NBER WORKING PAPER SERIES

MEASURING THE IMPLICATIONS OF SALES AND CONSUMER INVENTORY BEHAVIOR

Igal Hendel
Aviv Nevo

Working Paper 11307
http://www.nber.org/papers/w11307

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2005

We wish to thank David Bell for the data and Michael Keane, Ariel Pakes, John Rust and seminar participants in several workshops for comments and suggestions on earlier versions of this work. The second author wishes to thank the Center for the Study of Industrial Organization at Northwestern University for hospitality and support and the Sloan Foundation for support through a Sloan Research Fellowship. We gratefully acknowledge support from the NSF (SES-0093967 and SES-0213976). The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring the Implications of Sales and Consumer Inventory Behavior
Igal Hendel and Aviv Nevo
NBER Working Paper No. 11307
April 2005
JEL No. G0

ABSTRACT

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 mis-measure 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 to other products by up to a factor of 5; and (iii) overestimate the substitution to the no purchase, or outside option, by over 200 percent.

Igal Hendel
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
igal@northwestern.edu

Aviv Nevo
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston IL 60208-2600
and NBER
nevo@northwestern.edu

1. Introduction

When goods are storable traditional static demand analysis is likely to mis-measure 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 effect to a permanent price reduction would be lower than static estimates suggest. Quantifying the impact of the intertemporal effect on demand estimation is important, for example, in antitrust analysis which is increasingly driven by empirical estimates of demand elasticities. The impact of inventories on cross price elasticities is less obvious, but at least as important for antitrust analysis. This paper presents a model of household behavior that captures the main features faced by the household: intertemporal price variation, which creates incentives to store, several brands, non-linear pricing and promotional activities like advertising and display. We structurally estimate the model in order to study the economic implications of storability on demand estimation.

Estimation of demand in industries with differentiated products is a central part of applied industrial organization. Recent papers in the academic literature have studied a variety of industries including automobiles, retail products and computers (Bresnahan, 1987; Hausman, Leonard and Zona, 1994; Berry, Levinsohn and Pakes, 1995; as well as many others). Most applications have neglected intertemporal effects that could result, for example, from the durability or storability of the product. The estimation is typically performed assuming that the demand for the product is independent of the history of prices and previous purchases. We propose a framework to incorporate the dynamics generated by product storability into the estimation of demand. Our goal is to compare the estimates obtained from a dynamic model to those achieved by standard static methods.

In most demand applications (e.g., merger analysis or computation of welfare gains from introduction of new goods) we want to measure responses to permanent price changes. Static demand estimation methods will mis-measure these responses for two reasons: First, by neglecting dynamics these model are mis-specified. They do not correctly control for the relevant history like past prices, and inventories. If current prices are correlated with past prices, and hence with current inventories, the omitted variables lead to biased estimates of price sensitivity. Second, even adding the right controls static estimation cannot separate between short-run and long run responses.

Storability also has implications for how sales should be treated in the consumer price index. If consumers stockpile, then ignoring the fact that they can substitute over time will yield a bias similar to the bias generated by ignoring substitution between goods as relative prices change (Feenstra and Shapiro, 2001). A final motivation to study stockpiling behavior, is to understand sellers' pricing incentives when products are storable.

Previous work in the Marketing and Economics literature documented buying patterns at the household and store level that are consistent with stockpiling. We review these papers below and in Section 2.3 we provide a brief summary of similar findings for the data we use here. Our goal in this paper is to measure the implications of this behavior. Towards this goal we present a model of household behavior. 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. Households purchase for current consumption and to build inventories.

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 sub-markets of a large mid-west city. In addition we follow the purchases of roughly 1,000 households who shopped in those stores over a period of 104 weeks. We know exactly which product was bought, where it was bought and how much was paid. We also know 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. In order to reduce the dimensionality, we separate the probability of choosing any brand-size combination into the probability of choosing a brand conditional on quantity, 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 consistently 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, allowing for a large degree of consumer heterogeneity and nests standard static choice models. We discuss below the assumptions necessary to validate this procedure, the limitations of the method and potential suitability to model demand for other products.

The estimates suggest that ignoring dynamics can have strong 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 underestimate 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. These findings imply that an analysis based on static elasticity estimates will underestimate price-cost margins and under predict 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. He shows that in his model the equilibrium 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 et al. (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 et al. (2000) study the probability of a product being on sale as a function of its attributes. They report that sales are more likely for more popular products and in periods of high demand. Warner and Barsky (1995), Chevalier, et al. (2003) and MacDonald

(2000) study the relation between seasonality and sales. Aguirregabiria (1999) studies retail inventory behavior and its effect on prices.

The closest paper to ours is Erdem, Imai and Keane (2003). They were the first to structurally estimate a consumer inventory model. They construct a structural 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 than ours, but allows the inclusion of random effects in product preferences. Our method, on the other hand, allows for more observed heterogeneity, and due to the computational simplicity can handle a larger choice set. Modeling and estimation differences render each method better suited for different applications. Their method would be difficult to apply in the industry we study. Erdem et al. study the role of price expectations and differences between short run and long run price responses. To evaluate the role of expectations, they compare consumers responses' to price cuts, both allowing for the price cut to affect future price expectations, and holding expectations fixed. Interestingly, they are able to separate the price and the expectation effect of a sale on demand. They also use the estimates to simulate consumer responses to short run and long run price changes. In contrast, our interest is in comparing long run elasticities to those obtained through standard static methods. We compare the models in more detail in Section 4.

There is a 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) find effects of the same magnitude as Blattberg, Eppen and Lieberman (1981). Neslin and Schneider Stone (1996) summarize that the estimates of accelerated purchase are between 14 and 50 percent.

The subsequent literature concentrated on splitting the price responses to sales. As Blattberg, Eppen and Lieberman (1981) point out "the effect of a deal will be greatly overstated" due to stockpiling. Gupta (1988) decomposes price responses into brand switching, acceleration of purchases and stockpiling. His approach is to estimate separate and independent choice models of brand, timing and quantity purchased, never linking the three related decisions in a single framework. In his coffee sample, a 85% of the price response corresponds to brand switching, 14% is due to demand acceleration and 2% stockpiling. Chiang (1991) reports similar findings; he abstracts from estimating the timing of purchase. The goal of Chiang (1991) is to distinguish between the primary demand expansion effect of promotions (created by expanding sales of the category) versus the secondary effect (created by brand switching). A more detailed and comprehensive study by Bell, Chiang and Padmanabhan (1999) shows that the primary demand expansion is larger than the previous studies reported. Ailawadi, Scott 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 and household-level data. The first was collected using scanning devices in nine supermarkets, belonging to different chains, in two separate sub-markets in a large mid-west city. From the aggregate data we know for each detailed product (brand-size) in each store in each week the price charged, (aggregate) quantity sold and promotional activities that took place. The second component of the data set is at the household-level. We observe the purchases of roughly 1,000 households over a period of 104 weeks. We know when a household visited a supermarket and how much they spent each visit. The data includes purchases in 24 different product categories for which we know exactly which product each household bought, where it was bought and how much was paid.

