0.37	0	0.02	0
34	0.28	0.01	0.11	0.01
35	0.26	0	0	0.02
36	0.23	0	0.01	0
37	0.44	0.08	0.06	0.05
38	1.44	0.31	0.45	0.98
39	0.83	0	0.59	0.62
40	0.38	0.02	0.02	0.04
41	0.3	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

(continued)

Table 5.6 (continued)

Brand number	Sample market share (percentage)	RCL	RCL + QualityLearning	RCL + QualityLearning + HCS
45	0.5	0.23	0.09	0.24
46	0.75	0.18	0.23	0.36
47	0.61	0.1	0.08	0.19
48	0.53	0.04	0.14	0.17
49	0.49	0.12	0.1	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 square root of sum of squared deviations of predicted market share to sample market share, *RCL* random-coefficient logit model, *Learning* using expected product attributes for untried ("unused") brands in the utility function, *HCS* heterogeneous choice set

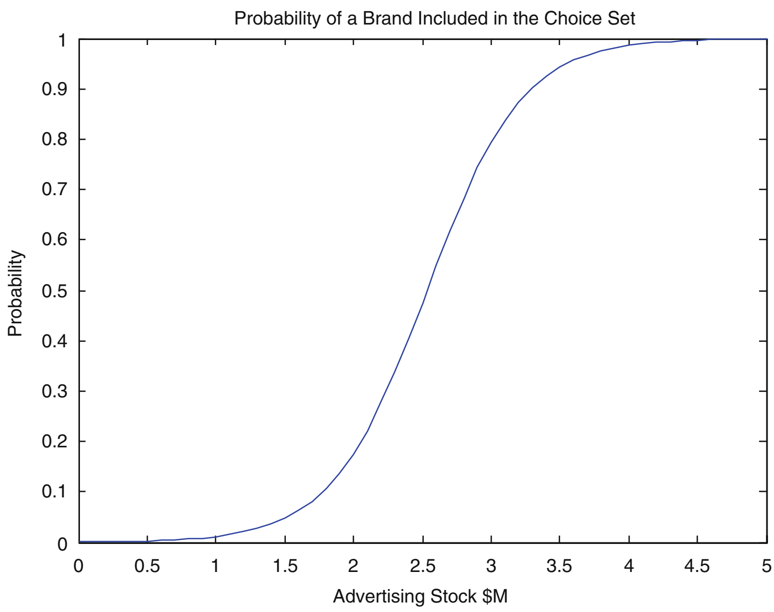

Fig. 5.4 Relationship between advertising stock and the probability of being included in the choice set

The past choice variables are still positive and significant, suggesting that consumers form persistent habits in cereal purchases. Compared with 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.

5.5.1.4 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, which equals 2*k−2lnL, where k is the number of parameters and lnL is the log likelihood. The AIC imposes a penalty 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, 201386, and for the third, 164516. Hence, according to the AIC, the third specification fits the data best.

Second, we 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.[23] 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 $E_{\max} = -\frac{1}{51}\log(\frac{1}{51})$ [24] 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 $\bar{I} = (I_1 - I_2)/N$, 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 $\bar{I}$ predicts the data best.[25] Applying the formula to our data, we find that the $\bar{I}$ for the first specification is −11.5, for the second one, −13.2, and for the third, 0.8.[26] Again, the third specification represents the best fit.

[23] For a more detailed discussion of the measure, refer to Betancourt and Clague (1981, Section 4.6). 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.

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

[25] 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 our goal is to compare only the three specifications, we do not calculate the other measures. Interested readers should refer to Betancourt and Clague's book for more information.

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

Table 5.7 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

RCL random−coefficient logit model, *Learning* using expected product attributes for untried ("unused") brands in the utility function, *HCS* heterogeneous choice set

Next we construct a variable to check how well the three specifications predict aggregate consumer behavior. Using the parameter estimates, we first predict consumer brand choice on each shopping occasion, which is the brand that generates the highest utility for the consumer on that shopping trip. Assuming that the consumer would purchase the same quantity of cereal as in the data, we can then calculate the consumer expenditure on that shopping trip. Summing up the consumer expenditures for each brand in the sample period, we get the predicted brand sales and brand market shares. Then we 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 5.7, the third specification generates a smaller market share prediction error than the first two.

In summary, introducing limited information about brand existence into the model improves the data fit and better captures consumer behavior. Therefore, we 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. The pricing decision for a brand depends on the price elasticity of demand. Advertising provides product information and affects 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 5.5, above, we calculate the own and cross price elasticities for the top 25 brands,[27] which are reported in Table 5.8. The formula for computing the price elasticities is in Appendix 5.4. The price elasticities are evaluated at the median of each brand's price and the sample market shares.

