Measuring the Implications of Sales and Consumer Stockpiling Behavior[1]

Igal Hendel
University of Wisconsin, Madison and NBER

and

Aviv Nevo
University of California, Berkeley and NBER

June 2003

ABSTRACT

Temporary price reductions (sales) are common for many goods and naturally result in large increases in the quantity sold. In previous work we found that the data support the hypothesis that these increases are, at least partly, due to stockpiling. In this paper we quantify the extent of stockpiling and assess its economic implications. We construct and 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, may: (i) overestimate own price elasticities by 30 percent; (ii) underestimate cross-price elasticities to other products by up to a factor of 4; and (iii) overestimate the substitution to the no purchase, or outside option, by up to 150 percent.

---

[1]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. We gratefully acknowledge support from the NSF (SES-0093967 and SES-0213976). Comments are very welcome, please direct them to igal@ssc.wisc.edu or nevo@econ.berkeley.edu.

1. Introduction

Many non-durable consumer products exhibit occasional short-lived price reductions, sales. In a previous paper (Hendel and Nevo, 2002) we documented purchasing patterns in the presence of sales, at the household and the store level. We argued that these purchasing patterns are due, at least partly, to stockpiling. When prices go down consumers buy for future consumption. Stockpiling has implications for the interpretation of demand estimates of storable products. In this paper, we present a dynamic model of household behavior. Our model captures the main features faced by the household: variation in prices over time, which create incentives to store, several closely related brands, non-linear pricing and promotional activities like advertising and display. We structurally estimate the model in order to study the economic implications of stockpiling behavior.

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; Hendel, 1999; Nevo, 2001; as well as many others). Virtually all the applications (including our own work) have neglected dynamics. The estimation is performed assuming that the demand for the product is independent of the history. We propose a framework to incorporate the dynamics dictated by stockpiling into the estimation of demand for a storable product. Our goal is assess and quantify the implications of stockpiling on demand estimation. In particular, we aim to compare the estimates we obtain from a dynamic model to those achieved by the 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 long run changes in prices. In contrast, static demand estimation methods will miss the target for two reasons: First, by neglecting dynamics these model are misspecified. They do not correctly control for the relevant history like past sales and prices, and inventories. Second, even adding all the right controls static estimation would capture reactions to short run price movements, which confound the long run price effect we are after with a short run stockpiling effect.[2] A simple back of the envelope calculation presented in Hendel and Nevo (2002) shows that neglecting dynamics may significantly overstate price sensitiveness.

Stockpiling has also 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.

In a previous paper (Hendel and Nevo, 2002) we documented buying patterns at the household and store level which are consistent with the predictions of a stockpiling model (see details in Section 2.3).[3] Since we (i) ignored many of the important aspects of the market (in order to get testable predictions) and (ii) did not attempt to estimate the model structurally, we were unable to assess the economic implications detailed above.

In this paper we structurally estimate a model of household demand. Households face uncertain future prices. In each period a household decides how much to buy, which brand to buy and how much to consume. These decisions are made to maximize the present expected value of future utility flows. Households purchase for two reasons: for current consumption and to build inventories. Consumers increase inventories when the difference between the current price and the expected future price is lower than the cost of holding inventory.

In order to estimate the model we use 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 over a period of 104 weeks. We know exactly which product was bought, where it was bought, how much was paid and whether a coupon was used. We also know when the households visited a supermarket but decided not to purchase a laundry detergent.

---

[2] This point has been made by Erdem, Imai and Keane (2003). See there for a comparison of short run and long run elasticities. Below we relate our work to theirs.

[3] Pesendorfer (2002) also finds evidence that is consistent with stockpiling.

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 use the stochastic structure of the model to show that the probability of choosing any brand-size combination can be separated into the probability of choosing a brand conditional on quantity, and the probability of choosing quantity. Furthermore, the probability of choosing a brand conditional on quantity does not depend in our model 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 very 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, which we believe are natural for the product in question, as well as the limitations of the method.

Our results suggest that ignoring the dynamics can have strong implications on demand estimates. By 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 0.22. Third, the estimates from the static model overestimate the substitution to the no purchase, or outside option, by up to 150 percent. These imply that if the a standard analysis is based on static elasticity estimates it will underestimate price-cost margins and under predict the effects of mergers.