Table 1 displays statistics of some household demographics, laundry detergents 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 we observe purchases. Since the household-level brand HHI is roughly 0.5, purchases are concentrated at two main brands. However, since preferred brands differ by household the 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 2. The top 11 brands account for roughly 90 percent of the quantity sold.

Most brand-size combinations have a regular price. In our sample 71 percent of the weeks the price is at the modal level, and above it only approximately 5 percent of the time. Defining a sale as any price at least 5 percent below the model price of each UPC 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 40 cents, the average is 67 cents, the 25 percentile is 20 cents and the 75 percentile is 90 cents. In percentage terms the median discount is 8 percent, the average is 12 percent, and the 25 and 75 percentiles are 4 and 16 percent, respectively. As shown in Table 1, there is some variation across brands in the percent quantity sold on sale.

Detergents come in several sizes. However, about 97 percent of the volume of liquid

[2]This definition of a sale would not be 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. Furthermore, the definition of a sale only matters for the descriptive analysis in this section. We do not use it in the structural econometric analysis below.

detergent sold was sold in 5 different sizes.[3] Sizes of powder detergent are not quite as standardized. Prices are non-linear in size. Table 3 shows the price per 16 oz. unit for several container sizes. The figures are computed by averaging the per unit price in each store over weeks and brands. The numbers suggest a per unit discount for the largest sizes. The figures in Table 3 are averaged across different brands and therefore might be slightly misleading since not all brands are offered in all sizes or at all stores. We also examined the pricing patterns for specific brands and the same patterns emerged.

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

In Hendel and Nevo (2002) we found several patters consistent with stockpiling. 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 non-sale

---

[3]Towards the end of our 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. with 50 oz.

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

periods.[5] Both these effects are consistent with an inventory model since the longer the duration from the previous sale, on average, the lower the inventory each household currently has, making purchase more likely. Second, 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 non-sale purchases, in both duration from previous purchase and duration to next purchase. In order 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 the following per period utility

$$u(c_{ht}, \nu_{ht}; \theta_h) + \alpha_h m_{ht}$$

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, θ_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 stochastic shock, ν_{ht}, introduces randomness in the consumer's needs, unobserved by the researcher. For simplicity we assume the shock to utility is additive in consumption, $u(c_{ht}, \nu_{ht}; \theta_h) = u(c_{ht} + \nu_{ht}; \theta_h)$. High realizations of ν_{ht} decrease the household's need. The product is offered in *J* different varieties, or brands. Consumers face random and potentially non-linear prices.

Therefore, consumers in each period decide which brand to buy, how much to buy and how

[5] Pesendorfer (2002) also finds that duration from previous sale affects demand during sales.

much to consume.[6] Since the good is storable quantity not consumed is stored as inventory. 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 and not how many containers).[7] We denote a purchase of 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

$$V(s_0) = \max_{\{c_h(s_t), d_{hjx}(s_t)\}} \sum_{t=0}^{\infty} \delta^t E\left[u(c_{ht}, \nu_{ht}; \theta_h) - C_h(i_{h,t+1}) + \sum_j d_{hjxt}(\alpha_h p_{jxt} + \xi_{hjx} + \beta_h a_{jxt} + \epsilon_{hjxt}) | s_0\right]$$

$$s.t. \quad 0 \le i_{ht}, \quad 0 \le c_{ht}, \quad 0 \le x_{ht}, \quad \sum_{j,x} d_{hjxt} = 1 \qquad \textbf{(1)}$$

$$i_{h,t+1} = i_{ht} + x_{ht} - c_{ht}$$

where s_t denotes the state at time *t*, $\delta > 0$ is the discount factor, $C_h(i)$ is the cost of storage, ξ_{hjx} is an idiosyncratic taste for brand *j* that could be a function of brand characteristics, size and vary across consumers, $\beta_h a_{jxt}$ captures the effect of advertising variables on the consumer choice, and ϵ_{hjxt}, is a random shock to consumers' choice. Notice, the latter is size specific, namely, different sizes get different draws introducing randomness in the size choice as well.

Given the assumptions that follow, the state at time *t*, 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 ϵ's. All functions in equation (1) are allowed to vary by household (as we will see in the results section). In order to simplify notation we drop the subscript *h* in what follows. 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 over time.

[6]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 prices since this is the main alternative explanation to why consumers buy more during sales; we want to make sure that are results are not driven by assuming it away. Moreover, reduced form results in Ailawadi, Scott and Neslin (1998), Hendel and Nevo (2002) and Bell and Boztug (2004) report consumption effects for several products.

[7]In our data more than 97 percent of the purchases are for a single unit so multi-unit purchases is an important issue in this industry. In principle our model could allow for multiple purchases by expanding the choice set to allow for bundles.

Assumption A1: ν_t is independently distributed over time.

In principle serial correlation in ν_t can be allowed, but at a significant increases the computational burden since the expectation in equation (1) will also be taken conditional on ν_t (and potentially past shocks as well).

Assumption A2: 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 since 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 expectations. With our assumption a consumer sees the current prices at the store and computes expected future prices. A higher order process would require that she recall prices at previous visits. In the application we explore higher order processes, but for now we maintain the first order assumption.

Assumption A2 also assumes that the price process is exogenous, namely, conditional on the control variables the process is independent of the unobserved random components. Given that we have 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 not a concern for the product we study below, but if present, seasonality should be modeled into the price process.

Assumption A3: ϵ_{jxt} is independently and identically distributed extreme value type 1.

We will use this assumption below to significantly reduce the computational burden. This assumption can be relaxed. We can allow for correlation in the unobserved shocks between brands by assuming that the ϵ 's are distributed according to some distributions in the generalized extreme value class.[8] For simplicity we present what follows only for the more restricted i.i.d. 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 of capturing the expected value of the future differences in utility at the time of consumption. This approach is valid as long as (i) brand-specific differences in the utility from consumption enter linearly in the utility function,[9] and (ii) discounting is low. For example, suppose $U(c_1,...,c_J) = \sum_j \psi_j c_j$, where c_j is quantity consumed of brand j and ψ_j is a taste parameter (as in Erdem et al., 2003). The term ξ_{jx} captures the utility from the future consumption of the x units of brand j, expected at the time of purchase. With discounting the previous analogy becomes less straightforward,[10] but since the products we study have an inter-purchase cycle of weeks the role of discounting can be neglected, to a first approximation.

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 the marginal utility from consuming one brand depends on the consumption of another brand. For most products for which scanner data is available (like detergents, tuna, ketchup, sugar, coffee) and other storables like gasoline, consumption independence seems like a reasonable assumption. Notice that interactions

---

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

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

[10] Two issues arise. First, since 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.

between brands are ruled out in standard static discrete choice models.

4. Econometrics

The structural estimation is based on the nested algorithm proposed by Rust (1987), but has to deal with issues specific 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, since 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 a 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 above algorithm. First, we do not observe inventory since both the 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.[11] 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 v), 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 the next period observed purchase. Repeating this procedure we obtain the likelihood of observing the whole sequence of purchases for each household. In order to start this procedure we need a value for the initial inventory. We do so by starting at an arbitrary initial level, and using

[11] Alternatively, we could assume that weekly consumption is constant, for each household over time, and estimate it by the total purchase over the whole period divided by the total number of weeks. Results using this approach are presented in Hendel and Nevo (2002).