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

Table 5.8 Estimated price elasticities for top 25 brands based on IV estimation

Brand	1	2	3	4	5	6	7	8	9	10	11	12
1	−2.428	0.367	0.146	0.01	0.099	0.338	0.22	0.315	0.003	0.673	0.034	0.002
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
3	0.12	0.594	−1.545	0.015	0.157	0.54	0.352	0.503	0.005	1.075	0.054	0.002
4	0.092	0.455	0.179	−3.703	0.12	0.414	0.269	0.385	0.004	0.823	0.042	0.002
5	0.07	0.349	0.137	0.009	−2.762	0.317	0.206	0.295	0.003	0.63	0.032	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
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
8	0.249	1.234	0.485	0.032	0.326	1.121	0.73	−2.209	0.01	1.331	0.113	0.005
9	0.047	0.232	0.091	0.006	0.061	0.211	0.137	0.196	−1.679	0.42	0.021	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
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
12	0.022	0.109	0.043	0.003	0.029	0.099	0.064	0.092	0.001	0.197	0.01	−2.798
13	0.153	0.756	0.298	0.02	0.2	0.688	0.448	0.64	0.006	1.368	0.069	0.003
14	0.022	0.108	0.043	0.003	0.029	0.099	0.064	0.092	0.001	0.196	0.01	0
15	0.033	0.165	0.065	0.004	0.044	0.15	0.098	0.14	0.001	0.299	0.015	0.001
16	0.044	0.217	0.085	0.006	0.057	0.197	0.128	0.183	0.002	0.392	0.02	0.001
17	0.116	0.573	0.225	0.015	0.152	0.52	0.339	0.484	0.005	1.035	0.052	0.002
18	0.034	0.168	0.066	0.004	0.045	0.153	0.1	0.142	0.001	0.305	0.015	0.001
19	0.078	0.385	0.152	0.01	0.102	0.35	0.228	0.326	0.003	0.697	0.035	0.002
20	0.168	0.831	0.327	0.022	0.22	0.755	0.492	0.703	0.007	1.503	0.076	0.003
21	0.027	0.136	0.053	0.004	0.036	0.123	0.08	0.115	0.001	0.245	0.012	0.001
22	0.014	0.07	0.028	0.002	0.019	0.064	0.042	0.059	0.001	0.127	0.006	0
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
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
25	0.031	0.155	0.061	0.004	0.041	0.141	0.092	0.131	0.001	0.28	0.014	0.001

13	14	15	16	17	18	19	20	21	22	23	24	25
0.1	0.001	0.001	0.065	0.261	0.027	0.115	0.469	0.004	0.004	0.011	0.003	0.001
0.18	0.003	0.001	0.077	0.306	0.032	0.135	0.551	0.005	0.004	0.013	0.003	0.003
0.159	0.002	0.001	0.014	0.058	0.006	0.025	0.104	0.001	0.001	0.002	0.001	0.004
0.122	0.002	0.001	0.043	0.173	0.018	0.076	0.312	0.003	0.002	0.008	0.002	0.004
0.093	0.001	0.001	0.014	0.055	0.006	0.024	0.099	0.001	0.001	0.002	0.001	0.001
0.387	0.005	0.003	0.007	0.027	0.003	0.012	0.049	0	0	0.001	0	0.002
0.281	0.004	0.002	0.047	0.188	0.02	0.083	0.338	0.003	0.003	0.008	0.002	0.006
0.33	0.005	0.003	0.007	0.027	0.003	0.012	0.048	0	0	0.001	0	0.008
0.062	0.001	0.001	0.01	0.041	0.004	0.018	0.074	0.001	0.001	0.002	0	0.001
0.187	0.003	0.002	0.013	0.054	0.006	0.024	0.097	0.001	0.001	0.002	0.001	0.004
0.059	0.001	0	0.036	0.142	0.015	0.063	0.256	0.002	0.002	0.006	0.002	0.001
0.029	0	0	0.01	0.042	0.004	0.018	0.075	0.001	0.001	0.002	0	0.001
−3.629	0.003	0.002	0.024	0.096	0.01	0.042	0.172	0.002	0.001	0.004	0.001	0.005
0.029	−1.094	0	0.052	0.206	0.022	0.091	0.371	0.003	0.003	0.009	0.002	0.001
0.044	0.001	−1.351	0.008	0.034	0.004	0.015	0.061	0.001	0	0.001	0	0.001
0.058	0.001	0	−2.409	0.017	0.002	0.008	0.031	0	0	0.001	0	0.001
0.153	0.002	0.001	0.036	−3.788	0.015	0.063	0.256	0.002	0.002	0.006	0.002	0.003
0.045	0.001	0	0.008	0.031	−1.305	0.014	0.056	0	0	0.001	0	0.001
0.103	0.001	0.001	0.01	0.038	0.004	−2.384	0.069	0.001	0.001	0.002	0	0.002
0.222	0.003	0.002	0.092	0.01	0.041	0.166	−3.958	0.001	0.004	0.001	0.002	0.005
0.036	0.001	0	0.167	0.017	0.074	0.3	0.003	−1.434	0.007	0.002	0.004	0.001
0.019	0	0	0.148	0.015	0.065	0.265	0.002	0.002	−0.944	0.002	0.004	0
0.153	0.002	0.001	0.113	0.012	0.05	0.203	0.002	0.002	0.005	−3.178	0.003	0.003
0.034	0	0	0.087	0.009	0.038	0.156	0.001	0.001	0.004	0.001	−1.171	0.001
0.041	0.001	0	0.359	0.037	0.158	0.645	0.006	0.005	0.016	0.004	0.009	−1.409

5.5.2 Estimation with Instrumental Variables

Using both sets of instruments (nutrition of competing brands and cost factors), we report the estimates of $\bar{\beta}$ in column IV of Table 5.5, above. To test the endogeneity of price and advertising, we run an ordinary least squares (OLS) regression of Eq. (5.21) after we obtain δ_{jt} in step (3) 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 estimates are not significantly different from the estimates with IV. 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 whereas in the estimations without IV, 37,858 transactions are used—we will conduct policy experiments using the estimates without IV.