Before we proceed we quantify the potential gains from dynamic behavior for the type of products we study. By quantifying the potential gains we want to get a sense of the incentives to stockpile generated by the observed price fluctuations. To do so we compare the actual amount paid by each household in the data to what they would have paid, for the same bundle of products, if prices were drawn randomly from the distribution of prices observed in the same locations they shopped. This is only an approximation which might underestimate the potential gains from exploiting sales because it takes actual behavior as fully optimal, while some consumers might have a cost to fully optimizing and therefore rationally decide to not fully exploit the gains from sales. On the other hand by keeping the purchased bundle constant it may overestimate the gains. In our data the average household pays 12.0 percent less for detergents than if they were to buy the exact same bundle at the average price. Replicating the exercise across other products we find an average saving of 12.7 percent.[4] Some households save little, i.e., they are essentially drawing prices at random, while others save a lot (the 90th percentile save 23 percent). Assuming savings in the 24 categories we examine represent saving in groceries in general, the total amount saved by the average household in our sample, over two years, is 500 dollars (with 10th and 90th percentiles of 150 and 860 dollars, respectively). These numbers show non-negligible incentives for households to time their purchases.

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. Hosken et al. (2000) study the probability of a product being put 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) also study the relation between seasonality and sales. The effect we study complements the seasonality they focus on. The same is also true for Aguirregabiria (1999), who studies retail inventory behavior. His

[4] This is for the 24 products in our data set. These products account for 22 percent of their total grocery expenditure.

paper is about firm's inventory policy and its effect on prices, while our focus is on consumers' inventory policies given the prices they face. 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.[5]

The closest paper to ours is Erdem, Imai and Keane (2003). They were the first to structurally estimate a consumer inventory model in the economics literature.[6] 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 assume that all brands are consumed proportionally to the quantity in storage. Together with the assumption that brand differences in quality enter linearly in the utility function this implies that only the total inventory and a quality weighted inventory matter as state variables, instead of the whole vector of brand inventories. The estimation method used by Erdem et. al. is more computationally burdensome, but more flexible in modeling of unobserved product heterogeneity. Our method can, in practice, more flexibly control for observed heterogeneity, and due to the computational simplicity can handle a larger choice set. Modeling and estimation differences between their method and our method render each better suited for different applications. In particular, with current computational constraints their method would be difficult to apply in the industry we study. In addition to the modeling and computational differences we differ in the focus. Their focus is on 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

[5]There is also a large marketing literature on the effects of sales, or more generally promotions, which we do not try to survey here. See Blattberg and Neslin (1990) and references therein.

[6]In the marketing literature there were attempts to estimate an inventory model in a very rudimentary set up, for example assuming consumption is constant (e.g. Gonul and Srinivassan, 1996). In the economics literature Boizot et. al. (2001) estimated the implications of an inventory model, but not the model itself.

through standard static methods. We compare the models in more detail in Section 4.

2. Data, Industry and Preliminary Analysis

2.1 *Data*

We use a scanner data set 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. For each detailed product (brand-size) in each store in each week we know 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, how much was paid, and whether a coupon was used.

Table 1 displays statistics of some household demographics, characteristics of household laundry detergents purchases (the product we focus on below) and store visits in general. The typical (median) household buys an 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 the purchases are concentrated at two main brands, which differ by household (because as we will see below 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*

We focus on laundry detergents. 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,[7] 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 we can see in Table 1, there is some variation across brands in the percent quantity sold on sale.

Detergents come in several different sizes. However, about 97 percent of the volume of liquid detergent sold was sold in 5 different sizes.[8] Sizes of powder detergent are not quite as standardized, and have small deviations across the sizes of liquid detergents. 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, therefore, also examined the pricing patterns for specific brands and essentially the same patterns emerged.

The figures in Table 3 average across sale and non-sale periods. Therefore, in principle, the pattern observed in the first column of Table 3 could be driven by more (and/or larger) sales for the larger sizes instead of quantity discounts. Indeed columns 2 through 5 of Table 3 confirms that the larger sizes have more frequent sales and larger discounts. However, these are not enough to explain the results in the first column. Indeed the quantity discounts can also be found in the "regular", non-

[7]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.