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, $(d_1,...,d_T)$ as a function of the observed state variables, $(p_1,....,p_T)$ is

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

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 the integration over the set of ϵ 's that induce d_t as the optimal choice and is given by

$$Pr(d_{jx}|p_t, i_t, \nu_t) = \frac{\exp\Big(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \underset{c}{Max}\big\{u(c+\nu_t) - C(i_{t+1}) + \delta E(V(s_{t+1})|\, d_{jx}, c, s_t)\big\}\Big)}{\sum_{k,y} \exp\Big(\alpha p_{kyt} + \xi_{ky} + \beta a_{kyt} + \underset{c}{Max}\big\{u(c+\nu_t) - C(i_{t+1}) + \delta E(V(s_{t+1})|\, d_{ky}, c, s_t)\big\}\Big)} \qquad \textbf{(3)}$$

where *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, 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 above 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's. As we show below, we do not need to solve the dynamic programming problem in order 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 discussed above 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 having 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 the 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.[12]

4.2 The Three Step Procedure

We now show that the break-up of the likelihood follows from the primitives of the model.

4.2.1 Step 1: Estimation of the "Static" Parameters

In order to simplify the computation of the likelihood we note that we can write the probability of choosing brand j and size x 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 us since we need to solve the consumer's dynamic programming problem in order to compute $Pr(d_{jt}=1|p_t,x_t,i_t,\nu_t)$. However, given the above model and

[12]We are aware of two instances in the literature that use similar ideas. First, one way to estimate a static nested logit model is to first estimate the choice within a nest, compute the inclusive value and then estimate the choice among nests using the inclusive values (Train, 1986). Second, in a dynamic context a similar idea was proposed independently 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.

assumptions we can compute this conditional probability without solving the dynamic programming problem. We rely on two results. First, the optimal consumption, conditional on the quantity purchased, is not brand specific. Thus, the term $\underset{c_t}{Max}\{u(c_t+\nu_t)-C(i_{t+1})+\delta EV(s_{t+1})|d_{jt}=1,x_t,c_t)\}$ is independent of brand choice. This in itself is not enough to deliver the result since the distribution of ϵ might be a function x. If this is the case we would have to fully solve the model in order to figure out the conditional distribution of the error. Therefore, we use Assumption A3 to show that after the appropriate cancellations in equation (3) we obtain

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

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 the brand fixed effects) by maximizing the product, over time and households, of $Pr(d_{jt}=1|p_t,x_t)$. This amounts to estimating a static brand choice model including in the choice set only the brands offered in the same size of the actual purchase.

4.2.2. Step 2: Inclusive Values

In order to compute the likelihood of a sequence of quantity purchases we show, in the next section, that we can simplify the state space of the dynamic programming problem. In order to do so, we compute the "inclusive values" for each size (quantity) and their transition probabilities from period to period. The inclusive value can be thought of as a quality adjusted price index for all brands of size x and is given by

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

All the information needed to compute the inclusive values and their transition probabilities is contained in the first stage estimates. Note that since the parameters might vary with consumer characteristics the inclusive values are consumer specific.

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. In order for this to work we make the following assumption regarding the transition probabilities, F, of the inclusive values.

Assumption A4: $F(\omega_t|s_{t-1})$ can be summarized by $F(\omega_t|\omega_{t-1})$.

This assumption says that the distribution of next period inclusive values depends only on the current inclusive values and not the whole vector of prices. So instead of keeping track of the prices of ten brands times four sizes (roughly the dimensions in our data), we only have to follow four quality adjusted prices.

The main loss in defining the process this way is that transition probabilities are a function of the inclusive values exclusively. Two price vectors that yield the same vector of inclusive values will have the same transition probabilities to next period state. A more general transition model will allow the transition to be different. 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.[13] A less computationally demanding fix would be to have the distribution depend on additional current information but not the full vector of prices. For instance, we can identify from the data groups (or categories) of current prices that lead to the same distribution of future inclusive values.

4.2.3 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 following simplified problem. In the simplified problem the consumer chooses whether to purchase, and what quantity to purchase. Her flow utility of

[13]Notice that if we assume the inclusive values depend on the whole vector of past prices only the third stage becomes more computational demanding. However, the split remains valid.

purchasing quantity x is given by $\omega_{xt}+\epsilon_{xt}$. Recall that ω_{xt} is the summary of utility of expected by the consumer in the original problem, before knowing her precise realizations of ϵ_{jxt}. The consumer in the simplified problem observes only a summary of the state variables (given by the inclusive values) and decides the quantity to purchase. The Bellman equation associated with the simplified problem is

$$V(i_t, \omega_t, \epsilon_t, \nu_t) = \underset{\{c,d_x\}}{Max}\left\{u(c+\nu_t) - C(i_{t+1}) + \sum_x d_x(\omega_{xt}+\epsilon_{xt}) + \delta E\left[V(i_{t+1}, \omega_{t+1}, \epsilon_{t+1}, \nu_{t+1}) | i_t, \omega_t, \epsilon_t, \nu_t, c, x\right]\right\}.$$

Using the estimated inclusive values, and the transition probabilities, we can solve the simplified problem. We apply the nested algorithm to compute the likelihood of purchasing a size (quantity). The following claim shows the equivalence of the likelihood computed from the two problem, thus, the validity of using the simpler problem for our estimation.

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

The are two steps to proving this claim (see the Appendix). The intuition is as follows. First note, that the solution to the original and the simplified problem are characterized by the same $E(V)$. The original problem is characterized by the following Bellman equation

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

The equivalence is not obvious, because in one case the consumer observes the whole state s_t when she decides what to purchase, while on the other case she picks a quantity, x, having observed only a summary of the utility from buying that quantity, namely ω_{xt}. The key to the equivalence is in Assumptions A3, which allows us to obtain a summary of the expected utility from all products of a particular size and in Assumption A4, enabling us to write the transitions probabilities as a function of only the inclusive values.

Once we have shown the equivalence of $E(V)$ we still need to show that we can use the solution to the reduced problem to derive the likelihood of purchase of quantity x in the original problem. This comes almost directly out of Assumption A3 (as shown in the Appendix).

Despite the reduction in the number of state variables, the value function is still computationally burdensome to solve. To solve it we use value function approximation with policy function iteration. We closely follow Benitez-Silva et al. (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 use least squares to 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. It 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 Betrsekas 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

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 and size effects functions of household demographics, as well as allowing for household specific brand effects.

The identification of the dynamic parameters, estimated in the third stage, 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 over the sample period. If inventory and consumption were observed then identification would follow the standard arguments (see Rust, 1996 and Magnac and Thesmar, 2002, 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 tells 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, we would then 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. Consider another 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, the pair: 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 costs levels determine inter-purchase duration and the extent to which consumers can exploit price reductions. Higher storage costs reduce consumers' ability to benefit from sales and make the average duration between purchases shorter. In the extreme case of no storage (or very high storage cost), inter-purchase duration depend exclusively on current prices, since 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 identified from the relation between purchases, prices and previous purchases. Indeed to evaluate the fit of the model we will compare the predictions of the model to the observed inter-purchase duration.

