Econometrica, Vol. 74, No. 6 (November, 2006), 1637–1673

MEASURING THE IMPLICATIONS OF SALES AND CONSUMER INVENTORY BEHAVIOR

BY IGAL HENDEL AND AVIV NEVO[1]

Temporary price reductions (sales) are common for many goods and naturally result in large increases in the quantity sold. Demand estimation based on temporary price reductions may mismeasure the long-run responsiveness to prices. In this paper we quantify the extent of the problem and assess its economic implications. We structurally estimate a dynamic model of consumer choice using two years of scanner data on the purchasing behavior of a panel of households. The results suggest that static demand estimates, which neglect dynamics, (i) overestimate own-price elasticities by 30 percent, (ii) underestimate cross-price elasticities by up to a factor of 5, and (iii) overestimate the substitution to the no-purchase or outside option by over 200 percent. This suggests that policy analysis based on static elasticity estimates will underestimate price–cost margins and underpredict the effects of mergers.

KEYWORDS: Long-run price elasticities, stockpiling, demand anticipation, discrete choice models, differentiated products, storable goods.

1. INTRODUCTION

WHEN GOODS ARE STORABLE, traditional static demand analysis is likely to mismeasure long-run own- and cross- price elasticities. A temporary price decrease may generate a large demand increase. However, if part of the increase is due to intertemporal substitution, then the (long-run) own-price response to a permanent price reduction would be smaller than static estimates suggest. On the other hand, the direction of the bias on cross-price elasticities is ambiguous. Static demand models are misspecified because they do not control for relevant history like past prices and inventories, and therefore estimates of price sensitivity might be biased. Even adding the right controls, static estimation confounds long- and short-run price effects. Measuring the long term response is relevant for most applications. For example, demand estimates are central in antitrust analysis and for computing welfare gains from new goods.

The distinction between long-run and short-run responses dates back to the Lucas and Rapping (1969) labor market study. Labor supply in their model responds to short-run wage changes differently than to long-run changes. In the context of product market demand estimation, most applications have neglected intertemporal effects (Bresnahan (1987), Hausman, Leonard, and Zona (1994), and Berry, Levinsohn, and Pakes (1995), as well as many others). Dynamic effects could be important, for example, for either durable or storable products. Previous work in the marketing and economics literature, which

[1]We wish to thank David Bell for the data, Michael Keane, Ariel Pakes, John Rust, a co-editor, Glenn Ellison, four anonymous referees, and seminar participants in several workshops for comments and suggestions. We gratefully acknowledge support from the NSF (Grants SES-0093967 and SES-0213976). The second author also thanks the Sloan Foundation for support through a Sloan Research Fellowship.

we survey below, documented purchasing patterns consistent with consumer stockpiling behavior. In an interesting paper, Erdem, Imai, and Keane (2003) study dynamic demand of a storable product (ketchup). The consumer inventory problem we study is very close to their setup. We propose a demand model that is computationally simpler and flexibly incorporates observable heterogeneity. The model enables us to split the estimation into a brand choice and a quantity choice. Most of the parameters are estimated consistently without solving the consumer's dynamic problem, and the remaining dynamic parameters are then estimated in a computationally simpler second stage.

We present a model in which households maximize the present expected value of future utility flows, facing uncertain future prices. In each period a household decides how much to buy, which brand to buy, and how much to consume. Quantities not consumed are stored, at a cost, and can be consumed in the future. The consumer balances the cost of holding inventory and potential future savings. The model is estimated using weekly scanner data on laundry detergents. These data were collected using scanning devices in nine supermarkets, belonging to different chains, in two submarkets of a large midwest city. In addition, we follow the purchases of households who shopped in those stores over a period of two years. The data have information on which product was bought, where it was bought, and how much was paid. It is also known when the households visited a supermarket but decided not to purchase a laundry detergent.

The structural estimation follows the nested algorithm proposed by Rust (1987). We have to make two adjustments. First, inventory, one of the endogenous state variables, is not observed by us. To address this problem we generate an initial distribution of inventory, and update it period by period using observed purchases and the (optimal) consumption prescribed by the model. Second, the state space includes prices (and promotional and advertising variables) of all brands in all sizes and therefore is too large for practical estimation. To reduce the dimensionality, we separate the probability of choosing any brand–size combination into the probability of choosing a brand conditional on the quantity purchased, and the probability of choosing each specific quantity. Given the stochastic structure of the model, we show that the probability of choosing a brand conditional on quantity does not depend on dynamic considerations. Therefore, we can estimate many of the parameters of the model without solving the dynamic programming problem. We estimate the remaining parameters by solving a nested algorithm in a much smaller space, considering only the quantity decision. This procedure enables us to estimate a general model, which allows for a large degree of consumer heterogeneity and nests standard static choice models. We discuss below the assumptions necessary to validate this procedure and the limitations of the method.

The estimates suggest that ignoring dynamics can have implications on demand estimates. Comparing estimates of the demand elasticities computed from a static model and the dynamic model, we find the following. First, the static model overestimates own-price elasticities by roughly 30 percent. Second, the static model underestimates cross-price elasticities to other products. The ratio of the static cross-price elasticities to those computed from the dynamic model is as low as 0.2. Third, the estimates from the static model overestimate the substitution to the no-purchase or outside option by 200 percent.

Estimates of demand elasticities are typically used in a first-order condition, for example, of a pricing game, to compute price–cost margins (PCM). For single product firms, the magnitude of the bias of the implied margin is proportional to the bias in the own-price elasticities. Our findings suggest that for single product firms, the PCM computed from the dynamic estimates will be roughly 30 percent higher than those computed from static estimates. The bias is even larger for multiproduct firms because according to the dynamic estimates the products are closer substitutes than the static estimates suggest. Policy analysis based on static elasticity estimates will underestimate price–cost margins and underpredict the effects of mergers.

1.1. *Literature Review*

There are several empirical studies of sales in the economics literature. Pesendorfer (2002) studies sales of ketchup. In his model, he shows that the decision to hold a sale is a function of the duration since the last sale. His empirical analysis shows that both the probability of holding a sale and the aggregate quantity sold (during a sale) are a function of the duration since the last sale. Boizot, Robin, and Visser (2001) study dynamic consumer choice with inventory. They show that duration from previous purchase increases in current price and declines in past price, and quantity purchased increases in past prices. Hosken, Matsa, and Reiffen (2000) study the probability of a product being on sale as a function of its attributes. Aguirregabiria (1999) studies retail inventory behavior and its effect on prices.

The closest paper to ours is Erdem, Imai, and Keane (2003). They also structurally estimate a model of demand in which consumers can store different varieties of the product. To overcome the computational complexity of the problem, they make several simplifications that we discuss in detail in Section 4.3.3. Their estimation method is more computationally burdensome, but richer in unobserved heterogeneity. Our method, on the other hand, can handle patterns of observed heterogeneity more flexibly and, due to the computational simplicity, can handle a larger choice set. Erdem, Imai, and Keane study the role of price expectations. They compare consumers responses' to price cuts, both allowing for the price cut to affect future price expectations and holding expectations fixed. They also use the estimates to simulate consumer responses to short-run and long-run price changes. We compare the models in more detail in Section 4.

There is a vast literature in marketing that studies the effect of sales and its link to stockpiling (Shoemaker (1979), Blattberg, Eppen, and Lieberman (1981), Neslin, Henderson, and Quelch (1985), Currim and Schneider (1991), Gupta (1988, 1991), Chiang (1991), Bell, Chiang, and Padmanabhan (1999)). Blattberg, Eppen, and Lieberman (1981) report that in the four categories they study the increase in duration to next purchase during sales is 23–36 percent. Neslin, Henderson, and Quelch (1985), using similar data, find acceleration of purchases for coffee and bathroom tissue as a reaction to advertised price cuts. Subsequent work (Currim and Schneider (1991), Gupta (1988, 1991)) finds effects of the same magnitude as Blattberg, Eppen, and Lieberman (1981). Neslin and Schneider Stone (1996), in surveying the literature, report estimates of purchase acceleration that range between 14 and 50 percent.

The more recent literature concentrates on decomposing consumers' response to sales. Gupta (1988) decomposes price responses into brand switching, acceleration of purchases, and stockpiling. In his coffee sample, 85 percent of the price response corresponds to brand switching, 14 percent is due to demand acceleration, and 2 percent relates to stockpiling. Chiang (1991) reports similar findings. Bell, Chiang, and Padmanabhan (1999) show that the primary demand expansion (the expansion in quantity sold in the category) is larger than the previous studies reported. Ailawadi and Neslin (1998) and Bell and Boztug (2004) study consumption effects to separate whether the sales expansions are caused by extra consumption or stockpiling. They report substantial consumption effects in their samples, explaining from 12 to 60 percent of increase in quantity sold.

2. DATA, INDUSTRY, AND PRELIMINARY ANALYSIS

2.1. *Data*

We use a scanner data set collected by Information Resources Inc. (IRI) from June 1991 to June 1993 that has two components: store-level and household-level data. The first data were collected using scanning devices in nine supermarkets, belonging to different chains, in two separate submarkets in a large midwest city. From the store-level data we know for each detailed product (brand–size) in each store in each week the price charged, (aggregate) quantity sold, and promotional activities. The second component of the data set is at the household level. We observe the purchases of households over a two year period. We know when a household visited a supermarket and how much they spent in each visit. The data include purchases in 24 different product categories for which we know which product each household bought, where it was bought, and how much was paid.

Table I displays statistics of some household demographics, laundry detergent purchases (the product we focus on below), and store visits in general. The typical (median) household buys a single container of laundry detergent every 4 weeks. This household buys three different brands over the 104 weeks. Since the household-level brand HHI is roughly 0.5, purchases are concentrated at two main brands. However, because preferred brands differ by household, the

TABLE I
SUMMARY STATISTICS OF HOUSEHOLD-LEVEL DATA[a]

	Mean	Median	Std	Min	Max
Demographics					
Income (000's)	35.4	30.0	21.2	<10	>75
Size of household	2.6	2.0	1.4	1	6
Live in suburb	0.53	—	—	0	1
Purchase of laundry detergents					
Price ($)	4.38	3.89	2.17	0.91	16.59
Size (oz.)	80.8	64	37.8	32	256
Quantity	1.07	1	0.29	1.00	4
Duration (days)	43.7	28	47.3	1	300
Number of brands bought over the 2 years	4.1	3	2.7	1	15
Brand HHI	0.53	0.47	0.28	0.10	1.00
Store visits					
Number of stores visited over the 2 years	2.38	2	1.02	1	5
Store HHI	0.77	0.82	0.21	0.27	1.00

[a]For Demographics, Store visits, Number of brands, and Brand HHI, an observation is a household. For all other statistics, an observation is a purchase instance. Brand HHI is the sum of the square of the volume share of the brands bought by each household. Similarly, Store HHI is the sum of the square of the expenditure share spent in each store by each household.

market-level shares are not as concentrated. Finally, the typical household buys mainly at two stores, with most of the purchases concentrated at a single store.

2.2. *The Industry*

Laundry detergents come in two main forms: liquid and powder. Liquid detergents account for 70 percent of the quantity sold. Unlike many other consumer goods, there are a limited number of brands offered. The shares within each segment (i.e., liquid and powder) are presented in the first column of Table II. The top 11 brands account for roughly 90 percent of the quantity sold.