[8]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.

sale, price.

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.[9] 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. However, this is driven by 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 this section we summarize the preliminary analysis that suggests that stockpiling is a relevant phenomenon. This analysis is described in detail in Hendel and Nevo (2002). There we present a model similar to the one below, but ignore two important features of the data: non-linear pricing and product differentiation. We use the model to derive predictions regarding observed variables and test these predictions in the data.

The results support the model's predictions in the following ways. First, 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 periods.[10] Both these effects are predicted by the 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

---

[9] These variables both have several categories (for example, type of display: end, middle or front of aisle). We treat these variables as dummy variables.

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

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. Fourth, even though we do not observe the household inventory, by assuming constant consumption over time we can construct a measure of implied inventory. We find that this measure of inventory is negatively correlated with the quantity purchased and with the probability of buying. 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. All these finding are consistent with the predictions of the inventory model.

In the presence of stockpiling, standard demand estimation which neglects inventory behavior may be misleading. Our goal below is to get precise estimates of the magnitude of these effects.

3. The Model

3.1 The Basic Setup

We consider a model in which a 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 current marginal utility from consumption, θ_h is a vector of consumer-specific taste parameters, m_{ht} is utility from the outside good, which is multiplied by the marginal utility of income, α_h. The stochastic shock, ν_{ht}, introduces randomness in the consumer's needs, unobserved to 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, decrease demand and making it more elastic. The product is offered in J different varieties, or brands. The consumer faces random and potentially non-linear prices.

The good is storable. Therefore, the consumer at each period has to decide which brand to buy, how much to buy and how much to consume.[11] Quantity not consumed is stored as inventory. In the estimation we assume that the purchase amount, denoted by x_{ht}, is simply a choice of size (i.e., the consumer chooses which size box and not how many boxes). We denote a purchase of brand j and size x by $d_{hjxt} = 1$, where $x = 0$ stands for no purchase, and we assume $\sum_{j,x} d_{hjxt} = 1$.[12] 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

$$V(\Omega_0) = \max_{\{c_{ht}, d_{hjxt}\}} \sum_{t=0}^{\infty} \delta^t E\left[u(c_{ht}, \nu_{ht}; \theta_h) - C_h(i_{ht}) + \sum_j d_{hjxt}\left(\alpha_h p_{jxt} + \xi_{hjx} + \beta_h a_{jxt} + \epsilon_{hjxt}\right) | \Omega_0\right]$$

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

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

where Ω_t denotes the information at time t, $\delta > 0$ is the discount factor, $C_h(i_{ht})$ is the cost of storing inventory, ξ_{hjx} is a taste of brand j that could be a function of brand characteristics, size and could vary by consumer, $\beta_h a_{jxt}$ captures the effect of advertising variables on the consumer choice, and ϵ_{hjxt}, is a random shock that impacts the consumer's choice. Notice, the latter is size specific, namely, different sizes get different draws introducing randomness in the size choice as well. In equation (1) we want to emphasize that all functions are allowed to vary by household (as we will see in the results section). In order to simply notation we drop the subscript h in what follows.

The information set at time t consists of the current (or beginning of period) inventory, i_{t-1}, current prices, the shock to utility from consumption, ν_t, and the vector of ϵ 's. Consumers face two sources of uncertainty: future utility shocks and random future prices. We assume the consumer knows the current shock to utility from consumption, ν_t, which are independently distributed over

---

[11]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, we feel 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, and we want to make sure that are results are not driven by assuming it away. Moreover, reduced form results in Hendel and Nevo (2002) suggest consumption effects are present in the data.

[12]In our data more than 97 percent of the purchases are for a single unit. In principle our model could allow for multiple purchases, but we do not believe this is an important issue in this industry.

time. Prices are (exogenously) set according to a first-order Markov process, which we describe in Section 4.[13] Finally, the random shocks, ϵ_{jxt}, are assumed to be independently and identically distributed according to a type I extreme value distribution. We discuss in detail the model, its limitations and comparison to alternative methods in Section 4.3.

Notice that product differentiation as it appears in equation (1) takes places exclusively 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 inventories of each brand we only need to keep track of total quantity in inventory, regardless of brand.