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 above procedure provides (i) an intuitive interpretation of the determinants of substitution patters and (ii) a shortcut to separate long run from short run prices responses. The basic insight is that in order to purge elasticities from timing effects – as a first approximation – one should 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 is 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 A4. Also critical for the analysis is Assumption A3 that assumes the error terms, ϵ_{jxt} are identically and independently distributed type I extreme value, both across choices and over time. As we mentioned above we can relax the independence across brands by allowing for a GEV distribution. Here we discuss ways to allow for heterogeneity that is persistent over time.

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 preferences must be uncorrelated over time. Suppose there are different (unobservable) types in the population that vary in their brand preferences, or in their sensitivity to price and advertising. The probability of brand choice conditional on size purchased and *conditional* on the type will still be independent of the various dynamic components. So one might think that the proposed split is still valid. 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 amounts to essentially solving the dynamic problem.

The main consequence of not having persistent preference shocks is that we rule out unobserved heterogeneity in brand preferences[14] that might generate interesting dynamics. For instance, a consumer with a high unobserved shock preference for Tide is likely not to react to deals in other brands and wait 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. Although the method does not allow for random effects, we can capture the same effects by allowing for heterogeneity by including observed household heterogeneity, by classifying the households into types (Buchinsky, Hahn and Hotz, 2005), by allowing for differences in the choice based on past behavior and by allowing for household-brand specific fixed effects. Since the household-brand dummy variables are estimated in a simple static logit they do not substantially increase the computational cost and we can allow for as much heterogeneity as there is in the data.[15] Notice we only need to estimate preferences for the brands actually purchased by each household. For brands that a household never bought during 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

[14]Notice that unobserved heterogeneity can be introduced in the dynamic part of the model.

[15]The estimation of fixed effects is in principle problematic in non-linear 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 ML. Since 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.

at prices outside the ones observed in the sample range. But that is a data limitation: the data do not inform us about out of sample prices anyway. Which of these methods is more appropriate depends on the particular application.

4.3.3 Alternative Methods

The estimation of the full model without the assumptions that lead to 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 space, which includes the inventories of all brands held by the household as well as the price vector of all brands in all sizes; plus all other promotional activities for each product/size combination.

Erdem et al. (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 proportionally 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. 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 concentrate on estimating the price process of the dominant container size, and assume that price differentials per ounce with other containers is distributed i.i.d. This simplifies the states and transitions from current to future prices, since only the prices of the dominant size are relevant state variables. Finally, to further simplify the state space they do not control for other promotional activities.

The main advantage of their method is to allow for unobserved preference heterogeneity. In contrast, our dynamic problem (third step) is considerably simpler and therefore can in practice be more flexible. For instance, we allow for unobserved heterogeneity in the dynamic estimation. Moreover, since most of the parameters of the model are estimated in the first step, without solving the dynamic programming problem, we can allow for a richer model that includes observed heterogeneity (demographics and fixed effects) and promotional activities.

4.4 Computing Standard Errors

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

In order 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 price process in the inclusive values we assume the following transition

$$Pr(\omega_{1,t}, \ldots, \omega_{S,t} \mid \omega_{1,t-1}, \ldots, \omega_{S,t-1}) =$$

$$N(\gamma_{10} + \gamma_{11}\omega_{1,t-1} + \ldots + \gamma_{1,S}\omega_{S,t-1}, \sigma_1) \cdot \ldots \cdot N(\gamma_{S0} + \gamma_{S1}\omega_{1,t-1} + \ldots + \gamma_{SS}\omega_{S,t-1}, \sigma_S)$$

where S is the number of different sizes and N denotes the normal distribution. We also explore below adding more lags. 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. Below we describe various ways in which we tested the robustness of the results. The estimation was performed using a sample of 221 households, 17,335 observations, where an observation is a visit to the store. The households were selected based on two criteria: (i) they made more than 10 observed 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.[16]

5.1 Parameter Estimates

The parameter estimates are presented in Tables 4-6. Table 4 presents the estimates from the first stage, which is a (static) Logit choice of brand 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 a household-brand specific effects. This is doable in our sample since 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, as in the literature on consideration sets (Chiang, Chib and Narasimhan, 1999). This reduces the number of choices significantly since most households purchase a small number of brands (the median households purchases 3 brands. See Table 1). Absence from the choice set is equivalent to a sufficiently negative (household specific) brand dummy variable.

The different columns in Table 4 present specifications which vary in the variables included. We conclude three things from this table. First, we note the effect of including feature/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.[17] The size of the change is intuitive. It implies

---

[16]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 for a small share of total sales). We think it is safer to keep them out of the sample, as they are the candidates to purchase detergent in alternative stores.

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

that the large effect on quantity sold seemingly associated with price changes are largely driven by the feature/display promotional activity. This suggests that one has to be careful in interpreting estimates that do not properly control for promotional activities other than sales.

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 et al. (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, is non-white and if the family has more than 4 people. These interactions are highly significant. The signs on the coefficients make sense. Larger families, non-white (which in our sample have lower income) and suburban shoppers are more price sensitive.

In order to allow for heterogeneity in brand preferences we interact, in column (v), the brand dummy variables with the demographic variables. Together the demographics variables generate 8 different "types" of households. In columns (viii)-(x) we allow for household-brand specific effects.

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 their preference proportional to the size of the container she is purchasing.

Table 5 reports the estimates of the price process. As explained above this process was estimated using the inclusive values (given in equation (5)) computed from the estimates of column (x) in Table 4. This index varies by household, since 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 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. In such a case the consumer is repeatedly surprised by the effect of advertizing (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. Under this alternative we have to compute the inclusive value net of the advertising effect. We opted for the internally consistent version of advertising.

The first four columns of Table 5 display the estimated parameters, for the two main sizes 64 and 128 ounces, 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 below). Within each set the first two columns displays results for a first-order Markov process. The point estimates suggest that the lagged value of own size is the most important in predicting the future prices. We see that letting the price process vary by household type seems to be important, as the results vary.

Considering the supply-side there are good reasons to believe that prices will not follow a first-order process.[18] In order to explore alternatives to the first-order Markov assumption, in the next set of columns we include the sum of 5 additional lags. We also estimated, but do not display, a specification which allows these 5 lags to enter with separate coefficients. Since these coefficients are similar for different lags we do not display this specification. These additional lags do not significantly improve the fit. Therefore, we concluded that the additional lags in the Markov process are not worth the extra computational complexity they entail.

Table 6 reports the results from the third stage, the dynamics quantity choice. We allow for 6 different types of households that vary by market and family size. For each type we allow for different utility and storage cost parameters. We also include size fixed effects that are allowed to vary by type, which are not reported in the table. Most of the parameters are statistically significant, at standard significance levels, and their implications are reasonable. Larger households have higher

[18]We note that the process we are trying to estimate is the process households use to form expectations. It is reasonable to believe that households do not remember more than one lag.

values of the coefficient on consumption, γ, which implies that holding everything else constant they consume more. Households that live in the suburban market (where houses are on average larger) have lower storage costs. Within each market larger households typically have lower storage costs.