Most brand–size combinations have a regular price. In the sample, the price is at the modal level for 71 percent of the weeks and above it only approximately 5 percent of the time. Defining a sale as any price at least 5 percent below the modal price of each UPC (universal product code) in each store,[2] we find that in our sample 43 and 36 percent of the volume sold of liquid and powder detergent, respectively, was sold during a sale. The median discount during a sale is $0.40, the average is $0.67, the 25th percentile is $0.20, and the

[2]This definition of a sale is not appropriate in cases where the "regular" price shifts due to seasonality or any other reason. This does not seem to be the case in this industry. The definition of a sale matters only for the descriptive analysis in this section. We do not use it in the structural estimation.

TABLE II

BRAND VOLUME SHARES AND FRACTION SOLD ON SALE[a]

	Liquid					Powder				
	Brand	Firm	Share	Cumulative	% on Sale	Brand	Firm	Share	Cumulative	% on Sale
1	Tide	P & G	21.4	21	32.5	Tide	P & G	40	40	25.1
2	All	Unilever	15	36	47.4	Cheer	P & G	14.7	55	9.2
3	Wisk	Unilever	11.5	48	50.2	A & H	C & D	10.5	65	28
4	Solo	P & G	10.1	58	7.2	Dutch	Dial	5.3	70	37.6
5	Purex	Dial	9	67	63.1	Wisk	Unilever	3.7	74	41.2
6	Cheer	P & G	4.6	72	23.6	Oxydol	P & G	3.6	78	59.3
7	A & H	C & D	4.5	76	21.5	Surf	Unilever	3.2	81	11.6
8	Ajax	Colgate	4.4	80	59.4	All	Unilever	2.3	83	
9	Yes	Dow Chemical	4.1	85	33.1	Dreft	P & G	2.2	86	15.2
10	Surf	Unilever	4	89	42.5	Gain	P & G	1.9	87	16.7
11	Era	P & G	3.7	92	40.5	Bold	P & G	1.6	89	1.1
12	Generic	—	0.9	93	0.6	Generic	—	0.7	90	16.6
13	Other	—	0.2	93	0.9	Other	—	0.6	90	19.9

[a]Columns labeled Share are shares of volume (of liquid or powder) sold in our sample, columns labeled Cumulative are the cumulative shares, and columns labeled % on Sale are the percent of the volume, for that brand, sold on sale. A sale is defined as any price at least 5 percent below the modal price, for each UPC in each store. A & H = Arm & Hammer; P & G = Procter and Gamble; C & D = Church and Dwight.

75th percentile is $0.90. In percentage terms, the median discount is 8 percent, the average is 12 percent, and the 25th and 75th percentiles are 4 and 16 percent, respectively. As shown in Table II, there is some variation across brands in the share of the quantity sold on sale.

Detergents come in several sizes. However, about 97 percent of the volume of liquid detergent sold was sold in five different sizes.[3] Sizes of powder detergent are not quite as standardized. Prices are nonlinear in size. The first column in Table III shows the price per 16 oz. unit for several container sizes. Since not all sizes of all brands are sold in all stores, reporting the average price per unit for each size is misleading. Instead, we report the ratio of the size dummy variable to the constant, from a regression of the modal price per 16 ounce regressed on size, brand and store dummy variable.

The data record two types of promotional activities: *feature* and *display*. The *feature* variable measures whether the product was advertised by the retailer (e.g., in a retailer bulletin sent to consumers that week). The *display* variable captures whether the product was displayed differently than usual within the

[3]Toward the end of the sample, ultra detergents were introduced. These detergents are more concentrated and therefore a 100 oz. bottle is equivalent to a 128 oz. bottle of regular detergent. For the purpose of the following numbers we aggregated 128 oz. regular with 100 oz. ultra and 68 oz. regular with 50 oz. ultra.

TABLE III
QUANTITY DISCOUNTS AND SALES[a]

	Quantity Discount (%)	Quantity Sold on Sale (%)	Weeks on Sale (%)	Average Sale Discount (%)	Quantity Share (%)
Liquid					
32 oz.	—	2.6	2.0	11.0	1.6
64 oz.	18.1	27.6	11.5	15.7	30.9
96 oz.	22.5	16.3	7.6	14.4	7.8
128 oz.	22.8	45.6	16.6	18.1	54.7
256 oz.	29.0	20.0	9.3	11.8	1.6
Powder					
32 oz.	—	16.0	7.7	14.5	10.1
64 oz.	10.0	30.5	16.6	12.9	20.3
96 oz.	14.9	17.1	11.5	11.7	14.4
128 oz.	30.0	36.1	20.8	15.1	23.2
256 oz.	48.7	12.9	10.8	10.3	17.3

[a]All cells are based on data from all brands in all stores. The column labeled Discount presents percent quantity discount (per unit) for the larger sizes, after correcting for differences across stores and brands (see text for details). The columns labeled Quantity Sold on Sale, Weeks on Sale, and Average Sale Discount present, respectively, the percent quantity sold on sale, percent of weeks a sale was offered, and average percent discount during a sale, for each size. A sale is defined as any price at least 5 percent below the modal. The column labeled Quantity Share is the share of the total quantity (measured in ounces) sold in each size.

store that week.[4] The correlation between a sale—defined as a price below the modal—and being featured is 0.38. Conditional on being on sale, the probability of being featured is less than 20 percent, while conditional on being featured, the probability of a sale is above 93 percent. The correlation with *display* is even lower, at 0.23, due to the large number of times that the product is displayed but not on sale. Conditional on a display, the probability of a sale is only 50 percent. If we define a sale as the price less than 90 percent of the modal price, both correlations increase slightly, to 0.56 and 0.33, respectively.

2.3. *Preliminary Analysis*

Further descriptive analysis of the data used in this paper can be found in Hendel and Nevo (2006), where we find several patterns consistent with consumer inventory behavior. In line with the marketing findings (summarized above), using the aggregate data we find that duration since previous sale has a positive effect on the aggregate quantity purchased, both during sale and nonsale periods. Both these effects are consistent with an inventory model because the longer is the duration from the previous sale, on average, the lower is the inventory each household currently has, making purchase more likely. Second,

[4]Both these variables have several categories (for example, type of display is end, middle, or front of aisle). We aggregate across these classifications to feature/no feature and display/no display.

we find that indirect measures of storage costs are negatively correlated with households' tendency to buy on sale. Third, both for a given household over time and across households, we find a significant difference between sale and nonsale purchases in both duration from previous purchase and duration to next purchase. To take advantage of the low price, during a sale households buy at higher levels of current inventory. Namely, duration to previous purchase is shorter during a sale. Furthermore, during a sale households buy more and therefore, on average, it takes longer until the next time their inventory crosses the threshold for purchase. Finally, we find that the pattern of sales and purchases during sales across different product categories is consistent with the variation in storage costs across these categories.

3. THE MODEL

3.1. *The Basic Setup*

Consumer h obtains per period utility

$$u(c_{ht}, \nu_{ht}; \boldsymbol{\theta}_h) + \alpha_h m_{ht},$$

where c_{ht} is the quantity consumed of the good in question, ν_{ht} is a shock to utility that changes the marginal utility from consumption, $\boldsymbol{\theta}_h$ is a vector of consumer-specific taste parameters, m_{ht} is the consumption of the outside good, and α_h is the marginal utility from consuming the outside good. The product is offered in J different varieties or brands. The total consumption of all varieties in period t is $c_{ht} = \sum_j c_{jht}$. The stochastic shock ν_{ht} introduces randomness into the consumer's needs, unobserved by the researcher. For simplicity we assume the shock to utility is additive in consumption: $u(c_{ht}, \nu_{ht}; \boldsymbol{\theta}_h) = u(c_{ht} + \nu_{ht}; \boldsymbol{\theta}_h)$. High realizations of ν_{ht} decrease the household's need, decrease demand and make it more elastic. Consumers face random and potentially nonlinear prices.

Consumers in each period decide which brand to buy, how much to buy, and how much to consume.[5] Because the good is storable, quantity not consumed is stored as inventory. As shown in the Appendix, consumption is not affected by which brand is in storage. In the estimation we assume that the purchased amount, denoted by x_{ht}, is simply a choice of size (i.e., consumers choose container size, not how many containers).[6] We denote a purchase of

[5]Instead of making consumption a decision variable, we could assume an exogenous consumption rate, either deterministic or random. Both these alternative assumptions, which are nested within our framework, would simplify the estimation. However, it is important to allow consumption to vary in response to price changes because this is the main alternative explanation to why consumers buy more during sales. Reduced form results in Ailawadi and Neslin (1998), Hendel and Nevo (2006), and Bell and Boztug (2004) report consumption effects for several products.

[6]More than 97 percent of the purchases are for a single unit. In principle, the model could allow for multiple purchases by expanding the choice set to allow for bundles.

brand j and size x by $d_{hjxt} = 1$, where $x = 0$ stands for no purchase, and we assume $\sum_{j,x} d_{hjxt} = 1$. We denote by p_{jxt} the price associated with purchasing x units (or size x) of brand j. The consumer's problem can be represented as

$$
\begin{aligned}
(1)\qquad & V(s_1) \\
&= \max_{\{c_h(s_t), d_{hjx}(s_t)\}} \sum_{t=1}^{\infty} \delta^{t-1} E\Big[u(c_{ht}, \nu_{ht}; \boldsymbol{\theta}_h) - C_h(i_{h,t+1}; \boldsymbol{\theta}_h) \\
&\qquad\qquad + \sum_j d_{hjxt}(\alpha_h p_{jxt} + \xi_{hjx} + \beta_h a_{jxt} + \varepsilon_{hjxt}) \Big| s_1 \Big] \\
&\text{s.t.}\quad 0 \le i_{ht}, \quad 0 \le c_{ht}, \quad 0 \le x_{ht}, \quad \sum_{j,x} d_{hjxt} = 1, \\
&\qquad\quad i_{h,t+1} = i_{ht} + x_{ht} - c_{ht},
\end{aligned}
$$

where s_t denotes the state at time t, $\delta > 0$ is the discount factor, $C_h(i; \theta_h)$ is the cost of storage, ξ_{hjx} is an idiosyncratic taste for brand j that could be a function of brand characteristics or size and varies across consumers, $\beta_h a_{jxt}$ captures the effect of advertising variables on consumer choice, and ε_{hjxt} is a random shock to consumer choice. The latter is size-specific, namely, different sizes get different draws, introducing randomness in the size choice as well. To simplify notation we drop the subscript h in what follows.

At time t, the state s_t consists of the current (or beginning-of-period) inventory i_t, current prices, the shock to utility from consumption ν_t, and the vector of epsilons. In the estimation the parameters in equation (1) are allowed to vary by household. Consumers face two sources of uncertainty in the model: future utility shocks and random future prices. We make the following assumptions about the distribution of these variables.

ASSUMPTION 1: *ν_t is independently distributed over time and across consumers.*

In principle, serial correlation in ν_t can be allowed, but at a significant increase in computational burden.

ASSUMPTION 2: *Prices (and advertising) follow an exogenous first-order Markov process.*

The assumption of a first-order Markov process reduces the state space. It is somewhat problematic because it is hard to come up with a supply model that yields equilibrium prices that follow a first-order process. On the other hand, a first-order process is a reasonable assumption about consumers' memory and formation of expectations. With our assumption a consumer sees the current prices at the store and forms expectations regarding future prices. A higher order process would require that she recall prices on previous visits. In the application we explore higher order processes, but for now we maintain the first-order assumption.