5.6 Counterfactual Experiments

We conduct three counterfactual experiments to evaluate some of the brand marketing strategies and a hypothetical food policy change. In the first two experiments, we choose Brand 28 as an example because it was newly introduced into the market in January 2003. Figure 5.5 summarizes Brand 28's average monthly prices, sales, and advertising in the estimation sample. Marketing managers are usually

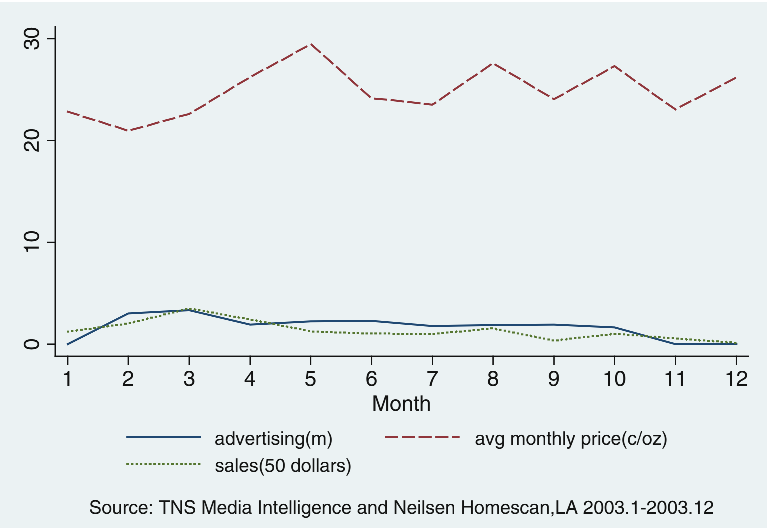

Fig. 5.5 Average monthly advertising, price and sales for Brand 28

Table 5.9 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 observed in data

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, we explore the effect of a hypothetical policy change—banning cereal advertising targeted to children—on consumer choices. A caveat should be borne in mind when we interpret the results of the experiments: the strategies of other firms are kept unchanged when we simulate the results, and thus the optimal responses of rival firms are not taken into account.[28]

5.6.1 Pricing Strategy for Brand 28

We first vary Brand 28's price from its observed price by +1, +5, −1, and −5 %, separately. Each time under the new pricing scheme, we calculate every household's simulated choices and aggregate them to get brand market shares and sales. The resulting changes in market share and sales of Brand 28 are reported in Table 5.9. We can see that if the price is reduced by 5%, the sales improve by 2.3%, compared with 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, we calculate the changes in expenditures for different demographic groups after the price drops by 5%. We 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 children in the household. The results by demographic groups are shown in Table 5.10. Consumers with children and lower income respond more to the price cut than their counterparts, but the response does not vary with age groups.

Next we look at the average (weighted by volume) daily transaction prices of Brand 28 at its introductory stage (the first 3 months of 2003) and see whether 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 January to March

[28] 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.

Table 5.10 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

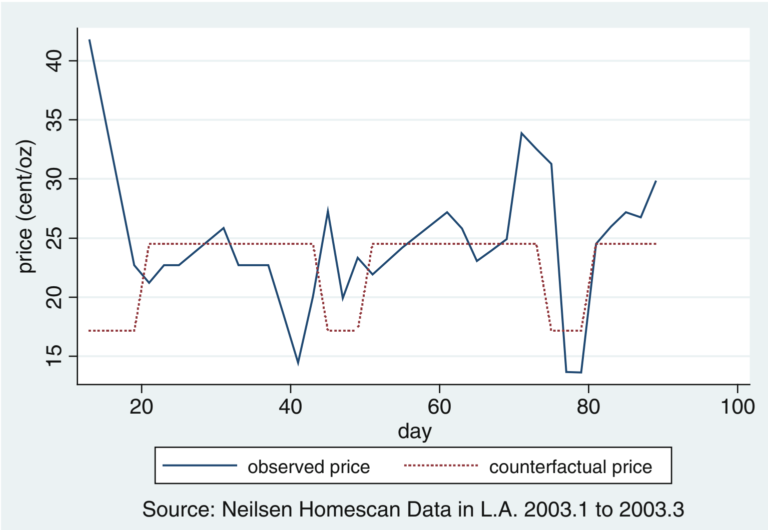

Fig. 5.6 Average daily transaction price for Brand 28

2003 is shown in Fig. 5.6. The initial price was very high, followed by a period of medium price level. Deep discounts happened twice when the price was about 60% of the average level. We 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 Fig. 5.4, above. With the new pricing strategy, we 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 a lower introductory price. Consumers who are lured into purchase by the low introductory prices will then form brand loyalty, and thus the brand manager can profit by setting the 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 attract 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.[29] 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 Sect. 5.5, above, if demand is estimated while 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.

5.6.2 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 continual 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 Fig. 5.5, above, that Brand 28's advertising was relatively smooth over time. In contrast, many advertisers of consumer packaged goods use pulsing strategies. For example, Dubé et al. (2005) find that pulsing is the optimal advertising strategy in the frozen entrée market. Naik et al. (1998) develop a model of dynamic advertising that shows that pulsing strategies can generate greater total awareness than the continual advertising when the effectiveness of advertising varies over time. Specifically, ad effectiveness declines during periods of continual advertising and is restored 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 campaign. They also show that the best advertising strategy for a major cereal brand is pulsing.