To get an idea of the economic magnitude consider the following. If the beginning of period inventory is 65 ounces (the median reported below) then buying a 128 ounce bottle increases the storage cost, relative to buying a 64 ounce bottle, by roughly \$0.25 to \$0.75, depending on the household type. As we can see from Table 3, the typical savings from non-linear pricing is roughly \$0.40, which implies that the high storage cost types would not benefit from buying the larger size while the low storage costs would. This is consistent with the observed purchasing patterns.

For this sample the estimated median inventory held is 65-69 oz., depending on the type. Larger households and households in the suburban market hold a higher inventory. There is more variation across the types at the higher end of the distribution. The mean weekly consumption is between 20 and 29 oz., for different types (with the 10th and 90th percentiles varying between 6 and 7, and 54 and 95, respectively). If we assumed the households had constant consumption, equal to their total purchases divided by the number of weeks, we get very similar average consumption.

5.2 Fit and Robustness

Ideally in order 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. So instead we focus on the model's prediction of inter-purchase duration. Figure 1 displays the distribution of the duration between purchases (in weeks). In addition to the simulation from the model and the empirical distribution, we also present the distribution predicted by a static model with constant probability of purchase. Overall our model traces the empirical distribution quite closely. The modal and the median inter-purchase time are predicted correctly. We also examined the survival functions and hazard rates of no-purchase, both have reasonably good fit.

We tested the robustness of the results in several ways. First, we explored a variety of methods to solve the dynamic programming problem. Besides 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 of taking this solution to the data. First, 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 for the cost of inventory.

5.3 Implications

In this section we present the implications of the estimates, and compare them to static ones. In Table 7 we present a sample of 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 path of each of the different products (brand and size). Since we are interested in the long run effects these changes were always permanent changes in the process (i.e. a change in the whole path of prices not just the current price). We then re-estimated the price process (although in reality 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 7. Cell entries *i*, *j*, where *i* indexes row and *j* column, present the percent change in market share of brand *i* with a one percent change in 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.[19] 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 towards other similar size

[19]At first glance the cross-price elasticities might seem low. However, we have a large number of products: different brand 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.

containers of a different brand.

Second, we note that 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 since we are looking at long run responses which reflect purely a consumption effect, presumably small for detergents. The substitution from say Tide 128 oz. to the outside option represent forgone purchases due to the higher price level. Namely, the proportion of purchases which due to the permanent increase in the price of Tide 128 oz. (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 short run price change. As we discuss next, we indeed find that short run estimates grossly 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. As we previously noted there are two reasons why these will differ. First, the coefficients estimated in the static model are (upward) biased and inconsistent because of 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 8 present the ratio of the static estimates to the dynamic estimates. Cell entries i, j, where i indexes row and j 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 one percent change in 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 above results presented in Tables 4-6. 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 over estimate 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. This highlights the concern that if one uses own price elasticities to infer markups (through a first order condition) one will underestimate the extent of market power. More on this below.

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 since the static model fails to account for the effect of inventory. A short run price increase is most likely to chase away consumers that can wait for a better price, namely those with high inventories. Therefore, the static model will over estimate the substitution to the no purchase option.

There are several effects impacting the cross-price elasticities to the other brands. As we noted above 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, since they represent reactions to a permanent change in the price of Tide.

For own and cross price effects towards the no-purchase option both the econometric bias and the difference between short- and long- run effects operate in the same direction. However, it is unclear which one of the effects will dominate for the cross price elasticities. 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. Indeed the results seem to suggest that the bias in the cross-price elasticities is greater than that in the own price elasticities.

One might wonder how representative the results from our sample are to the population. The answer depends on our focus and the question we wish to answer. We care about the ratio for the "typical" consumer from an academic point of view: how important is inventory behavior. We think that we are likely underestimating the importance of consumer inventory since many of the consumers that shop elsewhere (e.g., at Walmart or Costco) will not be in our sample and are the ones that are more likely to store inventory. Alternatively, we might care about the ratio for a typical shopper at the supermarkets in the sample. In other words, what is the long-run elasticity of the demand faced by the supermarkets. In order to examine this we estimated a static demand system using the aggregate data, we found similar results to the static demand estimated using the household sample. This leads us to conclude that the results from our sample are at least reasonably representative.

Estimates of the demand elasticities are typically used in one of two ways. First, they are used in a first order condition, typically from a Bertrand pricing game, in order to compute price cost margins (PCM). For single product firms it is straight-forward to see the magnitude of the bias: it is the same as the ratio of the own-price elasticities. Therefore, the figures in Table 9 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 multi-product firms since the dynamic model finds that the products are closer substitutes (and therefore a multi-product 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). The above analysis suggests that this exercise will tend to find evidence of collusion where there is none, since the PCM predicted by models without collusion will seem too low.

A second important use of demand estimates is for simulation of the effects of mergers (e.g., Hausman, Leonard and Zona, 1994; and Nevo, 2000). The figures in Table 9 suggest that estimates from a static model would tend to underestimate the effects of a merger, because they will tend to underestimate the substitution among products. Furthermore, because the static estimates over estimate 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. Conclusions and Extensions

In this paper we structurally estimate a model of household inventory holding. Our estimation procedure allows us to introduce features essential to modeling demand for storable products, like: product differentiation, sales, advertizing and non-linear prices. The estimates suggest that ignoring the dynamics dictated by the ability to time purchases can have strong implications on demand estimates. We find that static estimates overestimate own price elasticities, underestimate cross price responses to other products and overestimate the substitution towards no purchase.

Compared to the standard static discrete choice models heavily used in the recent IO literature we have two advantages. On the model side, our model endogenizes consumption and allows for consumer inventory. Regarding data, in contrast to most of the literature we estimate the model with weekly household data. The high frequency of the price variability is in principle a blessing for estimating substitution patterns. However, for products that are storable we argue that the quantity responses to short run prices changes may confound inventory effects and lead to biases.

An important implication of the model is that the likelihood of the observed choices can be split between a dynamic and static component. The latter we estimate in our first step quite richly with little additional computational cost. The dynamic component, estimated in the third step, requires the usual computation burden (of numerically solving the dynamic programming and numerically searching for the parameters that maximize the likelihood). However, the computational burdened 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 potentially arising in a static estimation. The shortcut simply involves estimating demand conditional on the actual quantity purchased. The shortcut may prove useful in making the ideas easier to use in applied and policy work.

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

$$\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt} + u(c_j^* + \nu_t) - C(i_t + x_t - c_j^*) + \delta E(V(s_t) \mid d_{jt} = 1, x_t, c_j^*) >$$

$$\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt} + u(c_k^* + \nu_t) - C(i_t + x_t - c_k^*) + \delta E(V(s_t) \mid d_{jt} = 1, x_t, c_k^*)$$

and therefore

$$u(c_j^* + \nu_t) - u(c_k^* + \nu_t) > \delta E(V(s_t) \mid d_{jt} = 1, x_t, c_k^*) - \delta E(V(s_t) \mid d_{jt} = 1, x_t, c_j^*) + C(i_t + x_t - c_k^*) - C(i_t + x_t - c_j^*)$$

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

$$u(c_j^* + \nu_t) - u(c_k^* + \nu_t) < \delta E(V(s_t) \mid d_{jt} = 1, x_t, c_k^*) - \delta E(V(s_t) \mid d_{jt} = 1, x_t, c_j^*) + C(i_t + x_t - c_k^*) - C(i_t + x_t - c_j^*),$$

which is a contradiction.□

We now justify the equivalence result claimed in Section 4.2.3.