This assumption is consistent with a more general model of differentiation in consumption as long as two conditions hold. The conditions needed are low discounting, which is very reasonable in our application given that we are using weekly data, and brand-specific differences in the utility from consumption enter linearly in the utility function. We discuss these conditions in more detail in Section 4.3.1.

4. Econometrics

The structural estimation is based on the nested algorithm proposed by Rust (1987), but has to deal with issues special to our problem. We start by providing a general overview of our estimation procedure and then discuss some of the 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, i.e., for any value of the state variables the consumer's optimal purchase and consumption.

[13] In principle, we can deal with the case where utility shocks, v, are correlated over time. However, this significantly increases the computational burden since the expectation in equation (1) will also be taken conditional on ν_t (and potentially past shocks as well). Also, in Section 4 we will show how we can allow for a higher order Markov process in prices.

However, since we do not observe the random shocks, which are state variables, from our perspective the decision rule is stochastic. Assuming a distribution for the unobserved shocks we derive a likelihood of observing the decisions of each consumer (conditional on prices and inventory). We nest this computation of the likelihood into a non-linear search procedure that finds the values of the parameters that maximize the likelihood of the observed sample.

We face two main hurdles in implementing the above algorithm. First, we do not observe inventory since both the initial inventory and consumption decisions are unknown. We deal with the unknown inventories by using the model to derive the optimal consumption in the following way.[14] Assume for a moment that the initial inventory is observed. Therefore, we can use the procedure described in the previous paragraph to obtain the likelihood of the observed purchases, and the optimal consumption levels (which will 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. The standard procedure is to use the estimated distribution of inventories itself to generate the initial distribution. In practice we do so by starting at an arbitrary initial level, and using part of the data (the first few observations) to generate the distribution of inventories implied by the model.

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

$$Pr(d_1...d_T|p_1...p_T) = \int \prod_{t=1}^{T} Pr\Big(d_t|p_t, i_{t-1}(d_{t-1},...,d_1,\nu_{t-1},...\nu_1,i_0),\nu_t\Big)dF(\nu_1,...\nu_T)\,dF(i_0). \qquad \textbf{(2)}$$

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 includes more information than

[14] 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).

prices, for instance, promotional activities. The above probability implicitly incorporates the first order Markov assumption on prices and the independence (over time) assumptions on ν and ϵ. Given the assumption that ϵ_{jxt} follows an i.i.d. extreme value distribution,

$$Pr(d_j|p_t,i_{t-1},\nu_t) = \frac{\exp\left(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \max_{c_t}\left\{u(c_t+\nu_t) - C(i_t) + \delta EV(\Omega_{t+1};d_t,c_t,p_t)\right\}\right)}{\sum_{k,x}\exp\left(\alpha p_{kxt} + \xi_{kx} + \beta a_{kxt} + \max_{c_t}\left\{u(c_t+\nu_t) - C(i_t) + \delta EV(\Omega_{t+1};d_t,c_t,p_t)\right\}\right)} \quad \textbf{(3)}$$

where $EV(\cdot)$ is the expected future value given today's state variables and today's decisions. Note that the summation in the denominator is over all brands and all sizes.

The second problem is the dimensionality of the state space. If there were only a few brand-size combinations offered at a small number of prices, then the above would be computationally feasible. In the data, over time, households buy several brand-size combinations, which are offered at many different prices. The state space includes not only the individual specific inventory and shocks, but also the prices of all brands in all sizes and their promotional activities. The state space and the transitions probabilities across states, in full generality, make the above standard approach computationally infeasible.

We therefore propose the following three-step procedure. The validity and limitations of the method are detailed in the next section. 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 logit (or varying parameters logit) 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" for each size (quantity) and their transition probabilities from period to period. This allows us, in the final step, to apply the nested algorithm discussed above to a simplified problem in order to estimate the rest of the parameters. 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. For the products we study this is a considerable reduction in the dimension of the state space. 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.

Intuitively, the logit structure enables the decomposition 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 key 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.[15]

4.2 The Three Step Procedure

We now discuss the details of the estimation. We show that the break-up of the problem follows from the primitives of the model, namely, it is consistent with the problem, not an approximation.