To mimic the pulsing strategy, we reschedule Brand 28's advertising by equally dividing the 2003 total ad expenditure into the 6 odd months and setting the budget to zero in the 6 even months (Fig. 5.7 plots the observed advertising versus the counterfactual pulsing advertising). Then we 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's entering the consumer choice set in the first 2 months after its introduction. In the observed data, the advertising expenditure for Brand 28 in January is zero, but in the pulsing strategy it is \$3.3 million. The increase in the advertising expenditure in January raises the probability that an average consumer (with mean income, mean age, and mean presence of children)

[29] See "Eating Well," *New York Times,* September 22, 1993.

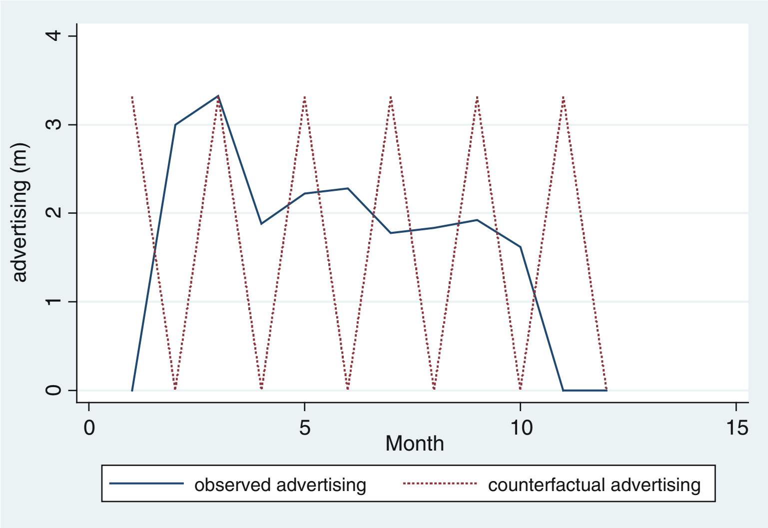

Fig. 5.7 Monthly advertising for Brand 28

Table 5.11 Change in expenditure by demographic group under pulsing strategy

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

will be aware of Brand 28 from almost zero to 89.7%. In February the pulsing strategy increases the advertising expenditure of Brand 28 from $2.99 million to $3.14 million, and it raises the probability of an average consumer's 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 from 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 its overall market share. We also examine how different consumer groups respond to the pulsing strategy. The results, in Table 5.11, suggest that consumers with higher income and with children are more sensitive to the change in advertising strategy, but age does not matter.

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

Table 5.12 Change in segment share after ban on advertising for kid brands

	Δ in mktshare (%)
Kid	−5.98
Adult	2.01
Family	3.96

Table 5.13 Effects of 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

5.6.3 Effects of Banning Child-Oriented Cereal Advertising

We do not directly observe the value of ad dollars used for marketing toward children. To measure the effects of an advertising ban, we approximate the ban of child-oriented cereal advertising by eliminating the advertising expenditures for kids' cereal brands while holding other factors unchanged. In the experiment, we replace the ad stock of these brands with zero and calculate how the brand market shares change. The total changes for each brand segment (family, adult, kid) are summarized in Table 5.12. After the hypothetical policy change, the total market share of kids' brands goes down by about 6%, of which 2% goes to the adult brands and 4% goes to the family brands.

Then we look at how the policy change affects the nutritional intake and the expenditures of different consumer groups. The results are summarized in Table 5.13. Overall, after the ban of child-oriented cereal advertising, consumers consume more fiber and less sugar, which is better for their health. Consumers who are younger, with lower income, and with children 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. However, after the ban, consumers of all demographic groups have to increase their expenditures because they consume more adult and family cereals, which are more expensive than kids' cereals.

5.7 Conclusion

Using ready-to-eat cereal as an example of experience goods, we consider limited information on both product existence and product quality in a dynamic model. 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.

We have two main findings pertaining to the value of information. First, failure to account for limited information about a product's existence may significantly underestimate price elasticity. In our data, consumers are indeed sensitive to the price of the brands they know, but by assumption they cannot respond to price cuts in the brands they do not know. This finding implies that informative advertising that expands consumer choice set promotes competition because it allows price-sensitive consumers to choose among more brands. Second, advertising is much more effective on new consumers than on old consumers, which is consistent with the argument that advertising is mainly informative and not persuasive (at least in the RTE cereal market). The strong habit formation found in our data emphasizes the importance of the first-time experience and the information generated from it.

Both findings have useful implications for public policy. Since manufacturers' advertising is driven by private gains, informative advertising may be under-provided if part of the value of informative advertising is public (e.g., the value of condoms in reducing public health risk), if new entrants cannot afford informative advertising, or if manufacturers anticipate the procompetitive effect of informative advertising and collude to keep consumers uninformed of all choices. In these cases, public policies may play an active role in presenting available choices to consumers and encouraging competition among firms. By helping consumers make a smarter choice of first-time experience, these public policies can have a long-lasting effect on consumer welfare, thanks to habit formation.

On the other hand, manufacturers' advertising can be overprovided if advertising signals high quality in a dimension that is easy to tell by experience (say, the taste of the cereal) but remains silent on dimensions that are hard to know (say, the health consequences of eating sugary cereals). Since we do not model this complication, the ban of advertising on sugary cereals appears welfare reducing *from the consumer's point of view* because it leads to a smaller choice set and a less informative choice within the choice set. However, if advertising misleads consumers to choose sugary cereals—either because consumers are unaware of the unhealthfulness of the advertised food or because they like the sugary taste and do not consider their future health—limiting consumer's choice set could be beneficial to consumers.