The dynamic problem defined in equation (1) has an associated Bellman equation

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

From Assumptions A1 and A3 both ϵ and ν are i.i.d., and current actions affect future utility only through end of period inventory, so we can write $E\left[V(s_{t+1}) \mid i_t, p_t, c, d_{jx}\right]$. In other words, the expectation of *V* given the state and current behavior is a function of current price and end of period inventories. Lets denote such 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 $V^e(i_{t+1}, p_t)$. Using the independence of ϵ, *v* and *p* we get

$$V^e(i_t,p_{t-1}) =$$

$$\int\left[\underset{\{c,d_{jx}\}}{Max}\left\{u(c+\nu_t) - C(i_{t+1}) + \sum_{j,x} d_{jx}(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt}) + \delta V^e(i_t + x - c, p_t)\right\}\right] dF(\epsilon)dF(\nu)dF(p_t|p_{t-1}).$$

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

$$V^e(i_t,p_{t-1}) = \int\left[\underset{d_{xj}}{Max}\left(\sum_{j,x} d_{jx}(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt}) + M(s_t,x)\right)\right] dF(\epsilon)dF(\nu)dF(p_t|p_{t-1}),$$

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

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

Using Assumption A4, 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):

$$V^e(i_t,\omega_{t-1}) = \int \log\left(\sum_x \exp\left(\omega_{xt} + \underset{c}{Max}\, u(c+\nu_t) - C(i_{t+1}) + \delta V^e(i_{t+1},\omega_t)\right)\right) dF(\nu)dF(\omega_t|\omega_{t-1}).$$

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

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

$$V(i_t,\omega_{t-1},\epsilon_t,\nu_t) = \underset{\{c,d_x\}}{Max}\left\{u(c+\nu_t) - C(i_{t+1}) + \sum_x d_x(\omega_{xt} + \epsilon_{xt}) + \delta V^e(i_{t+1},\omega_t)\right\}.$$

This can be seen by taking expectations on the right hand side. Therefore, $V^e(i_t,\omega_{t-1})$ characterizes the solution to both problems.

Finally, we want to show that we can rely on the likelihood computed from the simplified problem to derive the likelihood of purchase of quantity x in the original problem. We want to show that

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

Notice

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

while (abusing notation we denote $i_t(x)$ to remind the reader that the choice of x impacts end of period inventories):

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

$$\frac{\sum_j \exp\big(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + M(s_t, x)\big)}{\sum_{j,y} \exp\big(\alpha p_{jyt} + \xi_{jy} + \beta a_{jyt} + M(s_t, x)\big)} = \frac{\exp\big(M(s_t, x)\big)\exp\Big(\log\Big(\sum_j \exp\big(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt}\big)\Big)\Big)}{\sum_y \exp\big(M(s_t, y)\big)\exp\Big(\log\Big(\sum_j \exp\big(\alpha p_{jyt} + \xi_{jy} + \beta a_{jyt}\big)\Big)\Big)} =$$

$$\frac{\exp\big(\omega_{xt} + M(s_t, x)\big)}{\sum_y \exp\big(\omega_{yt} + M(s_t, y)\big)} = Pr(x_t | \omega_t, i_t, \nu_t)$$

where the last equality follows from the fact the M depends on ω but not on p.

References

Aguirregabiria, V. (1999) “The Dynamics of Markups and Inventories in Retailing Firms,” *The Review of Economic Studies*, 66, 275-308.

Aquirregabiria, V. (2005) “Nonparametric Identification of Behavioral Responses to Counterfactual Policy Interventions in Dynamic Discrete Decision Processes,” *Economics Letters*, forthcoming.

Ailawadi, K.L. and S. Neslin (1998). “The Effect of Promotion on Consumption: Buying More and Consuming it Faster,” *Journal of Marketing Research*, (89) pp. 390-98.

Bell, D., J. Chiang and V. Padmanabhan (1999), “The Decomposition of Promotional Response: An Empirical Generalization,” *Marketing Science*, 18, 504-26.

Bell, D. and Y. Boztug (2004). “The effect of Inventory on Purchase Incidence: Empirical Analysis of Opposing Forces of Storage and Consumption,” mimeo Wharton.

Benitaz-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,” Yale University, mimeo.

Berry, S., J. Levinsohn, and A. Pakes (1995), “Automobile Prices in Market Equilibrium,” *Econometrica*, 63, 841-890.

Bertsekas, D. and J. Tsitsiklis (1996), *Neuro-Dynamic Programming*, Athena Scientific.

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.

Blattberg, R. and S. Neslin (1990), *Sales Promotions*, Prentice Hall.

Boizot, C., J.-M. Robin, M. Visser (2001), “The Demand for Food Products. An Analysis of Interpurchase Times and Purchased Quantities,” *Economic Journal,* 111(470), April, 391-419.

Bresnahan, T. (1987), “Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War,” *Journal of Industrial Economics*, 35, 457-482.

Chevalier, J., A. Kashyap and P. Rossi (2003), “Why Don’t Prices Rise During Peak Demand

Periods? Evidence from Scanner Data," *American Economics Review*, 93, 1, 15-37.

Chiang J. (1991), "A Simultaneous Approach to the Whether, What and How Much to Buy Questions," *Marketing Science*, 10, 297-315.

Currim I. and L. Schneider (1991), "A Taxonomy of Consumer Purchase Strategies in a Promotion Intense Environment," *Marketing Science* 10, 91-110.

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.

Feenstra, R. and M. Shapiro (2001), "High Frequency Substitution and the Measurement of Price Indexes," in R. Feenstra and M. Shapiro, eds., *Scanner Data and Price Indexes*, Studies in Income and Wealth Vol. 64, Chicago: National Bureau of Economic Research..

Gupta S. (1988), "Impact of Sales Promotions on When, What, and How much to Buy," *Journal of Marketing Research*, 24, 342-55.

Hausman, J., G. Leonard, and J.D. Zona (1994), "Competitive Analysis with Differentiated Products," *Annales D'Economie et de Statistique*, 34, 159-80.

Hendel, I. (1999), "Estimating Multiple Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, 66, pp. 423-446.

Hendel, I. and A. Nevo (2002), "Sales and Consumer Stockpiling," working paper. (available at http://emlab.berkeley.edu/~nevo/)

Hosken, D., D. Matsa, and D. Reiffen (2000) "How do Retailers Adjust Prices: Evidence from Store-Level Data," working paper.

MacDonald, J. (2000), "Demand, Information, and Competition: Why Do Food Prices Fall At Seasonal Demand Peaks?," *Journal of Industrial Economics*, 48 (1), 27-45.

Magnac, T. And Themar D. (2002), "Identifying Dynamic Discrete Decision Processes," *Econometrica,* 70, 801-816.

McFadden, D. (1981), "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*, pp. 198-272, Cambridge: MIT Press.

Melnikov, O. (2001), "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," Cornell University, mimeo.

Murphy K. And R. Topel (1985), "Estimation and Inference in Two Step Econometric Models, *Journal of Business and Economics Statistics*, 370-379.