4.2.1 Step 1: Estimation of the "Static" Parameters

In the first step, we estimate part of the preference parameters using a static model of brand choice conditional on the size purchased. We now show that the static estimation is valid in the context of our model.

The probability in equation (2) can be used to form a likelihood, but it requires solving for $EV(\cdot)$, which implies solving the dynamic programming problem. Instead, we use a simpler approach. We can write

$$Pr(d_t|p_t,i_{t-1},\nu_t) \equiv Pr(d_{jt}=1,x_t|p_t,i_{t-1},\nu_t) = Pr(d_{jt}=1\,|p_t,x_t,i_{t-1},\nu_t)Pr(x_t|p_t,i_{t-1},\nu_t).$$

In general, this does not help us since we need to solve the consumer's dynamic programming problem in order to compute $Pr(d_{jt}=1\,|p_t,x_t,i_{t-1},\nu_t)$. However, given the primitives of our model,

---

[15] We are aware of two instances in the literature where a similar idea was used. 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 Melinikov (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.

conditional on the size purchased the optimal consumption is the same regardless of which brand is chosen (see proof in the Appendix). Since the brand of the inventory does not affect future utility, i.e., $EV(\Omega_{t+1};d_{jt}=1,x_t,c_t)=EV(\Omega_{t+1};x_t,c_t)$, then the term $\max_{c_t}\{u(c_t+\nu_t)-C(i_t)+\delta EV(\Omega_{t+1};d_{jt}=1,x_t,c_t)\}$ is independent of brand choice, thus, after the appropriate cancellations in equation (3) we obtain

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

where the summation is over all brands available in size x_t at time t. Thus, we can factor the probability in equation (2) into the probability of observing the brand choices and the probability of observing the sequence of quantity (size) choices.

Our approach is to estimate the marginal utility of income, the vector of parameters β and the parameters that enter ξ_{jx} by maximizing the product, over time and households, of $Pr(d_{jt}=1|p_t,x_t)$. To compute this probability we do not need to solve the dynamic programming problem, nor do we need to generate an inventory series. This amounts to estimating a brand choice logit model using only the choices with the same size as the size actually purchased. Next, we estimate the rest of the parameters of the model by maximizing the likelihood of observing the sequence of quantity purchases by each household.

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, in the second step, using the estimates from the first stage, we compute the "inclusive values" for each size (quantity) and their transition probabilities from period to period. Below we show how these inclusive values are used. The inclusive value for each size

$$\omega_{xt}=\log\left\{\sum_k \exp(\alpha p_{kxt}+\xi_{kx}+\beta a_{kxt})\right\}. \qquad \textbf{(4)}$$

can be thought of as a quality adjusted price index for all brands of size *x*. All the information needed to compute the inclusive values and their transition probabilities is contained in the estimates from the first stage. Note, that since the parameters might vary with consumer characteristics the inclusive values are consumer specific.

As we show below the original problem can be written such that the state space collapses to a single index per size, therefore reducing the computational cost. For example, 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. We assume that the inclusive values follow a Markov process and estimate, using the results of step one, the following transition

$$Pr(\omega_{1,t},...,\omega_{S,t}|\omega_{1,t-1},...,\omega_{S,t-1}) =$$

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

where S is the number of different sizes and $N(\cdot,\cdot)$ denotes the normal distribution.

This transition process is potentially restrictive, but can be generalized (for example, to include higher order lags) and tested in the data. The main loss is that transition probabilities have to be defined in a somewhat limited fashion. Two price vectors that yield the same vector of inclusive values will have the same transition probabilities to next period state, while a more general model will allow these to be different. In reality, however, we believe this is not a big loss since it is not practical to specify a much more general transition process.

4.2.3 Step 3: The Simplified Dynamic Problem

In the third, and final, step we feed the inclusive values, and the estimated transition probabilities, into the nested algorithm to compute the likelihood of purchasing a size (quantity). We now justify this step.