There are other reasons why the counterfactual predictions on the ban of advertising should be taken with caution. In all the counterfactual experiments, we do not consider the competitive responses of other firms to the change in brand strategies. Nor do we account for the fact that firms may change the way they promote kids' brands once the government regulation comes into play. To control for these responses, we would need to solve the firm's profit maximization problem. In a model with brand loyalty on the consumer side, the firm's problem should involve dynamic optimization: the firm considers not only 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 chapter focuses only on the role of manufacturers. A vertical competition model is needed to analyze the role of retailers.

5. Appendixes

5.1. Controlling for Unobserved Consumer Heterogeneity

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

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

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

Let $v_{ij} = \mu_{ij} + \omega_j$. $\mu_{ij} : N(0, \varsigma_{ij}^2)$, and $\omega_j = E(v_{ij})$ is a constant. Assuming ε_{ijt} has a generalized extreme value distribution, then we can write the probability that consumer i will choose j conditional on $\mu_{i1},\ \mu_{i2}, \ldots \mu_{i51}$, and choice set C_{it} as

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

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

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

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

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

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

To reduce the number of parameters to be estimated, we allow ω_j to vary across brand segment, and ς_{ij}^2 to vary across both brand segment and whether the household has children. There are a total of eight parameters to estimate for unobserved consumer-brand preferences, of 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 the family, adult, or kid segment, and the second subscript denotes whether there are any children in the household; two are location parameters: ω_A and ω_K, where the subscript denotes whether the brand belongs to the adult or kid segment. ω_F is normalized to zero.

5.2. Choice Set Simulation Details

In the simulation, we assume that choice set is a function of brand advertising and purchase experience, as shown in Eqs. (5.1) and (5.2). The specific choice set simulation process is outlined as follows.

Step 1. Calculate $q_{ijt}(\varphi)$ 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 $u_{ijt}^r < q_{ijt}$, 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_{it}^r. After simulating the choice set, we 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_{it}^r. (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 $p_{ijt}^r = \prod_{j \in C_{it}^r} q_{ijt} \prod_{k \notin C_{it}^r} (1 - q_{ikt}) \times P^r(j|C_{it})$, consumer i's unconditional probability of choosing brand j at time t in the rth simulation.
Step 6. Draw the random numbers u_{ijt}^r repeatedly for R times, and each time repeat steps 2–5.
Step 7. Calculate the simulated choice probability $\hat{p}_{ijt} = \frac{1}{R} \sum_{r=1}^{R} p_{ijt}^r$..

5.3. Contraction Mapping Details

In the instrumental variable estimation, we need to find the δ that makes 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 calculated 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 his choice set on the shopping occasion and chooses the one that yields the highest utility.

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

To obtain the values of δ that solve $\sigma_j(\delta^h, \Pi, \Sigma, \kappa, \lambda, \gamma) = S_j$, 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 (2008).

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 ∀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 $\sigma_j = \int \sum_{C_i \in \Omega_j} \prod_{l \in C_i} q_{ilt} \prod_{k \notin C_i} (1 - q_{ikt}) P(j|C_i) f(v) dv$,where, and Ω_j denotes the set of choice sets that include j.

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

where

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

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

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

Since all elements in the integral are nonnegative, 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.

5.4. *Price Elasticity Calculation*

Suppressing the time subscript, we can write the consumer utility function as

$$U_{ij} = \alpha_i p_j + \Upsilon_j g \beta_{\Upsilon i} + \varepsilon_{ij}$$

where $\alpha_i = \alpha + \Pi_3 g D_i + \Sigma_{33} \cdot v_3$, 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

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

where p_{ij} represents the probability that consumer i will choose brand j.

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

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

5. Commentary: Explaining Market Dynamics: Information Versus Prestige

Mead Over

Information is valuable to cereal manufacturers, who pay for advertising. Information is valuable to consumers, who reveal by their expenditure response that they attend to advertising. Information is valuable to nutrition activists, as a policy instrument to manipulate in the paternalistic hope that consumers deprived of advertising for sugary cereals will feed their children less sugar. And finally, information is valuable to the authors of the chapter, because using more of it enables them to explain more of the variation in market shares across the cereal brands and to predict more plausibly the reaction of consumers to price or advertising interventions for an individual brand or by a government consumer protection agency

Advertising is one of the industries whose business model involves the packaging and delivery of information. In contrast to the commercial publishing industry, wherein the author and originator of the information profits when the consumer values the information enough to buy the book, profits of the advertising industry derive from the advertiser's willingness to pay to subsidize information provision to the consumer. The distinction is due to the fact that consumers of books value them for their own sake, whereas consumers of information about advertised products use that information to inform their expenditures on those products. In an imperfectly competitive market for ready-to-eat cereals, cereal manufacturers are willing to subsidize consumers' information acquisition in order to differentiate brands from one another and reduce consumers' price elasticity of demand for their own brands.

The chapter deploys a variety of interesting microeconomic modeling and computationally intense econometric techniques to exploit a large data set on consumer purchases of ready-to-eat cereals and estimate the potential effect of a specific type of government intervention in this market: a ban on the advertising of children's cereal. The authors conclude that such a ban would indeed be effective in reallocating consumer expenditure away from the least healthful types of cereals and toward more healthful, more expensive brands, but it would induce consumers to spend more on cereal than they would without the ban. But one wonders whether the extraordinarily complex econometric paraphernalia the authors would really be required to show these impacts of advertising.