Neslin S., C. Henderson and J. Quelch (1985), "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science*, 4(2), 147-165.

Neslin, S. and L. Schneider Stone (1996), "Consumer Inventory Sensitivity and the Postpromotion Dip," *Marketing Letters*, 7, 77-94.

Nevo, A. (2000), "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 31(3), 395-421.

Nevo, A. (2001),"Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), 307-342.

Pesendorfer, M. (2002), "Retail Sales. A Study of Pricing Behavior in Supermarkets,"*Journal of Business*, 75(1), 33-66.

Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999-1033.

Rust, J. (1996), "Numerical Dynamic Programming in Economics," in H. Amman, D. Kendrick, and J. Rust (eds.), *Handbook of Computational Economics*, Volume 1, 619-729.

Train, K. (1986), *Qualitative Choice Analysis: Theory, Econometrics and Application to Automobile Demand*, Cambridge: MIT Press.

Warner, E. and R. Barsky (1995), "The Timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays,"*Quarterly Journal of Economics*;110(2), 321-52.

Table 1
Summary Statistics of Household-level Data

	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

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.

Table 2
Brand Volume Shares and Fraction Sold on Sale

	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

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.

Table 3
Quantity Discounts and Sales

	price/ discount ($ / %)	quantity sold on sale (%)	weeks on sale (%)	average sale discount (%)	quantity share (%)
Liquid					
32 oz.	1.08	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.	0.61	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

All cells are based on data from all brands in all stores. The column labeled *price/discount* presents the price per 16 oz. for the smallest size and the 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.

Table 4
First Step: Brand Choice Conditional on Size

	(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	-0.56	-1.40	-0.73
	(.022)	(.038)	(.043)	(.050)	(.052)	(.055)	(.056)	(.085)	(.092)	(.098)
*suburban dummy				-0.33	-0.30	-0.34	-0.33	-0.26	-0.46	-0.20
				(.055)	(.061)	(.055)	(.056)	(.113)	(.127)	(.127)
*non-white dummy				-0.34	-0.39	-0.38	-0.33	-0.35	-0.34	-0.27
				(.075)	(.083)	(.076)	(.076)	(.152)	(.166)	(.168)
*large family				-0.23	-0.13	-0.21	-0.22	-0.46	-0.39	-0.43
				(.080)	(.107)	(.080)	(.082)	(.181)	(.192)	(.195)
feature			1.06	1.05	1.07	0.92	0.93	1.09		1.05
			(.095)	(.096)	(.097)	(.099)	(.100)	(.123)		(.126)
display			1.19	1.17	1.20	1.14	1.15	1.55		1.52
			(.069)	(.070)	(.071)	(.071)	(.072)	(.093)		(.093)
brand dummy var		✓	✓	✓	✓					
* demographics					✓					
*size						✓				
brand-size dummy var							✓			
brand-HH dummy var								✓		
*size									✓	✓

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 in parentheses.

Table 5
Second Step: Estimates of the Price Process

	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}
$\omega_{1,t-1}$	.003	-.014	.005	.014	-.023	-.005	-.019	.007
	(.012)	(.011)	(.014)	(.014)	(.017)	(.014)	(.019)	(.015)
$\omega_{2,t-1}$	.413	.033	.295	.025	.575	-.003	.520	.011
	(.007)	(.010)	(.008)	(.007)	(.013)	(.010)	(.016)	(.013)
$\omega_{3,t-1}$	.003	-.034	.041	-.006	.027	-.072	.051	-.018
	(.007)	(.007)	(.009)	(.009)	(.020)	(.016)	(.025)	(.020)
$\omega_{4,t-1}$	.029	.249	.026	.236	-.018	.336	-.018	.274
	(.008)	(.008)	(.008)	(.017)	(.020)	(.016)	(.021)	(.017)
$\sum_{\tau=2}^{5} \omega_{1,t-\tau}$			-.003	-.012			-.008	-.003
			(.005)	(.004)			(.006)	(.005)
$\sum_{\tau=2}^{5} \omega_{2,t-\tau}$			.089	.006			.073	-.004
			(.003)	(.002)			(.005)	(.004)
$\sum_{\tau=2}^{5} \omega_{3,t-\tau}$			-.008	-.009			-.004	-.016
			(.003)	(.003)			(.008)	(.006)
$\sum_{\tau=2}^{5} \omega_{4,t-\tau}$			-.013	.018			-.008	.056
			(.003)	(.003)			(.007)	(.005)

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 4. 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, households in market 1 with large families. Results for other types are available from the authors.

Table 6
Third Step: Estimates from the Nested DP Problem

household type:	1	2	3	4	5	6
coefficient on:						
Cost of inv - linear	4.44	1.36	1.03	4.06	1.73	1.42
	(10.32)	(0.13)	(0.10)	(0.36)	(0.26)	(0.14)
Cost of inv - quadratic	13.93	7.45	6.84	6.33	3.65	1.81
	(6.92)	(1.31)	(0.77)	(9.03)	(0.65)	(0.86)
Utility from consumption	0.61	1.49	2.02	0.63	1.68	1.86
	(0.12)	(0.67)	(0.52)	(0.05)	(0.26)	(0.16)
Log likelihood	-1350.6	-2700.7	-3441.0	-2266.3	-1458.8	-3017.8

Asymptotic standard errors in parentheses. Also included are size fixed effects, which are allowed to vary by household type. Types 1-3 live in the urban market, while 4-6 live in the suburban market. Within each market the type increases with family size.

Table 7
Long Run Own and Cross-Price Elasticities

#	Brand	Size (oz.)	All*	Wisk	Surf	Cheer	Tide	Private Label
1	All*	32	0.316	0.126	0.031	0.039	0.118	0.000
2		64	0.465	0.089	0.049	0.030	0.081	0.006
3		96	0.677	0.082	0.028	0.027	0.092	0.003
4		128	-2.485	0.147	0.083	0.052	0.105	0.007
5	Wisk	32	0.083	0.757	0.044	0.012	0.150	0.006
6		64	0.078	0.606	0.043	0.012	0.106	0.003
7		96	0.058	0.689	0.048	0.019	0.141	0.011
8		128	0.127	-2.887	0.083	0.023	0.130	0.005
9	Surf	32	0.052	0.065	0.707	0.019	0.333	0.005
10		64	0.141	0.087	0.897	0.023	0.147	0.005
11		96	0.144	0.097	0.714	0.014	0.191	0.001
12		128	0.196	0.151	-3.436	0.038	0.212	0.008
13	Cheer	64	0.163	0.047	0.024	0.831	0.280	0.001
14		96	0.152	0.012	0.007	0.965	0.522	0.001
15		128	0.249	0.088	0.056	-3.338	0.434	0.003
16	Tide	32	0.068	0.081	0.040	0.020	1.053	0.002
17		64	0.044	0.052	0.022	0.023	0.871	0.001
18		96	0.043	0.063	0.011	0.026	1.085	0.001
19		128	0.068	0.087	0.036	0.042	-2.681	0.001
20	Solo	64	0.064	0.076	0.028	0.021	0.148	0.002
21		96	0.222	0.038	0.024	0.035	0.074	0.000
22		128	0.125	0.129	0.057	0.042	0.299	0.001
23	Era	32	0.028	0.170	0.030	0.029	0.368	0.000
24		64	0.028	0.100	0.035	0.019	0.276	0.008
25		96	0.029	0.178	0.023	0.029	0.333	0.001
26		128	0.058	0.185	0.058	0.030	0.488	0.014
27	Private	64	0.118	0.115	0.065	0.036	0.074	0.233
28	Label	128	0.167	0.256	0.095	0.019	0.065	-2.611
29	No purchase		0.005	0.001	0.003	0.001	0.009	0.000