Assumption 2 implies that the price process is exogenous, namely, conditional on the control variables, the process is independent of the unobserved random components. Given that we use household-level data and can control for many variables, this assumption is reasonable. The main concern might be seasonality or periods of high demand when the likelihood of a sale changes. This is probably not a concern for the product we study herein, but if seasonality is present, it should be modeled into the price process.

ASSUMPTION 3: *ε_{jxt} is independently and identically distributed extreme value type* 1.

This assumption significantly reduces the computational burden. It can be relaxed. We can allow for correlation in the unobserved shocks between brands by assuming that the epsilons are distributed according to some distributions in the generalized extreme value (GEV) class.[7] For simplicity, we present the independently and identically distributed case.

Product differentiation as it appears in equation (1) takes place at the moment of purchase. Taken literally, product differences affect the behavior of the consumer at the store but different brands do not give different utilities at the moment of consumption. This assumption helps reduce the state space. Instead of the whole vector of brand inventories, only the total quantity in stock matters, regardless of brand.

Differentiation at purchase, represented by the ξ_{hjx} term in equation (1), is a way to capture the expected value of the future differences in utility from consumption. The term ξ_{jx} captures the expected utility from the future consumption of the x units of brand j at the time of purchase. This approach is valid as long as (i) brand-specific differences in the utility from consumption enter linearly in the utility function[8] and (ii) discounting is low.[9]

For the brand-specific differences in the utility from consumption to be linear, utility differences from consuming the same quantities of different brands must be independent of the bundle consumed. Thus, we rule out interactions in consumption that arise if, for instance, the marginal utility from consuming Cheerios depends on the consumption of Trix. These interactions between brands are ruled out in standard static discrete choice models.

[7]We have to restrict the distribution to a subset of the GEV family: those with a GEV generating function that is separable across sizes.

[8]Preferences need not be linear. We actually allow for a nonlinear utility from consumption, $u(c)$. Only the brand-specific differences need to enter linearly.

[9]Two issues arise. First, because the timing of consumption is uncertain, the present value of the utility from consumption becomes uncertain ex ante. Second, with discounting, the order in which the different brands already in storage are consumed becomes endogenous.

4. ECONOMETRICS

The structural estimation is based on the nested algorithm proposed by Rust (1987), but has to deal with issues special to our problem. We start by providing a general overview of our estimation procedure and then discuss more technical details.

4.1. *An Overview of the Estimation*

Rust (1987) proposes an algorithm based on nesting the (numerical) solution of the consumer's dynamic programming problem within the parameter search of the estimation. The solution to the dynamic programming problem yields the consumer's deterministic decision rules (in terms of purchases and consumption). However, because we do not observe the random shocks, which are state variables, from our perspective the decision is stochastic. Assuming a distribution for the unobserved shocks, we derive the likelihood of observing each consumer's decision conditional on prices and inventory. We nest this computation of the likelihood into the search for the values of the parameters that maximize the likelihood of the observed sample.

We face two hurdles in implementing the algorithm. First, we do not observe inventory since both initial inventory and consumption decisions are not observed. We deal with the unknown inventories by using the model to derive the optimal consumption in the following way. Assume for a moment that the initial inventory is observed. We can use the procedure described in the previous paragraph to obtain the likelihood of the observed purchases and the optimal consumption levels (which depend on ν), and, therefore, the end-of-period inventory levels. For each inventory level, we can again use the procedure of the previous paragraph to obtain the likelihood of observed purchase in the subsequent period. Repeating this procedure, we obtain the likelihood of observing the whole sequence of purchases for each household. To start this procedure, we need a value for the initial inventory. The standard approach is to use the estimated distribution of inventories to generate the initial distribution. In practice, we do so by starting at an arbitrary initial level and using part of the data (the first few observations) to generate the distribution of inventories implied by the model.

Formally, for a given value of the parameters, the probability of observing a sequence of purchasing decisions, $(d_1, \dots, d_T)$ as a function of the observed state variables $(p_1, \dots, p_T)$ is

$$\text{(2)} \qquad \Pr(d_1 \cdots d_T | p_1 \cdots p_T) = \int \prod_{t=1}^{T} \Pr\big(d_t | p_t, i_t(d_{t-1}, \dots, d_1, \nu_{t-1}, \dots, \nu_1, i_1), \nu_t\big)\, dF(\nu_1, \dots, \nu_T)\, dF(i_1).$$

Note that the beginning-of-period inventory is a function of previous decisions—the previous consumption shocks and the initial inventory. Note also that p_t includes all observed state variables, not just prices, for instance, promotional activities denoted a_t in equation (1). The probability inside the integral on the right-hand side of equation (2) represents integration over the set of epsilons that induce d_t as the optimal choice and is given by

$$\text{(3)} \qquad \Pr(d_{jx}|p_t, i_t, \nu_t) = \frac{\exp(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + M(s_t, j, x))}{\sum_{k,y} \exp(\alpha p_{kyt} + \xi_{ky} + \beta a_{kyt} + M(s_t, k, y))},$$

where $M(s_t, j, x) = \text{Max}_c\{u(c+\nu_t) - C(i_{t+1}) + \delta E(V(s_{t+1})|d_{jx}, c, s_t)\}$ and $E(V)$ is the expectation of the future V as a function of today's state and actions. Note that the summation in the denominator is over all brands and all sizes.

The second problem is the dimensionality of the state space. Households in our sample buy several brand–size combinations, which are offered at many different prices. The state space includes not only the individual-specific inventory and shocks, but also the prices of all brands in all sizes and their promotional activities. The state space and the transitions probabilities across states, in full generality, make the standard approach computationally infeasible.

We therefore propose the following three-step procedure. The first step consists of maximizing the likelihood of observed brand choice *conditional* on the size (quantity) bought in order to recover the marginal utility of income α and the parameters that measure the effect of advertising, β and ξ. As we show below, we do not need to solve the dynamic programming problem to compute this probability. We estimate a static discrete choice model, restricting the choice set to options of the same size (quantity) actually bought in each period. This estimation yields consistent, but potentially inefficient, estimates of these parameters. In the second step, using the estimates from the first stage, we compute the *inclusive values* associated with each size (quantity) and their transition probabilities from period to period. Finally, we apply the nested algorithm to the simplified dynamic problem. We estimate the remaining parameters by maximizing the likelihood of the observed sequence of sizes (quantities) purchased. The simplified problem involves quantity choices exclusively. Rather than have the state space include prices of all available brand–size combinations, it includes only a single "price" (or inclusive value) for each size.

Intuitively, the time independent nature of the shocks enables the decomposition of the likelihood of the individual choices into two components that can be separately estimated. First, at any specific point in time, when the consumer purchases a product of size x, we can estimate her preferences for different brands. Second, we can estimate the parameters that determine the dynamic (storing) behavior of the consumer by looking at a simplified version of the problem, which treats each size as a single choice.[10]

4.2. *The Three-Step Procedure*

We now show that the breakup of the likelihood follows from the primitives of the model.

STEP 1—Estimation of the "Static" Parameters: To simplify the computation of the likelihood, note that the probability of choosing brand j and size x can be written as

$$\Pr(d_{jt}=1, x_t|p_t, i_t, \nu_t) = \Pr(d_{jt}=1|p_t, x_t, i_t, \nu_t)\Pr(x_t|p_t, i_t, \nu_t).$$

In general, this does not help because we need to solve the consumer's dynamic programming problem to be able to compute $\Pr(d_{jt}=1|p_t, x_t, i_t, \nu_t)$. However, given the preceding model and assumptions, we can compute this conditional probability without solving the dynamic programming problem. We rely on two results. The optimal consumption, conditional on the quantity purchased, is not brand-specific (see Lemma 1 in the Appendix). Thus, the term $M(s_t, j, x)$ is independent of brand choice. This in itself is not enough to deliver the result because the conditional distribution of epsilon might be a function of x. If indeed the actual choice (given the observed state, i.e., prices) conveyed information about the distribution of the epsilons, we would have to fully solve the model to figure out the conditional distribution of the error (more details in Section 4.3.2). Assumption 3 implies that, after the appropriate cancellations,

$$\Pr(d_{jt}=1|x_t, i_t, p_t, \nu_t) = \frac{\exp(\alpha p_{jxt} + \xi_{kx} + \beta a_{kxt})}{\sum_k \exp(\alpha p_{kxt} + \xi_{kx} + \beta a_{kxt})} = \Pr(d_{jt}=1|x_t, p_t), \tag{4}$$

where the summation is over all brands available in size x_t at time t.

Thus, our approach is to estimate the parameters that affect brand choice (i.e., marginal utility of income, the vector of parameters β, and brand-specific and potentially household-specific fixed effects) by maximizing the product, over time and households, of $\Pr(d_{jt}=1|p_t, x_t)$. This amounts to estimating a

[10]We are aware of two instances in the literature that use similar ideas. First, one way to estimate a static nested logit model is first to estimate the choice within a nest, compute the inclusive value, and then to estimate the choice among nests using the inclusive values (Train (1986)). Second, in a dynamic context a similar idea was proposed by Melnikov (2001). In his model (of purchase of durable products) the value of all future options enters the current no-purchase utility. He summarizes this value by the inclusive value.

static brand choice model where the choice set includes only brands offered in the same size as the actual purchase.

STEP 2—Inclusive Values: To compute the likelihood of a sequence of purchased quantities, we show in the next section that we can simplify the state space of the dynamic programming problem. The state variables of the simplified problem are the inclusive values of each size (quantity), which can be thought of as a quality adjusted price index for all brands of size x. They are given by

$$\omega_{xt} = \log\left\{\sum_k \exp(\alpha p_{kxt} + \xi_{kx} + \beta a_{kxt})\right\}. \tag{5}$$

All the information needed to compute the inclusive values and their transition probabilities from period to period is contained in the first stage estimates.

We show below that the dynamic problem can be rewritten in terms of inclusive values, so that the state space collapses to a single index per size. For this to work, we make the following assumption:

ASSUMPTION 4: *$F(\omega_t|s_{t-1})$ can be summarized by $F(\omega_t|\omega_{t-1})$.*

Instead of keeping track of the prices of ten brands times four sizes (roughly the dimensions in our data), we have to follow only four quality adjusted prices. If the marginal utility from income and the brand preferences are household-specific, then the inclusive values and their future distribution $F(\omega_t|\omega_{t-1})$ will be household-specific. A household-specific process requires that we solve the dynamic program separately for each household. In the application, we group households into six types, letting the process vary by type.

The main loss when the process is defined this way is that transition probabilities are exclusively a function of the inclusive values. Two price vectors that yield the same vector of inclusive values will have the same transition probabilities to next period's state. This restriction is testable and to some extent can be relaxed. In the regression of current on previous inclusive values, we can add vectors of previous prices. Under our assumption, previous prices should not matter independently once we control for the vector of current inclusive values. A full fix is to allow the distribution of the inclusive values to depend on the whole vector of current prices. This would naturally undo part of the computational advantage of the inclusive values.[11]

[11] If we assume that the inclusive values depend on the whole vector of prices the third stage becomes more computationally demanding, but the split remains valid.