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

$$V(i_{t-1},p_t,\epsilon_t,\nu_t) =$$

$$\underset{\{c_t,d_{jxt}\}}{Max}\left\{u(c_t+\nu_t)-C(i_t)+\sum_{j,x} d_{jxt}(\alpha p_{jxt}+\xi_{jx}+\beta a_{jxt}+\epsilon_{jxt}) + \delta E\left[V(i_t,p_{t+1},\epsilon_{t+1},\nu_{t+1})|i_{t-1},p_t,\epsilon_t,\nu_t,c_t,x_t,d_{jt}\right]\right\}.$$

Note that since both ϵ and ν are identically and independently distributed, and the choice of brand j does not affect future utility, then we can write $E\left[V(i_t,p_{t+1},\epsilon_{t+1},\nu_{t+1})|i_{t-1},p_t,c_t,x_t\right]$. Moreover, consumption and purchases, c_t and x_t, affect future expected utility only through i_t. Thus we can rewrite this expectation as a function of current prices and inventory exclusively, namely $EV(i_t,p_t)$.

Using the independence of ϵ ,v and p

$$EV(i_t,p_t) =$$

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

By Lemma 1 (in the Appendix) optimal consumption depends on the quantity purchased but not of the brand chosen; then $Max_{c_t} u(c_t+\nu_t) - C(i_t) + \delta EV\left(i_{t+1},p_{t+1}\right)$ varies by size, *x*, but is independent of choice of brand *j*. Therefore,

$$EV(i_t,p_t) =$$

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

which is equal to (McFadden, 1981)

$$\int \log\left(\sum_{j,x} \exp\left(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \underset{c_t}{Max} \left\{ u(c_t+\nu_t) - C(i_t) + \delta EV\left(i_{t+1},p_{t+1}\right) \right\} \right) \right) dF(\nu)dF(p_{t+1}|p_t).$$

Using the definition of the inclusive values given in equation (4) this last expression can be written as

$$\int \log\left(\sum_{x} \exp\left(\omega_{xt} + \underset{c_t}{Max}\, u(c_t+\nu_t) - C(i_t) + \delta EV\left(i_{t+1},p_{t+1}\right) \right) \right) dF(\nu)dF(p_{t+1}|p_t).$$

Furthermore, if we assume, as we did above, that the transition probabilities, $F(p_{t+1}\,|\,p_t)$ can be fully summarized by $F(\omega_{t+1}\,|\,\omega_t)$, then $EV(\cdot)$ can be written as a function of ω_t and i_t instead of (p_t and i_t).

Using this result and substituting the definition of the inclusive value into equation (3) we can write

$$Pr(x_t\,|\,i_{t-1},p_t,\nu_t) = Pr(x_t\,|\,i_{t-1},\omega_t,\nu_t) = \frac{\exp\left(\omega_{xt} + Max_{c_t}\left\{u(c_t+\nu_t) - C(i_t) + \delta EV(i_{t+1},\omega_t;x_t,c_t)\right\}\right)}{\sum_{x} \exp\left(\omega_{xt} + Max_{c_t}\left\{u(c_t+\nu_t) - C(i_t) + \delta EV(i_{t+1}\,\omega_t;x_t,c_t)\right\}\right)}.$$

It is this probability that we use to construct a likelihood function in order to estimate the remaining parameters of the model.

The likelihood is a function of the value function, which despite the reduction in the number of state variables, is still computationally burdensome to solve. To solve the dynamic programming problem 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 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 computed 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 Discussion

In this section we discuss the limitations and advantages of our method, as well as alternative methods. Before doing so we address what features of the data identify the model.

The identification of the static parameters is the standard one. Variation over time in prices and advertising enable the estimation of household's sensitivity to price and to promotional activities, namely identification of 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.

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 (containers sizes) 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). 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, then we would conclude that consumers are purchasing for immediate consumption and not stockpiling. 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 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.

These are just two examples of the type of variation in the data that allows us to identify the parameters. Our model extends these ideas. 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 (i.e., a 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 stockpiling, based on the impact of previous purchases on current behavior (see Hendel and Nevo, 2002, for details). 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. On the one hand, the static parameters recovered in stage one, determine the substitutability across brands, namely brand choice. On the other, 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) an approximation, or a shortcut, to separate long run from short run prices responses. The basic insight is that in order to capture responses to long run price changes – 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. Monte Carlo experiments will help us assess whether this is a useful shortcut to improve demand elasticities estimates.

We discuss next the merits and limitations of the proposed approach, vis-a-vis potential alternative approaches.