Since the authors generously provide the market shares of the top 50 brands as well as their average prices, brand-specific monthly advertising expenses, and market

M. Over
Center for Global Development, Washington, DC, USA
e-mail: mover@cgdev.org

Table 5.C.1 Ordinary least squares regression of logit of average market share on log price and advertising expenditures, by market segment

Number of obs = 50 F(8, 41) = 10.73 Prob > F = 0.0000 R-squared = 0.6768 Adj R-squared = 0.6137 Root MSE = .55338

Logit of (marketshare)	Coef.			P > |t|	[95%]	
Log(price):						
Adult	−.88	.69	−1.27	0.212	−2.28	.52
Family	−2.17	.54	−4.05	0.000	−3.26	−1.09
Kid	−.07	1.16	−0.06	0.951	−2.42	2.28
Advertising:						
Adult	.00086	.0002	4.44	0.000	.00047	.0013
Family	.00092	.0002	5.14	0.000	.00056	.0013
Kid	.00105	.0004	2.50	0.016	.00020	.0019
Constants:						
Adult	−3.78	2.56	−1.47	0.148	−8.96	1.40
Family	1.09	1.55	0.70	0.485	−2.04	4.22
Kid	−5.94	3.71	−1.60	0.117	−13.43	1.55

Source: This reviewer's estimates using the grouped data from Table 5.2

segment (in Table 5.2), one can calculate a descriptive ordinary least squares regression of (the logit of) market share on this grouped data. The results of this "naïve" regression are presented here in Table 5.C.1.

Although requiring very little effort beyond the tabulation of the average market shares, prices, and advertising expenditures for the 50 top brands, these results seem somewhat informative. The point estimates of the three estimated price coefficients, one for each of the three market segments, are all negative, as expected, with the one for family cereals being large (>2 in absolute magnitude) and statistically significant. Furthermore, all three advertising coefficients are highly statistically significant, suggesting that an extra million dollars of advertising increases market share by 0.86% for adult cereals, 0.92% for family cereals, and 1.05% for kid cereals. The category of kid cereals seems to respond more to advertising expenditures than the other two.

So why do more? What have the authors' prodigious efforts added to our knowledge of the ready-to-eat cereal market?

This chapter supports the proposition that "information is valuable to economic researchers" in three ways. First, by exploiting detailed information on the thousands of individual consumer transactions summarized in Table 5.2, the authors are able to relax several of the assumptions that are maintained by the above naïve analysis. In so doing, they demonstrate the value of that detailed information to the understanding of this complex market. Second, by bringing to bear an economic theory of decision making, the authors demonstrate that this theory itself has information content—because it helps explain the market data. Third, by combining the unusually detailed and granular data with this powerful theory, the authors are able to distinguish the two channels by which advertising hypothetically affects consumer behavior, the "information" channel and the "prestige" channel, and to demonstrate that it's the information that influences the consumer's behavior—not

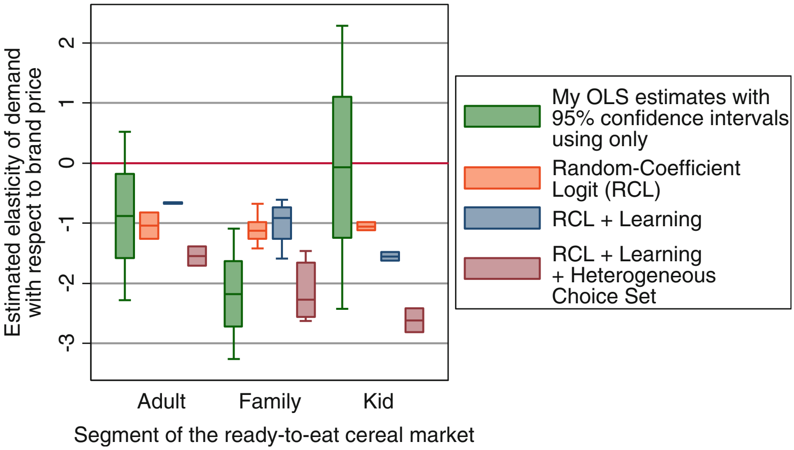

Fig. 5.C.1 Adding information either with more granular data or more theory-constrained economic structure increases both the precision and heterogeneity of estimated price elasticities across brands

the prestige. Fourth, by using information from the supply side of the cereal market, the authors are able to reject some types of endogeneity that would cast doubt not only on my naïve model, but also on their three principal models.

Consider the estimated price elasticities. Figure 5.C.1 displays for each of the three market segments the confidence intervals for my naively estimated price elasticities from Table 5.C.1 and the range of estimated elasticities for the top 10 cereal brands presented by the authors in their Table 5.8. There are two adult cereal brands in the top 10, six family brands and two kid brands. Note the extremely wide confidence intervals from my naïve estimates. Next to those confidence intervals (in green), my Fig. 5.C.1 displays the range of estimated price elasticities for each of the authors' three estimated models. Although the authors do not report confidence intervals, the point estimates of the brand-specific coefficient estimates from which these elasticities are derived (the first row of Table 5.5) are from 3 to 50 times larger than their estimated standard errors, suggesting tight confidence intervals for the elasticities. And the range of these reported estimates is also relatively tight within each market segment. Thus, one benefit of the information in the granular data appears to be tighter estimates of the brand-specific price elasticities.