STEP 3—The Simplified Dynamic Problem: In the third step we estimate the remaining dynamic parameters by maximizing the likelihood of purchases implied by the simplified problem. In the simplified problem, the consumer chooses whether and what quantity to purchase. Her flow utility of purchasing quantity x is given by $\omega_{xt} + \varepsilon_{xt}$. The inclusive value ω_{xt} is the utility expected by the consumer from all brands of size x in the original problem before knowing the realizations of ε_{jxt} (McFadden (1981)). The consumer in the simplified problem observes only a summary of the state variables (given by the inclusive values) and decides what quantity to purchase. The Bellman equation associated with the simpler problem is

$$\begin{aligned} &V(i_t, \omega_t, \varepsilon_t, \nu_t) \\ &\quad = \underset{\{c,x\}}{\mathrm{Max}} \big\{ u(c+\nu_t) - C(i_{t+1}) + \omega_{xt} + \varepsilon_{xt} \\ &\qquad\qquad + \delta E[V(i_{t+1}, \omega_{t+1}, \varepsilon_{t+1}, \nu_{t+1}) | i_t, \omega_t, \varepsilon_t, \nu_t, c, x] \big\}. \end{aligned}$$

Using the estimated inclusive values and the transition probabilities, we estimate the utility and the storage cost functions by applying the nested algorithm to compute the likelihood of purchasing a size (quantity). The following claim shows the equivalence of the likelihood computed from the original problem and that computed from the simplified one, thus establishing the validity of using the simpler problem for estimation.

CLAIM 1: $\Pr(x_t | i_t, p_t, \nu_t) = \Pr(x_t | i_t, \omega(p_t), \nu_t)$, *where the left-hand side is the probability of choosing x (summing over brand choices) in the original problem defined in equation* (3), *and the right-hand side is the probability of choosing x in the simplified problem.*

There are two steps to prove this claim (see the Appendix). The first step is to note that the solutions to the original and the simplified problems are characterized by the same $E(V)$. The original problem is characterized by the Bellman equation

$$\begin{aligned} V(s_t) = \underset{\{c,d_{jx}\}}{\mathrm{Max}} \Bigg\{ & u(c+\nu_t) - C(i_{t+1}) + \sum_{j,x} d_{jx}(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \varepsilon_{jxt}) \\ & + \delta E[V(s_{t+1}) | s_t, c, d_{jx}] \Bigg\}. \end{aligned}$$

The equivalence is not obvious, because in one case the consumer observes the whole state, s_t, when she decides what to purchase, whereas in the other case she picks a quantity x after having observed only the expected utility from that quantity, namely ω_{xt}. The key to the equivalence is in Assumption 3, which allows us to obtain a summary of the expected utility from all products of a particular size, and in Assumption 4, which enables us to write the transition probabilities as a function of the inclusive values exclusively.

The second step of the proof shows the equivalence of the probability of choosing size x in the original problem, defined by summing over the brand choice probabilities given in equation (3), and the choice probability implied by the simplified model, where size x is a single choice evaluated at the inclusive value generated by the vector p_t. This follows almost directly from the definition of the inclusive values (and Assumption 3).

Despite the reduction in the number of state variables, the value function of the simpler problem is still computationally burdensome to solve. To solve it we use value function approximation with policy function iteration. We closely follow Benitez-Silva, Hall, Hitsch, Pauletto, and Rust (2000), where further details can be found. Briefly, the algorithm consists of iterating the alternating steps of policy evaluation and policy improvement. The value function is approximated by a polynomial function of the state variables. The procedure starts with a guess of the optimal policy at a finite set of points in the state space. Given this guess, and substituting the approximation for both the value function and the expected future value, one can solve for the coefficients of the polynomial that minimize the distance (in a least squares sense) between the two sides of the Bellman equation. Next the guess of the policy is updated. This is done by finding, for every state, the action that maximizes the sum of current return and the expected discounted value of the value function. The expected value is computed using the coefficients found in the first step and the expected value of the state variables. We perform this step analytically. These two steps are iterated until convergence (of the coefficients of the approximating function). The output of the procedure is an approximating function that can be used to evaluate the value function (and the expected value function) at any point in the state space. See Bertsekas and Tsitsiklis (1996) and Benitez-Silva et al. (2000) for more details on this procedure, as well as its convergence properties.

4.3. *Identification, Heterogeneity, and Alternative Methods*

4.3.1. *Identification*

We informally discuss identification. The identification of the static parameters is standard. Variation over time in prices and advertising identify the household's sensitivity to price and to promotional activities, denoted by α and β. As we pointed out in Section 2.2, sales are not perfectly correlated with feature and display activity, and therefore the effects can be separately identified. Brand and size effects are identified in the first stage from variations in shares across products. Household heterogeneity in the static parameters is captured by making the sensitivity to promotional activities, brand effects, and size effects functions of household demographics, as well as allowing for household-specific brand effects.

The identification of the dynamic parameters is more subtle. The third stage involves the estimation of the utility and storage cost parameters that maximize the likelihood of the observed sequence of quantities purchased (container sizes) over the sample period. If inventory and consumption were observed, then identification would follow the standard arguments (see Rust (1996), Magnac and Thesmar (2002), and Aquirregabiria (2005)). However, we do not observe inventory or consumption, so the question is what feature of the data allows us to identify functions of these variables?

The data tell us about the probability of purchase conditional on current prices (i.e., the current inclusive values) and past purchases (amounts purchased and duration from previous purchases). Suppose that we see that this probability is not a function of past behavior. Then we would conclude that consumers are purchasing for immediate consumption and not for inventory. On the other hand, if we observe that the purchase probability is a function of past behavior and we assume that preferences are stationary, then we conclude that there is dynamic behavior. Regarding storage costs, consider the following example. Suppose we observe two consumers who face the same price process purchase the same amount over a given period. However, one of them purchases more frequently than the other. This variation will lead us to conclude that this consumer has higher storage costs.

More generally, given the process of the inclusive values, preferences and storage costs determine consumer behavior. For a given storage cost function, preferences determine the level of demand. In contrast, given preferences, different storage cost levels determine interpurchase duration and the extent to which consumers can exploit price reductions. Higher storage costs reduce consumers' benefit from sales and shorten the average duration between purchases. In the extreme case of no storage (or very high storage cost), interpurchase duration depends exclusively on current prices, because the probability of current purchase is independent of past purchases. This suggests a simple way to test the relevance of the inventory model, based on the impact of previous purchases on current behavior. Preferences and storage costs are recovered from the relationship between purchases, prices, and previous purchases. We have no reason to believe that costs and preferences are identified nonparametrically or even that flexible functional forms can be estimated.

The split between the quantity purchased and brand choice provides some insight into the determinants of demand elasticities. There are two sets of parameters that determine price responses. The static parameters recovered in stage one determine the substitutability across brands, while the utility and inventory cost parameters recovered in stage three determine the responsiveness to prices in the quantity dimension. Both sets of estimates are needed to simulate the responses to price changes.

The split provides both an intuitive interpretation of the determinants of substitution patterns as well as a shortcut to separate long-run from short-run price responses. The basic insight is that in order to purge elasticities from timing effects—as a first approximation—one can estimate demand at the individual level, conditional on the size of the purchase. This approximation might prove helpful when the full model is too complicated to estimate or the data are insufficient.

4.3.2. *Heterogeneity in brand preferences*

Three assumptions are critical to justify the split in the likelihood. We already discussed the implications and limitations of the first two—product differentiation is modeled as taking place at the time of purchase and Assumption 4. Also critical for the analysis is the assumption that the error terms ε_{jxt} are identically and independently distributed type I extreme value, both across choices and over time. As we mentioned, we can relax the independence across brands by allowing for a GEV distribution. Here we discuss ways to allow for persistent heterogeneity.

Our procedure allows for heterogeneity in brand preferences that is a function of observed household attributes and observed past behavior. However, we cannot allow for a persistent random component. One can see from equation (3) why the unobservable brand–size preferences must be uncorrelated over time. Suppose there are different (unobservable) types in the population that differ in their brand preferences or in their sensitivity to price and advertising. To compute the choice probabilities (unconditional on type), we have to integrate over the distribution of these types. The probability of brand choice conditional on size purchased and *conditional* on the type (of consumer) will still be independent of the various dynamic components. The problem, however, is that in order to compute the brand choice probability conditional on size (but unconditional on type) we have to integrate over the distribution of types *conditional* on the size purchased. Figuring out this distribution requires essentially solving the dynamic problem.

The main consequence of not having persistent preference shocks is that we cannot allow for random effects in the first stage of the estimation. Notice that unobserved heterogeneity can be introduced in the dynamic part of the model; indeed because of its computational simplicity, this can be done in a relatively general way. Random effects usually enter choice models in the brand preferences and in the marginal impact of price and other variables. Heterogeneity along these dimensions might generate interesting dynamics. For instance, consider the case of heterogeneity in brand preferences. A consumer with a high unobserved shock preference for Tide is likely not to react to deals on other brands and waits for deals on Tide if the shock is persistent. If her inventories are running low, she may prefer to purchase a small container of an alternative brand as she waits for Tide to go on sale.

Instead of random effects, we can allow for heterogeneity by including observed household heterogeneity, classifying the households into types, allowing for differences in the choice based on past behavior, and allowing for household-brand-specific fixed effects. Because the household-brand dummy variables are estimated in a static conditional logit, they do not substantially increase the computational cost.[12] Notice that we need only to estimate preferences for the brands actually purchased by each household. For brands that a household never bought in the sample (which are the majority), we just eliminate them from the household choice set. Absence from the choice set is equivalent to a sufficiently negative brand dummy. Clearly, we do not know if the household would buy this brand at prices outside those observed in the sample range, but that is a data limitation: the data do not inform us about out-of-sample prices anyway.

4.3.3. *Alternative methods*

The estimation of the full model without the assumptions that enable the split (of the likelihood) would not be tractable for most products that come in several sizes and brands. The dynamic problem would have an extremely large state, which includes the inventories of all brands held by the household as well as the price vector of all brands in all sizes, in addition to other promotional activities for each product–size combination.

Erdem, Imai, and Keane (2003) propose a different approach. Their solution to reduce the complexity of the problem is to assume that once the quantity to be consumed is determined, each brand in storage is consumed. Moreover, the consumption share of each brand is proportional to the share of that brand in storage. In other words, if the storage is made of 40% of one brand and 60% of the other, then consumption will be made up of both brands in those proportions. This, together with the assumption that brand differences in quality enter linearly in the utility function, implies that only the total inventory and a quality weighted inventory matter as state variables. To reduce the dimensionality of the price vector, they estimate only the price process of the dominant container size and assume that price differentials per ounce with other containers is independently and identically distributed. This simplifies the states and transitions from current to future prices, because only the prices of the dominant size are relevant state variables. Finally, to further simplify the state space, they ignore other promotional activities.