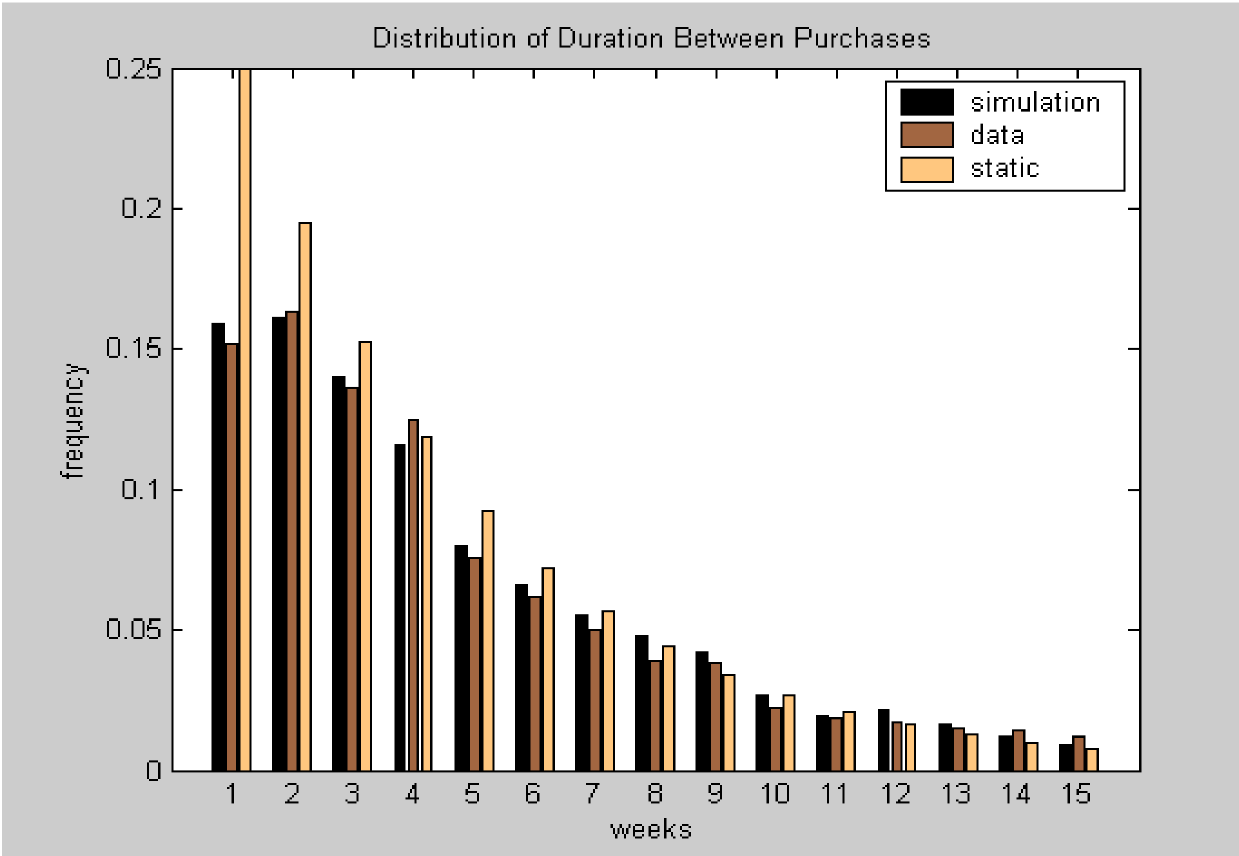
Cell entries *i*, *j*, where *i* indexes row and *j* column, give the percent change in market share of brand *i* with a one percent change in price of *j*. All columns are for a product 128 oz, the most popular size. Based on the results of Tables 4-6.
(*) Note that "All" is a name of a detergent produced by Unilever.

Table 8
Average Ratios of Elasticities Computed from a Static Model to Long Run Elasticities Computed from the Dynamic Model

			64 oz.						128 oz.					
#	Brand	Size (oz.)	All*	Wisk	Surf	Cheer	Tide	Private Label	All*	Wisk	Surf	Cheer	Tide	Private Label
1	All*	64	1.047	0.136	0.141	0.117	0.129	0.163	0.141	0.182	0.182	0.193	0.213	0.355
2		128	0.182	0.252	0.276	0.218	0.286	0.384	1.253	0.098	0.118	0.107	0.166	0.213
3	Wisk	64	0.137	1.227	0.127	0.161	0.123	0.133	0.162	0.222	0.149	0.250	0.270	0.215
4		128	0.244	0.263	0.228	0.318	0.271	0.278	0.078	1.429	0.076	0.145	0.203	0.116
5	Surf	64	0.138	0.134	1.361	0.160	0.124	0.136	0.182	0.184	0.126	0.186	0.236	0.289
6		128	0.258	0.218	0.173	0.271	0.250	0.174	0.119	0.107	1.205	0.085	0.162	0.139
7	Cheer	64	0.118	0.170	0.156	1.407	0.090	0.133	0.145	0.247	0.179	0.136	0.227	0.289
8		128	0.239	0.260	0.247	0.118	0.227	0.199	0.094	0.126	0.063	1.158	0.158	0.065
9	Tide	64	0.168	0.168	0.135	0.125	1.282	0.158	0.242	0.291	0.179	0.287	0.231	0.393
10		128	0.255	0.319	0.231	0.244	0.212	0.323	0.118	0.174	0.085	0.143	1.446	0.331
11	Solo	64	0.154	0.125	0.148	0.146	0.122	0.147	0.172	0.133	0.149	0.309	0.304	0.294
12		128	0.223	0.205	0.229	0.223	0.207	0.278	0.073	0.063	0.061	0.165	0.173	0.218
13	Era	64	0.210	0.127	0.129	0.123	0.101	0.179	0.463	0.178	0.163	0.202	0.208	0.364
14		128	0.295	0.228	0.244	0.225	0.169	0.330	0.176	0.079	0.090	0.104	0.104	0.222
15	Private	64	0.185	0.158	0.135	0.168	0.169	1.396	0.336	0.232	0.157	0.292	0.340	0.272
16	Label	128	0.307	0.292	0.349	0.334	0.418	0.311	0.164	0.126	0.132	0.103	0.300	1.327
17	No purchase		2.298	1.274	1.366	1.383	1.375	3.692	2.342	2.962	2.829	3.670	3.723	3.509

Cell entries *i*, *j*, where *i* indexes row and *j* 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 one percent change in 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 4-6.
(*) Note that "All" is a name of a detergent produced by Unilever.

Figure 1


Distribution of Duration Between Purchases
simulation
data
static
frequency
weeks
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15