The authors' basic model is a random-coefficients logit model (RCL) structured to assume that the choice sets for all consumers include all 50 brands (plus a 51st composite of all other brands) and characteristics of all brands are known. Figure 5.C.1 shows that the estimated elasticities for this model are roughly the same across the three market segments. (See the orange boxes in Fig. 5.C.1.) The authors' second model, whose elasticity estimates are represented by the blue boxes labeled "RCL + Learning," relaxes the assumption that all consumers know the characteristics of all brands. In this model the consumers again choose among all brands but only know the qualities of brands previously purchased. Advertising directly influences a brands market share. Thus, the impact of the economic theory on the estimated elasticities is to differentiate the three theoretically distinct markets, information that is useful to students of this ready-to-eat cereal market. Finally when the authors use an elaborate simulation model to require advertising to inform consumers of an unused brand's existence before it can affect their purchases (the assumption of heterogeneous choice sets), the estimated elasticities diverge even more across the three market segments (pink boxes) and also increase substantially in absolute magnitude. In the words of the authors, "[t]he estimated price elasticities in the ... specification [allowing a heterogeneous choice set] 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." Once more, economic theory has improved the fit of the model and contributed insight on the cereal market.

Variation in observed market shares, the naïve model contains substantial information. Its prediction error (defined by the authors as the square root of the sum of squared differences between the actual market share of Table 5.6 and the predicted share) equals 6.5, which is actually less than the 7.26 scored by the authors' random-coefficients model (bottom row of Table 5.6). However, both of the authors' more sophisticated models do better than my naïve model, scoring 5.28 and 3.81 respectively, and thus can be said to contain more valuable information.

Because they are able to simulate the consumers' choice sets each time on each visit to the grocery store, the authors can distinguish the two possible channels by which advertising might induce people to spend more on cereal—the information channel and the prestige channel. It's interesting that for this market, the authors find no support for the hypothesis that advertising persuades consumers to increase their consumption of ready-to-eat cereals that are familiar to them—which would be a prestige effect of advertising. Instead, advertising's role seems to be to induce consumers to try cereals that are unfamiliar. When they model this effect, the authors estimate much larger price elasticities (the pink boxes in Fig. 5.C.1). Since consumers have many choices in the cereal market, evidence that price elasticities are large in the children's cereal market and small in the adult cereal market suggests that the adults who purchase cereal for children see them as highly substitutable for one another, whereas they are loath to substitute one adult cereal for another. Adult cereal brands thus have more market power than children's brands.[30]

The authors' simulations of a ban on advertising for children's cereal and of a "pulsed" advertising strategy both raise the issue of the potential value to the public of government use of advertising. Using their third model, which incorporates

[30] The authors' finding of the highest price elasticity for children's cereal contrasts with the naïve model's failure to find any price effect on the market shares of children's cereals. A simple experimental manipulation of the price of a children's cereal would thus quickly demonstrate which of these two models is a more realistic portrayal of this market.

consumer learning and heterogeneous choice sets, and assuming that affected cereal manufacturers hold constant the prices of their brands, the authors simulate a ban on advertising and conclude that "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." It's possible to perform this same experiment with the naïve model, by first computing the fitted market shares from the OLS regression in the children's market and then computing them a second time after the value of advertising has been set to zero. The result from the naïve model is that the total market share of children's brands would decline from 17.7 to 9.5% of the market, a reduction of about 8.2%. Under the assumption of the independence of irrelevant alternatives (the well-known IIA assumption typically maintained in multinomial logit models), about 2.2 percentage points of this decline would be reflected by an increase in the adult segment and about 5.8% age points would go to the family segment. Despite the simplicity of the naïve model, these results are remarkably similar to those obtained by the authors.

In contrast to the ban on advertising of children's cereals, the possible effects on the market of pulsed advertising could not be analyzed with the naïve model. The authors have used their heterogeneous choice set model to show that spreading the same advertising dollars smoothly is less effective at increasing market share than would be a strategy of bunching the advertising in specific months. The superior effectiveness of pulsing seems to be due to the lack of a prestige effect of advertising in this market. The implication is that government public awareness campaigns that intend to improve people's awareness of alternatives—and subsequently depend on their good experience with these alternatives to motivate behavior—could also benefit from pulse advertising. Whether the reverse is true for public awareness campaigns that intend to enhance the prestige of certain behavior remains to be determined.

The authors allude in passing to monopolistic pricing strategies when they point out that a monopolist operates in the elastic portion of its demand curve. Under certain conditions one could go further and assert that a profit-maximizing firm in a monopolistic or monopolistically competitive market will set its price-cost margin equal to the inverse of the elasticity of demand. According to the authors' heterogeneous choice set model, the median elasticities in the adult, family, and children's market segments are about -1.5, -2.3, and -2.8, respectively. This suggests that typical markups of price over marginal cost in these three segments are 65, 49, and 38 % of marginal costs, respectively. Furthermore, markups on individual brands vary from 34 to 72% of marginal costs. This information is of only academic interest in the market for ready-to-eat cereals, imagine if a similar analysis of the pharmaceutical market revealed such information about the prices of pharmaceutical brands. Views on pharmaceutical pricing range from the idea that monopoly profits in the pharmaceutical market are unproductive "rent" gained from branding products that largely result from government-subsidized research to the position that these profits are a just return on pharmaceutical firms' own research investments and motivate their future research. An objective observer would grant that both views have some legitimacy in various parts of the market. But policy intervention on the prices of individual drugs is hampered by the secrecy with which pharmaceutical firms guard their cost information. To the extent that the techniques employed in this chapter could be used to reveal the apparent markups of pharmaceutical prices over costs, regulators would value this information as an input to the regulation of the monopoly prices of individual pharmaceutical products.