4.4. *Computing Standard Errors*

The standard errors computed in the third step of the estimation have to be adjusted because the inclusive values and the transition parameters were

[12]The estimation of fixed effects is, in principle, problematic in nonlinear models. There are basically two ways around this. First, extend the conditional logit approach (Chamberlain (1983)) to deal with household-brand fixed effects. Second, estimate the parameters using maximum likelihood. Because in our case T is large, assuming it grows asymptotically is not unreasonable, in which case the standard incidental parameters problem is not an issue. The estimation of hundreds of household-specific brand fixed effects is feasible because the likelihood function of conditional logit is well behaved.

estimated in the first and second steps. This can be done using the correction methods proposed by Murphy and Topel (1985). The key in this approach is to compute the derivative of the third step likelihood with respect to the parameters estimated in the first and second steps. These derivatives can be evaluated numerically: we perturb each parameter slightly, resolve the dynamic programming problem, and see the impact on the likelihood. This computation is simplified significantly because the derivatives with respect to the first stage parameters can be written as the derivatives of the likelihood with respect to the inclusive value times the derivative of the inclusive value with respect to the parameter. The latter is easy to compute analytically, which significantly reduces the number of perturbations required of the third step. In the results below the correction increased the standard errors by less than 10 percent for most parameters.

5. RESULTS

To estimate the model, we have to choose functional forms. We assume $u(c_t + \nu_t) = \gamma \log(c_t + \nu_t)$, $C(i_t) = \beta_1 i_t + \beta_2 i_t^2$, and ν_t is distributed log normal. For the evolution in the inclusive values, we assume that ω_{st} is distributed normal with mean $\gamma_{s0} + \gamma_{s1}\omega_{1,t-1} + \cdots + \gamma_{s4}\omega_{4,t-1}$ and standard deviation σ_s for $s = 1, \dots, 4$, where the actual size is 32 oz. times s. We explore below adding more lags, the normality assumption, and letting the process vary by household type. The dynamic programming problem was solved by parametric policy approximation. The approximation basis used is a polynomial in the natural logarithm of inventory and levels of the other state variables. We will describe various ways in which we tested the robustness of the results. The estimation was performed using a sample of 218 households. The households were selected based on the following criteria: (i) they made more than 10 purchases of detergents, (ii) but no more than 50 purchases, and (iii) at least 75 percent of their purchases of detergents were of liquid detergent.[13] The households in the sample visit a supermarket 45–103 times. These households made 3,772 purchases of liquid detergents. The third stage is estimated using visits 11–35 of each household. The first 10 visits are used to generate the distribution of initial inventories. The late visits are eliminated to avoid right censoring in a uniform way. In what follows, we limit the analysis to liquid detergent. This product is easier to handle, more concentrated, and offered in fewer container sizes.

[13]Only a couple of household were eliminated because they purchased more than 50 times. We dropped these households because their purchases were so frequent—in one case essentially every week and multiple units each time—that we did not think our model applied to them. The number of households that purchased less than 10 times is more numerous, but of little impact on the elasticities (as they amount to a small share of total sales). We think it is safer to keep them out of the sample because they are candidates to purchase detergent in alternative stores.

5.1. *Parameter Estimates*

The parameter estimates are presented in Tables IV–VI. Table IV presents the estimates from the first stage, which is a (static) conditional logit brand choice conditional on size. This stage was estimated using choices by all households in the sample, where the choice set was restricted to products of the same size as the observed purchase. We introduce heterogeneity in three ways. First, we interact price with observable household characteristics. Second, we allow for household–brand specific effects. This is doable in our sample because each household purchases a small number of brands and we have a relatively long time series. Third, we eliminate from the household choice set brands not purchased over the whole sample period by that household. This reduces the number of choices significantly because most households purchase a small number of brands (the median household purchases three brands; see Table I).

The different columns in Table IV present specifications that vary in the variables included. We conclude three things from this table. First, we note the effect of including feature and display on the price coefficient, which can be seen by comparing columns (ii) and (iii) (as well as (ix) and (x)). Once feature and display are included, the price coefficient is roughly cut by half, which implies that the price elasticities are roughly 50 percent smaller.[14] The size of the change is intuitive. It implies that the large effect on quantity sold seemingly associated with price changes is largely driven by the feature and display promotional activity. This suggests that one has to be careful when interpreting estimates that do not properly control for promotional activities other than sales.[15]

More importantly these effects highlight one of the advantages of our approach: we can easily control for various observed variables. An alternative approach proposed by Erdem, Imai, and Keane (2003) can, in principle, incorporate promotional activities, but due to the computational cost they do not do so in practice.

Second, we see the impact of heterogeneity in columns (iv)–(x), where we interact price with three demographic variables: dummy variables that equal one if the family lives in the suburban market, if the family is nonwhite, and if the family has more than four people. These interactions are highly significant. The signs on the coefficients make sense. Larger families, nonwhite people (who in our sample have lower income), and suburban shoppers are more price sensitive.

[14]This finding is not new to our paper and has been pointed out before in the marketing literature.

[15]The problem here is of price endogeneity. Prices are (negatively) correlated with promotional activities. If one does not control for promotional activity the price coefficient will be biased. Our control is relatively rough and in principle might not fully control for the full impact of promotional activity.

TABLE IV

FIRST STEP: BRAND CHOICE CONDITIONAL ON SIZE[a]

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
Price	−0.51	−1.06	−0.49	−0.26	−0.27	−0.38	−0.38	−0.57	−1.41	−0.75
	(0.022)	(0.038)	(0.043)	(0.050)	(0.052)	(0.055)	(0.056)	(0.085)	(0.092)	(0.098)
*Suburban dummy				−0.33	−0.30	−0.34	−0.33	−0.25	−0.45	−0.19
				(0.055)	(0.061)	(0.055)	(0.056)	(0.113)	(0.127)	(0.127)
*Nonwhite dummy				−0.34	−0.39	−0.38	−0.33	−0.34	−0.33	−0.26
				(0.075)	(0.083)	(0.076)	(0.076)	(0.152)	(0.166)	(0.168)
Large family				−0.23	−0.13	−0.21	−0.22	−0.46	−0.38	−0.43
				(0.080)	(0.107)	(0.080)	(0.082)	(0.181)	(0.192)	(0.195)
Feature			1.06	1.05	1.08	0.92	0.93	1.08		1.05
			(0.095)	(0.096)	(0.097)	(0.099)	(0.100)	(0.123)		(0.126)
Display			1.19	1.17	1.20	1.14	1.15	1.55		1.52
			(0.069)	(0.070)	(0.071)	(0.071)	(0.072)	(0.093)		(0.093)
Brand dummy variable		✓	✓	✓	✓					
*Demographics					✓					
*Size						✓				
Brand–size dummy variable							✓			
Brand-HH dummy variable								✓		
*Size									✓	✓

[a] Estimates of a conditional logit model. An observation is a purchase instance by a household. Options include only products of the same size as the product actually purchased. Asymptotic standard errors are shown in parentheses.

To allow for heterogeneity in brand preferences, we interact the brand dummy variables with the demographic variables in column (v). Together the demographics variables generate eight different "types" of households. In columns (viii)–(x) we allow for household–brand specific effects. This is the richest time invariant form of heterogeneity in brand preference possible in these data.

Third, we see the importance of allowing brand preferences to vary by size in columns (vi), (vii), (ix), and (x), where we interact the brand–dummy variables with size (either by multiplying the dummy by the size (in columns (vi), (ix), and (x)) or by allowing a full interaction). Notice that interacting brand effects with sizes makes the preference for each specific product proportional to the container size purchased. Namely, if a consumer prefers Tide, then it is reasonable to increase or rescale her preference by the size of the container she is purchasing.

Table V reports the estimates of the price process. This process was estimated using the inclusive values (given in equation (5)) computed from the estimates of column (x) in Table IV. The inclusive values can be considered a quality weighted price: for each household, all the different prices (and promotional activities) of all the products offered in each size are combined into

TABLE V

SECOND STEP: ESTIMATES OF THE PRICE PROCESS[a]

	Same Process for All Types				Different Process for Each Type			
	ω_{2t}	ω_{4t}	ω_{2t}	ω_{4t}	ω_{2t}	ω_{4t}	ω_{2t}	ω_{4t}
$\omega_{1,t-1}$	0.003	−0.014	0.005	0.014	−0.023	−0.005	−0.019	0.007
	(0.012)	(0.011)	(0.014)	(0.014)	(0.017)	(0.014)	(0.019)	(0.015)
$\omega_{2,t-1}$	0.413	0.033	0.295	0.025	0.575	−0.003	0.520	0.011
	(0.007)	(0.010)	(0.008)	(0.007)	(0.013)	(0.010)	(0.016)	(0.013)
$\omega_{3,t-1}$	0.003	−0.034	0.041	−0.006	0.027	−0.072	0.051	−0.018
	(0.007)	(0.007)	(0.009)	(0.009)	(0.020)	(0.016)	(0.025)	(0.020)
$\omega_{4,t-1}$	0.029	0.249	0.026	0.236	−0.018	0.336	−0.018	0.274
	(0.008)	(0.008)	(0.008)	(0.017)	(0.020)	(0.016)	(0.021)	(0.017)
$\sum_{\tau=2}^{5} \omega_{1,t-\tau}$			−0.003	−0.012			−0.008	−0.003
			(0.005)	(0.004)			(0.006)	(0.005)
$\sum_{\tau=2}^{5} \omega_{2,t-\tau}$			0.089	0.006			0.073	−0.004
			(0.003)	(0.002)			(0.005)	(0.004)
$\sum_{\tau=2}^{5} \omega_{3,t-\tau}$			−0.008	−0.009			−0.004	−0.016
			(0.003)	(0.003)			(0.008)	(0.006)
$\sum_{\tau=2}^{5} \omega_{4,t-\tau}$			−0.013	0.018			−0.008	0.056
			(0.003)	(0.003)			(0.007)	(0.005)

[a]Each column represents the regression of the inclusive value for a size (32, 64, 96, and 128 ounces, respectively) on lagged values of all sizes. The inclusive values were computed using the results in column (x) of Table IV. The four left columns impose the same process for each household type; the four right columns allow for a different process for each type. Reported results are only for households of type 3, that is, households in market 1 with large families. Results for other types are available from the authors.

a single index. This index varies by household, because the brand preferences are allowed to vary.

The inclusive value is a summary of all the variables that impact utility, in particular prices and advertising. In our model, the consumer realizes that advertising will impact her future choice and therefore she forms expectations regarding future advertising. It may seem awkward that expected advertising affects current behavior. Alternatively, we could assume that advertising does not have dynamic implications, i.e., we could compute an inclusive value without advertising. In such a case, the consumer is repeatedly surprised by the effect of advertising (namely, advertising affects the choice probability without the consumer foreseeing it will happen). Such specification implies an inconsistency between the consumers' expected and actual probability of purchase. We opted for the internally consistent version of advertising.

The first four columns of Table V display the estimated parameters for the two main sizes (64 and 128 oz.), restricting the process to be the same for all households. The last four columns allow a different process for different household types. The results are presented for households in the urban market with large families (type 3). Within each set, the first two columns display results for a first-order Markov process. The point estimates suggest that the lagged value of own size is the most important in predicting future prices. Letting the price process vary by household type changes the parameters of the price process, but seems to have little impact on the bottom line (the results displayed below). In principle, one could allow the process to differ by household, but that would mean solving a different dynamic problem for each individual. Given the shortcut, this might be feasible, but it would significantly increase the computational burden. Given the small impact on the final results in going from one process to six, we decided against this option.

Considering the supply side, there are good reasons to believe that prices will not follow a first-order process. To explore alternatives to the first-order Markov assumption, in the next set of columns we include the sum of five additional lags. We also estimated, but do not display, a specification that allows these five lags to enter with separate coefficients. These additional lags do not significantly improve the fit. Therefore, we stick to a first-order Markov process.

Finally, assuming the inclusive values are distributed normally might seem strange given that the product-specific prices tend to be at a modal nonsale price with occasional reductions. The inclusive values aggregate across several prices with such patterns and also promotional activity. Although the resulting variable is distributed in a bell-like shape, a formal test rejects the hypothesis that the inclusive values are distributed normal. In principle, this assumption can be relaxed, at a substantial computational cost.

Table VI reports third stage estimates. We allow for six different types of households that vary by market and family size. Each type is allowed different utility and storage cost parameters. We also include size fixed effects that are

TABLE VI

THIRD STEP: ESTIMATES OF DYNAMIC PARAMETERS[a]

Household Type:	1	2	3	4	5	6
	Urban Market			Suburban Market		
Household Size:	1–2	3–4	5+	1–2	3–4	5+
Cost of inventory						
Linear	9.24	6.49	21.96	4.24	4.13	11.75
	(0.01)	(0.02)	(0.09)	(0.01)	(0.17)	(5.3)
Quadratic	−3.82	1.80	−35.86	−8.20	−6.14	−0.73
	(29.8)	(1.77)	(0.19)	(0.03)	(1.69)	(1.53)
Utility from consumption	1.31	0.75	0.51	0.08	0.92	3.80
	(0.02)	(0.09)	(0.21)	(0.03)	(0.18)	(0.38)
Log likelihood	365.6	926.8	1,530.1	1,037.1	543.6	1,086.1

[a]Asymptotic standard errors are shown in parentheses. Also included are size fixed effects, which are allowed to vary by household type.

allowed to vary by type (not reported in the table). Most of the parameters are statistically significant at standard significance levels. Households that live in the suburban market (where houses are on average larger) have lower storage costs.

To get an idea of the economic magnitude, consider the following. If the beginning-of-period inventory is 30 oz. (the median reported below), then buying a 128 oz. bottle increases the storage cost, relative to buying a 64 oz. bottle, by roughly \$0.20–0.75, depending on the household type. As we can see from Table III, the typical savings from nonlinear pricing are roughly \$0.40, which implies that the high storage cost types would not benefit from buying the larger size, whereas the low storage cost types would.

For this sample the estimated median inventory held is 27–33 oz., depending on the type. There is more variation across the types at the higher end of the distribution. The mean weekly consumption is between 22 and 36 oz., for different types (with the 10th and 90th percentiles varying between 4 and 6, and 85 and 115, respectively). If we assume that the households had constant consumption equal to their total purchases divided by the number of weeks, we get very similar average consumption. Furthermore, we can create an inventory series by using the assumption of constant consumption and observed purchases. If we set the initial inventory for such a series so that the inventory will be nonnegative, then the mean inventory is essentially the same as the inventory simulated from the model.

5.2. *Fit and Robustness*

Ideally, to further test the fit of the model, we would compare the simulated consumption (and inventory) behavior to observed data. However, consumption and inventory are not observed. We focus instead on the model's

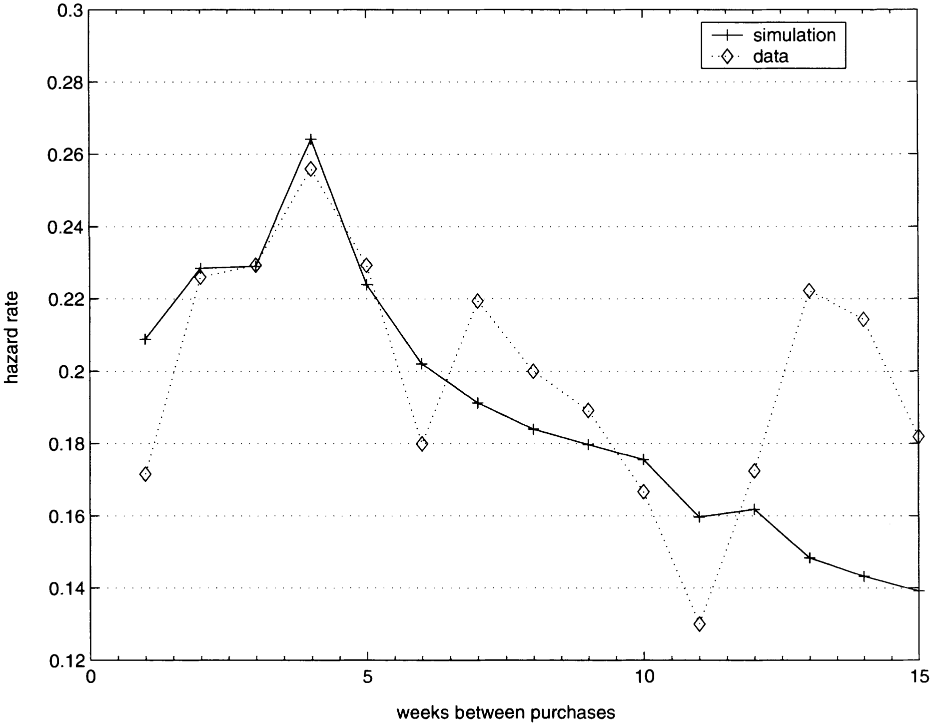

FIGURE 1.—Hazard rate of purchases.

prediction of interpurchase duration. Figure 1 displays the hazard rate of purchasing by duration week from last purchase. The hazard rate in this case is the probability that you purchase if you have not purchased up to now. If the probability of purchase is not duration dependent, than the hazard rate would be constant and equal to roughly 0.25 in our data. The data display a clear pattern: the probability of purchase gradually increases until week 4 and then gradually declines. The simulation from the model traces this pattern reasonably well. The predicted purchase probability one week after a purchase is higher than observed in the data. The model does a better job fitting the subsequent weeks—clearly better than a static model.

We tested the robustness of the results in several ways. First, we explored a variety of methods to solve the dynamic programming problem. In addition to the approximation method we used to generate the final set of results, we explored dividing the state space into a discrete grid. We then solved the dynamic programming problem over this grid and explored two ways to take this solution to the data. We divided the data into the same discrete grid and used the exact solution. We also used the exact solution on the grid to fit continuous value and policy functions, and used these to evaluate the data. We also explored a variety of functional forms for both the utility from consumption and the cost of inventory.

The estimated parameters did vary across several of these robustness tests. Indeed, even in our preferred specification the likelihood was relatively flat, which suggests that some of the identification issues we discussed are relevant in this data set. However, in the cases we examined, the implications discussed next were robust.

5.3. *Implications*

In this section we present the implications of the estimates, and compare them to static ones. In Table VII we present own- and cross-price long-run elasticities simulated from the dynamic model. The elasticities were simulated as follows. Using the observed prices, we simulated choice probabilities. We then generated the increased prices by adding a small amount to the observed price paths of each of the different products (brand and size). To evaluate long-run price responses, we generated permanent changes in the process (i.e., a change in the whole path of prices, not just the current price). We then reestimated the price process (although in practice the prices changes were small enough that the change in the price process was negligible) and solved for the optimal behavior given the new price process. Finally, we simulated new choice probabilities, used them to compute the change in choice probabilities relative to the initial values, and computed the price elasticities.

The results are presented in Table VII. Cell entries i and j, where i indexes row and j indexes column, present the percent change in market share of brand i with a 1 percent change in the price of j. All columns are for a size of 128 oz., the most popular size. The own-price elasticities are between -2.5 and -4.5. The cross-price elasticities also seem reasonable.[16] There are several patterns worth pointing out. First, cross-price elasticities to other brands of the same size, 128 oz., are generally higher. This is driven by the type-specific size fixed effects. Therefore, if the price of a product changes, consumers currently purchasing it are more likely to substitute toward similar size containers of a different brand.

Second, the cross-price elasticities to other sizes of the same brand are generally higher, sometimes over 20 times higher, than the cross-price elasticities to other brands. This suggests that these brands have a relatively loyal base. This pattern is driven by the heterogeneity estimated in the first step.

Third, the cross-price elasticities to the outside option, i.e., no purchase, are generally low. This is quite reasonable because we are looking at long-run responses, which reflect purely a consumption effect, presumably small for detergents. The substitution from say 128 oz. Tide to the outside option represents

[16]At first glance the cross-price elasticities might seem low. However, there is a large number of products: different brands in different sizes. If we were to look at cross-price elasticities across brands (regardless of the size), the numbers would be higher, roughly four times higher.

TABLE VII

LONG-RUN OWN- AND CROSS-PRICE ELASTICITIES[a]

Brand	Size (oz.)	All[b]	Wisk	Surf	Cheer	Tide	Private Label
All[b]	32	0.418	0.129	0.041	0.053	0.131	0.000
	64	0.482	0.093	0.052	0.033	0.085	0.006
	96	0.725	0.092	0.036	0.035	0.100	0.002
	128	−2.536	0.154	0.088	0.059	0.115	0.007
Wisk	32	0.088	0.702	0.046	0.012	0.143	0.006
	64	0.078	0.620	0.045	0.014	0.116	0.004
	96	0.066	0.725	0.051	0.022	0.135	0.009
	128	0.126	−2.916	0.083	0.026	0.147	0.005
Surf	32	0.047	0.061	0.977	0.024	0.369	0.003
	64	0.146	0.086	0.905	0.023	0.158	0.005
	96	0.160	0.101	0.915	0.016	0.214	0.001
	128	0.202	0.149	−3.447	0.039	0.229	0.008
Cheer	64	0.168	0.049	0.027	0.831	0.293	0.001
	96	0.167	0.015	0.008	0.982	0.470	0.001
	128	0.250	0.090	0.058	−4.341	0.456	0.003
Tide	32	0.071	0.085	0.050	0.022	1.007	0.002
	64	0.048	0.055	0.024	0.025	0.924	0.001
	96	0.045	0.063	0.016	0.026	1.086	0.001
	128	0.072	0.093	0.039	0.045	−2.683	0.001
Solo	64	0.066	0.070	0.027	0.021	0.150	0.002
	96	0.219	0.032	0.023	0.033	0.075	0.000
	128	0.127	0.125	0.060	0.043	0.302	0.001
Era	32	0.035	0.155	0.039	0.022	0.425	0.000
	64	0.030	0.103	0.039	0.018	0.304	0.008
	96	0.035	0.168	0.033	0.027	0.352	0.001
	128	0.054	0.192	0.061	0.029	0.513	0.014
Private	64	0.123	0.119	0.066	0.039	0.081	0.248
label	128	0.174	0.266	0.100	0.019	0.072	−2.682
No purchase		0.007	0.002	0.004	0.000	0.013	0.000

[a]Cell entries i and j, where i indexes row and j indexes column, give the percent change in market share of brand i with a 1 percent change in the price of j. All columns are for a 128 oz. product, the most popular size. The results are based on Tables IV–VI.

[b]Note that "All" is the name of a detergent produced by Unilever.

forgone purchases due to the higher price level. Namely, the proportion of purchases that, due to the permanent increase in the price of 128 oz. Tide (i.e., an increase in the support of the whole price process), lead to no purchase at all. In contrast, one would expect the response to a short-run price increase to be a lot larger. The reaction to a temporary price change includes not only the reduction in purchases, but also the change in the timing of the immediate purchase due to the temporary price change. As we discuss next, we indeed find that short-run estimates overestimate the substitution to the outside option. This highlights the bias from static estimates that our framework overcomes.