References

Ackerberg, D. (2001). Empirically distinguishing informative and prestige effects of advertising. *The RAND Journal of Economics, 32*(2), 316–333.

Ackerberg, D. (2003). Advertising, learning, and consumer choice in experience good markets: A structural empirical examination. *International Economic Review, 44*(3), 1007–1040.

Allenby, G., & Ginter, J. (1995). The effects of in-store display and feature advertising on consideration sets. *International Journal of Research in Marketing, 12*, 67–80.

Anand, BN., & Shachar, R. (2011). *Advertising, the Matchmaker*. *The RAND Journal of Economics*, 42(2), 205–245.

Andrews, R., & Srinivasan, T. (1995, Feburary). Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research, 32*, 30–41.

Becker, G., & Murphy, K. (1993). A simple theory of advertising as a good or bad. *Quarterly Journal of Economics, 108*, 942–964.

Berry, S. (1994). Estimating discrete choice models of product differentiation. *The RAND Journal of Economics, 25*(2), 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica, 63*(4), 841–890.

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy, 112* (1), 68–105.

Berry, S., Linton, O., & Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *Review of Economics Studies, 71*(3), 613 –654.

Betancourt, R., & Clague, C. (1981). *Capital utilization: A theoretical and empirical analysis*. Cambridge: Cambridge University Press.

Butters, G. (1977). Equilibrium distribution of prices and advertising. *The Review of Economic Studies, 44*, 465–492.

Chen, Y., & Kuan, C. (2002). The pseudo-true score encompassing test for non-nested hypotheses. *Journal of Econometrics, 106*, 271–295.

Chintagunta, P., Jiang, R., & Jin, G. (2009). Information, learning, and drug diffusion: The case of Cox-2 inhibitors. *Quantitative Marketing and Economics, 7*(4), 399–443.

Costantino, C. (2004). *Three essays on vertical product differentiation: Exclusivity, non-exclusivity and advertising*. Ph.D. dissertation, University of Maryland, College Park.

Crawford, G., & Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica, 73*, 1137–1174.

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

Dubé, J., Hitsch, G., & Manchanda, P. (2005). An empirical model of advertising dynamics. *Quantitative Marketing and Economics, 3*(2), 107–144.

Einav, L., Leibtag, E., & Nevo, A. (2010, June). Recording discrepancies in Nielsen Homescan data: Are they present and do they matter? *Quantitative Marketing and Economics, 8*(2), 207–239.

Eliaz, K., & Spiegler, R. (2011). Consideration sets and competitive marketing, *Review of Economic Studies, 78*(1), 235–262.

Erdem, T., & Keane, M. (1996). Decision-making under uncertainty: Capturing dynamic brand choices in turbulent consumer goods markets. *Marketing Science, 15*, 1–20.

Geweke, J., Keane, M., & Runkle, D. (1994). Alternative computational approaches to inferences in the multinomial probit model. *The Review of Economics and Statistics, 76*(4), 609–632.

Goeree, M. S. (2008). Limited information and advertising in the US personal computer industry. *Econometrica, 76*(5), 1017–1074.

Grossman, G., & Shapiro, C. (1984). Informative advertising with differentiated products. *The Review of Economic Studies, 51*, 63–81.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods* (pp. 209–237). Chicago: University of Chicago Press.

Hendel, I. (1999). Estimating multiple-discrete choice models. *The Review of Economic Studies, 66*(2), 423–446.

Hendel, I., & Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica, 74*(6), 1637–1673.

Hitsch, G. (2006). An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science, 25*(1), 25–50.

Keane, M. (1994). A computationally practical simulation estimator for panel data. *Econometrica, 61*(1), 95–116.

Kihlstrom, R., & Riordan, M. (1984). Advertising as a signal. *Journal of Political Economy, 92*, 427–450.

Mehta, N., Rajiv, S., & Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Science, 22*(1), 58–84.

Milgrom, P., & Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy, 94*, 796–821.

Naik, P., Mantrala, M., & Sawyer, A. (1998). Planning media schedules in the presence of dynamic advertising quality. *Marketing Science, 17*(3), 214–235.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy, 78*(2), 311–329.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy, 82*, 729–753.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica, 69* (2), 307–342.

Nevo, A., & Hendel, I. (2006). Sales and consumer inventory. *The RAND Journal of Economics, 37*(3), 543–561.

Osborne, M. (2006). Consumer learning and habit formation in packaged goods markets, Ph.D. thesis. Stanford University.

Roberts, J., & Lattin, J. (1997). Consideration: Review of research and prospects for future insights. *Journal of Marketing Research, 34*(3), 406–410.

Shum, M. (2004). Does advertising overcome brand loyalty? Evidence from the breakfast-cereals market. *Journal of Economics and Management Strategy, 13*(2), 241–272.

Stigler, G., & Becker, G. (1977). De Gustibus Non Est Disputandum. *The American Economic Review, 67*(2), 76–90.

Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B: Methodological, 35*, 643–666.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*, 307–333.