We next compare the long-run elasticities to elasticities computed from static demand models. There are two reasons why these will differ. First, the coefficients estimated in the static model are (upward) biased and inconsistent because of the omitted inventory and expectations. Even with the right controls, static estimation would reveal short-run responses. The dynamic model is needed to separate inventory responses to short-run price changes from consumption changes and brand substitutions in response to long-run price changes.

Table VIII presents the ratio of the static estimates to the dynamic estimates. Cell entries i and j, where i indexes row and j indexes column, give the ratio of the (short-run) elasticities computed from a static model divided by the long-run elasticities computed from the dynamic model. The elasticities for both models are the percent change in market share of brand i with a 1 percent change in the price of j. The static model is identical to the model estimated in the first step except that brands of all sizes are included as well as a no-purchase decision, not just products of the same size as the chosen option. The estimates from the dynamic model are based on the results presented in Tables IV–VI. The elasticities are evaluated at each of the observed data points, the ratio is taken and then averaged over the observations.

The results suggest that the static own-price elasticities overestimate the dynamic ones by roughly 30 percent. Part of this difference is driven by the bias in the estimates of the static model. The price coefficient estimated in the static model is roughly 15 percent higher than that estimated in the first stage of the dynamic model. This ratio varies slightly across brands and also across the main sizes, 64 and 128 oz.

In contrast, the static cross-price elasticities, with the exception of the no-purchase option, are smaller than the long-run elasticities. The effect on the no-purchase option is expected because the static model fails to account for the effect of inventory. A short-run price increase is most likely to chase away consumers who can wait for a better price, namely those with high inventories. Therefore, the static model will overestimate the substitution to the no-purchase option.

There are several effects that impact the cross-price elasticities to the other brands. As we previously noted, the coefficients estimated in the static model will tend to be upward biased. This suggests that the elasticities estimated from the dynamic model will be larger in magnitude. However, there is an additional effect due to the difference between long-run and short-run effects. Consider a reduction in the price of Tide. The static elasticities are computed from data that had temporary price reductions. Switchers are those households that are willing to switch, for example, from Cheer *and* have a low enough inventory at the time of the price change. In contrast, the long-run elasticities capture those households that are willing to substitute at all relevant levels of inventory, because they represent reactions to a permanent change in the price of Tide.

In the case of the own-price effect as well as in the case of cross-price effects toward the no-purchase option, the econometric bias and the difference

TABLE VIII

AVERAGE RATIOS OF ELASTICITIES COMPUTED FROM A STATIC MODEL TO LONG-RUN ELASTICITIES COMPUTED FROM THE DYNAMIC MODEL[a]

		64 oz.						128 oz.					
Brand	Size (oz.)	All[b]	Wisk	Surf	Cheer	Tide	Private Label	All[b]	Wisk	Surf	Cheer	Tide	Private Label
All[b]	64	1.03	0.13	0.14	0.12	0.13	0.15	0.14	0.17	0.17	0.18	0.21	0.34
	128	0.17	0.24	0.26	0.20	0.28	0.35	1.23	0.09	0.11	0.09	0.15	0.22
Wisk	64	0.14	1.20	0.13	0.17	0.12	0.13	0.16	0.22	0.14	0.22	0.25	0.20
	128	0.25	0.27	0.23	0.31	0.26	0.28	0.08	1.42	0.08	0.13	0.18	0.11
Surf	64	0.14	0.13	0.93	0.16	0.13	0.14	0.18	0.18	0.12	0.18	0.22	0.28
	128	0.25	0.22	0.18	0.27	0.25	0.18	0.12	0.11	1.20	0.08	0.15	0.14
Cheer	64	0.12	0.17	0.16	0.84	0.09	0.13	0.14	0.24	0.16	0.14	0.22	0.24
	128	0.25	0.26	0.26	0.12	0.23	0.22	0.09	0.12	0.06	0.89	0.15	0.07
Tide	64	0.16	0.17	0.13	0.13	1.26	0.15	0.22	0.28	0.16	0.26	0.22	0.37
	128	0.25	0.31	0.22	0.24	0.22	0.31	0.11	0.16	0.08	0.13	1.44	0.31
Solo	64	0.15	0.12	0.15	0.14	0.12	0.14	0.17	0.15	0.15	0.30	0.30	0.28
	128	0.23	0.20	0.24	0.21	0.21	0.25	0.07	0.07	0.06	0.16	0.17	0.21
Era	64	0.21	0.12	0.13	0.13	0.10	0.19	0.43	0.17	0.15	0.22	0.19	0.35
	128	0.31	0.22	0.24	0.25	0.17	0.38	0.19	0.08	0.09	0.11	0.10	0.22
Private	64	0.19	0.15	0.14	0.17	0.17	1.02	0.32	0.22	0.15	0.26	0.31	0.25
label	128	0.29	0.28	0.34	0.30	0.39	0.29	0.16	0.12	0.13	0.10	0.27	1.29
No purchase		2.12	1.13	1.15	1.40	1.27	2.39	1.80	7.60	2.26	14.11	2.38	10.86

[a]Cell entries i and j, where i indexes row and j indexes column, give the ratio of the (short-run) elasticities computed from a static model divided by the long-run elasticities computed from the dynamic model. The elasticities for both models are the percent change in market share of brand i with a 1 percent change in the price of j. The static model is identical to the model estimated in the first step, except that brands of all sizes are included as well as a no-purchase decision, not just products of the same size as the chosen option. The results from the dynamic model are based on the results presented in Tables IV–VI.

[b]Note that "All" is the name of a detergent produced by Unilever.

between short- and long- run effects operate in the same direction. Both overstate price responses. However, it is unclear which one of the effects (that work in opposite directions) dominates in the case of cross-price elasticities (toward other goods). It depends on the relative size of the two effects and whether the observed price variation was temporary or more permanent in nature. In our data, the latter effect dominates.

One might wonder about the representativeness of the results from our sample. The sample may not be representative of the whole population because it records detergent purchases in supermarkets only. Moreover, the sample we selected in the estimation may not reflect the whole population that purchases in supermarkets. The answer depends on the question one wishes to answer. We care about the ratio of elasticities for the typical consumer. We are likely underestimating the importance of consumer inventory because many of the consumers who shop elsewhere (e.g., at Walmart or Costco) will not be in our sample; these are the consumers more likely to buy for inventory. Alternatively, we might care about the typical shopper at the supermarkets in the sample. In other words, what is the long-run elasticity of the demand faced by the supermarkets. To evaluate how representative our selected sample is (of the whole population buying at these supermarkets), we estimated a static demand system using the aggregate data (namely, using total weekly sales at the store level, not just from the households in our sample). We found results similar to the static demand estimated using the household sample. This leads us to conclude that the results from our sample are reasonably representative.

Estimates of the demand elasticities are often used in one of two ways. They are used in a first-order condition, typically from a Bertrand pricing game, to compute price–cost margins (PCM). For single product firms, it is straightforward to see the magnitude of the bias: it is the same as the ratio of own-price elasticities. Therefore, the numbers in Table VIII suggest that for single product firms, the PCM computed from the dynamic estimates will be roughly 30 percent higher than those computed from static estimates. The bias is even larger for multiproduct firms because the dynamic model finds that the products are closer substitutes (and therefore a multiproduct firm would want to raise their prices even further) than the static estimates suggest.

PCM computed in this way are used to test among different supply models; in particular they are used to test for tacit collusion in prices (e.g., Bresnahan (1987) or Nevo (2001)). Static estimates will tend to find evidence of collusion where there is none, because the PCM predicted by models without collusion will seem too low.

Another use of demand estimates is to simulate the effects of mergers (e.g., Hausman, Leonard, and Zona (1994) and Nevo (2000)). The numbers in Table VIII suggest that estimates from a static model would tend to underestimate the effects of a merger, because they tend to underestimate the substitution among products. Furthermore, because the static estimates overestimate the substitution to the outside good if used to define the market, then they will tend to define it larger than a definition based on the dynamic estimates. In either case, the static estimates will favor approval of mergers.

6. CONCLUDING REMARKS

In this paper we structurally estimate a model of household inventory holding, allowing for product differentiation, sales, advertising, and nonlinear prices. The estimates suggest that ignoring the dynamics dictated by the ability to time purchases can lead to biased demand estimates. We find that static estimates overestimate own-price elasticities, underestimate cross-price responses to other products, and overestimate the substitution toward no purchase.

An implication of the model is that the likelihood of the observed choices can be split between a dynamic and a static component. We estimate the latter independently at little computational cost. The dynamic component requires the usual computation burden (of numerically solving the dynamic programming and numerically searching for the parameters that maximize the likelihood). However, the computational burden is substantially reduced by the split of the likelihood: we solve a simplified problem that involves only a quantity choice.

This split of the likelihood suggests a simple shortcut that can be used to reduce the biases that arise in a static estimation. The shortcut involves estimating demand conditional on the actual quantity purchased. In our case, the shortcut eliminates most of the short-run biases.

The estimates suggest differences between the long-run price effects computed from the dynamic model and a static one. Our static estimation used high frequency weekly consumer-level data. We also examined aggregate weekly data and got similar results. A common approach in the literature is to use monthly or quarterly aggregate data with the hope that aggregation over time smooths away the effects of temporary price reductions. An open question is how the results from static estimation using less frequent data compare to the dynamic results. A somewhat related question is whether there is a simple way to adjust the static estimation through time aggregation, much like the shortcut discussed in the previous paragraph, to attenuate the biases.

This paper neglected the supply side of the market. Future work could use estimates of demand to simulate profits from different pricing policies to evaluate, for instance, the profitability of observed sales. One of the key difficulties is setting up the agent's problem in a tractable way. A compact and tractable representation of the aggregate demand, as discussed in the previous paragraph, will be helpful.

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A., and NBER; igal@northwestern.edu

and

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A., and NBER; nevo@northwestern.edu.

Manuscript received March, 2005; final revision received May, 2006.

APPENDIX

In this appendix we provide proofs to some of the claims made in the text. First, we show that, conditional on size purchased, optimal consumption is the same regardless of which brand is purchased. Let $c_k^*(x_t, \nu_t)$ be the optimal consumption conditional on a realization of ν_t and purchase of size x_t of brand k.

LEMMA 1: $c_j^*(x_t, \nu_t) = c_k^*(x_t, \nu_t)$.

PROOF: Suppose there exists j and k such that $c_j^* = c_j^*(x_t, \nu_t) \neq c_k^*(x_t, \nu_t) = c_k^*$. Then

$$\begin{aligned}
&\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \varepsilon_{jxt} + u(c_j^* + \nu_t) - C(i_t + x_t - c_j^*) \\
&\quad + \delta E(V(s_t)|d_{jt} = 1, x_t, c_j^*) \\
&> \alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \varepsilon_{jxt} + u(c_k^* + \nu_t) - C(i_t + x_t - c_k^*) \\
&\quad + \delta E(V(s_t)|d_{jt} = 1, x_t, c_k^*)
\end{aligned}$$

and, therefore,

$$\begin{aligned}
&u(c_j^* + \nu_t) - u(c_k^* + \nu_t) \\
&> \delta E(V(s_t)|d_{jt} = 1, x_t, c_k^*) - \delta E(V(s_t)|d_{jt} = 1, x_t, c_j^*) \\
&\quad + C(i_t + x_t - c_k^*) - C(i_t + x_t - c_j^*).
\end{aligned}$$

Similarly, from the definition of $c_k^*(x_t, \nu_t)$,

$$\begin{aligned}
&u(c_j^* + \nu_t) - u(c_k^* + \nu_t) \\
&< \delta E(V(s_t)|d_{jt} = 1, x_t, c_k^*) - \delta E(V(s_t)|d_{jt} = 1, x_t, c_j^*) \\
&\quad + C(i_t + x_t - c_k^*) - C(i_t + x_t - c_j^*),
\end{aligned}$$

which is a contradiction. *Q.E.D.*

PROOF OF CLAIM 1: The dynamic problem defined in equation (1) has an associated Bellman equation

$$V(s_t) = \underset{\{c, d_{jx}\}}{\mathrm{Max}} \left\{ u(c + \nu_t) - C(i_{t+1}) + \sum_{j,x} d_{jx}(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \varepsilon_{jxt}) + \delta E[V(s_{t+1})|s_t, c, d_{jx}] \right\}.$$

From Assumptions 1 and 3, both ε and ν are independently and identically distributed. Moreover, current actions affect future utility only through end-of-period inventory, so we can write $E[V(s_{t+1})|i_t, p_t, c, d_{jx}]$. In other words, the expectation of V given the state and current behavior is a function of current price and end-of-period inventories. Let us denote such a function as $V^e(i_{t+1}, p_t)$.

Taking expectations of $V(s_t)$ given the information available at $t-1$ (which includes actions taken at $t-1$), we can find $V^e(i_t, p_{t-1})$. Using the independence of ε, ν, and p, we get

$$\begin{aligned} &V^e(i_t, p_{t-1}) \\ &\quad = \int \Bigg[\operatorname*{Max}_{\{c, d_{jx}\}} \Bigg\{ u(c+\nu_t) - C(i_{t+1}) + \sum_{j,x} d_{jx}(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \varepsilon_{jxt}) \\ &\qquad\qquad + \delta V^e(i_t + x - c, p_t) \Bigg\} \Bigg] dF(\varepsilon)\, dF(\nu)\, dF(p_t|p_{t-1}). \end{aligned}$$

By Lemma 1, optimal consumption depends on the quantity purchased, but not on the brand chosen; then $\text{Max}_c\{u(c+\nu_t) - C(i_{t+1}) + \delta V^e(i_t + x - c, p_t)\}$ varies by size x, but is independent of choice of brand j. Denote this function $M(s_t, x)$. Therefore,

$$\begin{aligned} V^e(i_t, p_{t-1}) = \int \Bigg[\operatorname*{Max}_{d_{xj}} \Bigg(&\sum_{j,x} d_{jx}(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \varepsilon_{jxt}) \\ &+ M(s_t, x) \Bigg) \Bigg] dF(\varepsilon)\, dF(\nu)\, dF(p_t|p_{t-1}). \end{aligned}$$

For the case of extreme value shocks (see McFadden (1981)) and using the definition of the inclusive values given in equation (5), this last expression can be written as

$$V^e(i_t, p_{t-1}) = \int \log\Bigg(\sum_x \exp(\omega_{xt} + M(s_t, x)) \Bigg) dF(\nu)\, dF(p_t|p_{t-1}).$$

Using Assumption 4, V^e can be written as a function of ω_t and i_t instead of p_t and i_t, leading to (spelling out the function M to remind the reader)

$$\begin{aligned} V^e(i_t, \omega_{t-1}) = \int \log\Bigg(&\sum_x \exp\big(\omega_{xt} + \operatorname*{Max}_c \{u(c+\nu_t) - C(i_{t+1}) \\ &+ \delta V^e(i_{t+1}, \omega_t)\}\big) \Bigg) dF(\nu)\, dF(\omega_t|\omega_{t-1}). \end{aligned}$$

The former functional equation can be used to find $V^e(i_t, \omega_{t-1})$.

Notice that $V^e(i_t, \omega_{t-1})$ also characterizes the problem

$$V(i_t, \omega_{t-1}, \varepsilon_t, \nu_t) \\ = \operatorname*{Max}_{\{c,x\}}\{u(c+\nu_t) - C(i_{t+1}) + \omega_{xt} + \varepsilon_{xt} + \delta V^e(i_{t+1}, \omega_t)\}.$$

This can be seen by taking expectations with respect to ε on the right-hand side. Notice that this last Bellman equation characterizes the optimal behavior of a consumer who chooses x (quantity purchased) in every period, enjoying a shock $\varepsilon_{xt} + w_{xt}$ from buying quantity x in period t.

Finally, we want to show that we can rely on $V^e(i_t, \omega_{t-1})$ to derive the likelihood of purchase of quantity x. We want to show that

$$\Pr(x_t|i_t, p_t, \nu_t) = \Pr(x_t|i_t, \omega(p_t), \nu_t).$$

Notice that

$$\Pr(x_t|i_t, \omega_t, \nu_t) \\ = \frac{\exp(\omega_{xt} + \mathrm{Max}_{c_t}\{u(c_t+\nu_t) - C(i_{t+1}) + \delta V^e(i_{t+1}, \omega_t)\})}{\sum_x \exp(\omega_{xt} + \mathrm{Max}_{c_t}\{u(c_t+\nu_t) - C(i_{t+1}) + \delta V^e(i_{t+1}, \omega_t)\})},$$

whereas

$$\begin{aligned}
&\Pr(x_t|p_t, i_t, \nu_t) \\
&= \frac{\sum_j \exp(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + M(s_t, x))}{\sum_j \exp(\alpha p_{jyt} + \xi_{jy} + \beta a_{jyt} + M(s_t, x))} \\
&= \frac{\exp(M(s_t, x)) \exp(\log(\sum_j \exp(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt})))}{\sum_y \exp(M(s_t, y)) \exp(\log(\sum_j \exp(\alpha p_{jyt} + \xi_{jy} + \beta a_{jyt})))} \\
&= \frac{\exp(\omega_{xt} + M(s_t, x))}{\sum_y \exp(\omega_{yt} + M(s_t, y))} \\
&= \Pr(x_t|\omega_t, i_t, \nu_t),
\end{aligned}$$

where the last equality follows from the fact the M depends on ω but not on p. *Q.E.D.*

REFERENCES

AGUIRREGABIRIA, V. (1999): "The Dynamics of Markups and Inventories in Retailing Firms," *The Review of Economic Studies*, 66, 275–308. [1639]

——— (2005): "Nonparametric Identification of Behavioral Responses to Counterfactual Policy Interventions in Dynamic Discrete Decision Processes," *Economics Letters*, 87, 393–398. [1653]

AILAWADI, K. L., AND S. NESLIN (1998): "The Effect of Promotion on Consumption: Buying More and Consuming It Faster," *Journal of Marketing Research*, 35, 390–398. [1640,1644]
BELL, D., AND Y. BOZTUG (2004): "The Effect of Inventory on Purchase Incidence: Empirical Analysis of Opposing Forces of Storage and Consumption," Mimeo, Wharton School, University of Pennsylvania. [1640,1644]
BELL, D., J. CHIANG, AND V. PADMANABHAN (1999): "The Decomposition of Promotional Response: An Empirical Generalization," *Marketing Science*, 18, 504–526. [1640]
BENITEZ-SILVA, H., G. HALL, G. HITSCH, G. PAULETTO, AND J. RUST (2000): "A Comparison of Discrete and Parametric Approximation Methods for Continuous-State Dynamic Programming Problems," Mimeo, Yale University. [1652]
BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. [1637]
BERTSEKAS, D., AND J. TSITSIKLIS (1996): *Neuro-Dynamic Programming*. Cambridge, MA: Athena Scientific. [1652]
BLATTBERG, R., G. EPPEN, AND J. LIEBERMAN (1981): "A Theoretical and Empirical Evaluation of Price Deals in Consumer Nondurables," *Journal of Marketing*, 45, 116–129. [1639,1640]
BOIZOT, C., J.-M. ROBIN, AND M. VISSER (2001): "The Demand for Food Products. An Analysis of Interpurchase Times and Purchased Quantities," *Economic Journal*, 111, 391–419. [1639]
BRESNAHAN, T. (1987): "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War," *Journal of Industrial Economics*, 35, 457–482. [1637,1667]
CHIANG, J. (1991): "A Simultaneous Approach to the Whether, What and How Much to Buy Questions," *Marketing Science*, 10, 297–315. [1640]
CHAMBERLAIN, G. (1983): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches and M. Intrilligator. Amsterdam: North-Holland. [1655]
CURRIM, I., AND L. SCHNEIDER (1991): "A Taxonomy of Consumer Purchase Strategies in a Promotion Intense Environment," *Marketing Science*, 10, 91–110. [1639,1640]
ERDEM, T., S. IMAI, AND M. KEANE (2003): "Consumer Price and Promotion Expectations: Capturing Consumer Brand and Quantity Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5–64. [1638,1639,1655,1657]
GUPTA, S. (1988): "Impact of Sales Promotions on when, What, and How Much to Buy," *Journal of Marketing Research*, 24, 342–355. [1640]
——— (1991): "Stochastic Models of Interpurchase Time with Time-Dependent Covariates," *Journal of Marketing Research*, 28, 1–15. [1640]
HAUSMAN, J., G. LEONARD, AND J. D. ZONA (1994): "Competitive Analysis with Differentiated Products," *Annales d'Economie et de Statistique*, 34, 159–180. [1637,1667]
HENDEL, I., AND A. NEVO (2006): "Sales and Consumer Stockpiling," *The RAND Journal of Economics*, forthcoming. [1643,1644]
HOSKEN, D., D. MATSA, AND D. REIFFEN (2000): "How Do Retailers Adjust Prices: Evidence from Store-Level Data," Working Paper 230, Federal Trade Commission Bureau of Economics. [1639]
LUCAS, R. E., AND L. A. RAPPING (1969): "Real Wages Employment, and Inflation," *The Journal of Political Economy*, 77, 721–754. [1637]
MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816. [1653]
MCFADDEN, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden. Cambridge, MA: MIT Press, 198–272. [1651, 1670]
MELNIKOV, O. (2001): "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," Mimeo, Cornell University. [1649]
MURPHY, K., AND R. TOPEL (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business & Economics Statistics*, 3, 370–379. [1656]
NESLIN, S., C. HENDERSON, AND J. QUELCH (1985): "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science*, 4, 147–165. [1639,1640]

NESLIN, S., AND L. SCHNEIDER STONE (1996): "Consumer Inventory Sensitivity and the Postpromotion Dip," *Marketing Letters*, 7, 77–94. [1640]

NEVO, A. (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 31, 395–421. [1667]

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342. [1667]

PESENDORFER, M. (2002): "Retail Sales. A Study of Pricing Behavior in Supermarkets," *Journal of Business*, 75, 33–66. [1639]

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033. [1638,1647]

——— (1996): "Numerical Dynamic Programming in Economics," in *Handbook of Computational Economics*, Vol. 1, ed by. H. Amman, D. Kendrick, and J. Rust. Amsterdam: North-Holland, 619–729. [1653]

SHOEMAKER, R. W. (1979): "An Analysis of Consumer Reactions of Product Promotions," in *Educators' Conference Proceedings*. Chicago: American Marketing Association, 244–248. [1639]

TRAIN, K. (1986): *Qualitative Choice Analysis: Theory, Econometrics and Application to Automobile Demand*. Cambridge, MA: MIT Press. [1649]