4.3.1 Limitations

Three assumptions are critical to the above procedure. First, the transition of the inclusive values is assumed to depend exclusively on previous inclusive values. Second, product differentiation is modeled as taking place at the time of purchase rather than consumption. Finally, the error term is assumed to be i.i.d. extreme value. We will not expand on the latter. The implications of the logit assumption are quite well understood, moreover, the computational simplicity of the method will enable us to enrich the error structure with brand effects. We now expand on the other two limitations.

Transitions

The specification we are currently using is quite rich, it allows for the dependence of inclusive values across sizes, namely the distributions depend on previous inclusive values of all sizes. We also experiment with higher order Markov processes (see Section 5.1). It is worth mentioning the process is household specific. Since buyers that visit stores with different frequencies will potentially face different transitions. In spite of the generality, the approach limits two price vectors that yield the same vector of (current) inclusive values to have the same transition probabilities to next period state, while a more general model could allow these to be different.

This assumption is testable and to some extent it can be relaxed, should it fail in the data. 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, in case the assumption fails in the data, 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.[16] 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 all lead to the same distribution of future inclusive values. Such formulation would be a compromise, as the inclusive values would be allowed to depend on some additional information beyond the current vector of inclusive values. The idea is to draw a map of regions within the state space that generate similar transitions.

Product Differentiation

Taken literally our model assumes that differentiation occurs at the time of purchase rather than during consumption. However, we think of differentiation at purchase, represented by the ξ_{hjx} term in equation (1), as a way of capturing the expected value of the future differences in utility from consumption. This approach is valid as long as (i) brand-specific differences in the utility from

[16] 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 and many parameters can be recovered in the first stage.

consumption enter linearly in the utility function,[17] and (ii) there is no discounting. 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 (e.g., Erdem et al. ,2003). When a consumer purchases *x* units of brand *j* (with no discounting) she will obtain $x\psi_j$ units of utility from future consumption. The term ξ_{jx} captures the utility from consuming the *x* units of brand *j*, expected at the time of purchase. With discounting the previous analogy becomes less straightforward,[18] 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. So if, for example, the utility of consumption of brand *j* depends on how much brand *k* is consumed then at the moment of purchase, one cannot compute the expected utility from *x* units of brand *j*. In order to deal with this sort of utility a vector of the inventories of all brands has to be included as state variables. In our model only total inventory is relevant. This is the main advantage of capturing product differentiation at the moment of purchase. It reduces the state space. Only the total quantity held in inventory matters as a state variable. In a more general model the whole vector of inventories, the quantity of each brand held in inventory, have to be carried as state variables.

Our framework is appropriate to study purchases of detergents, where consumption interactions do not seem important. For other products, where the marginal utility from consuming Cherios may depend on the consumption of Trix (namely, interactions are more important), the state space has to be expanded, to include all inventories. This can be done as long as either the choice set is small or we can aggregate the products into segments, each of which serve an independent

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

[18] Two issues arise. First, since the timing of consumption is uncertain, the present value of the utility from consumption becomes uncertain ex-ante. Nevertheless, we can compute the expected present value of the utility *x* units of brand *j*. Second, with discounting the order in which the different brands already in storage are consumed, becomes endogenous.

process or task as in Hendel (1999).

In defense of our approach we should mention not only the computational simplicity (discussed below) but also that although restrictive, the assumption of no interaction between brands is standard in static discrete choice models.

4.3.2 Advantages

The key advantage of our approach is that the state space can be substantially reduced, and some of the preference parameters can be recovered through the estimation of a static discrete choice of brand given size choice. In our setup $EV(\cdot)$ is a function of the total inventory and a vector of inclusive values (as many as product sizes). While in the unrestricted problem $EV(\cdot)$ depends, instead, on a vector of inventories (one for each brand), on the vector of prices (one for each brand-size) and promotional activities, like feature and display (potentially of all brands and sizes.)

The second advantage is that the static preferences parameters, those that determine product differentiation, are recovered through a static estimation, described in Step 1. Since this estimation is quite simple we can allow for a rich error structure, including brand and size effects, as well as controls for advertising and special displays. Furthermore, the framework is flexible enough to accommodate possible generalizations. For example, we can allow for purchases of multiple brands on the same trip.

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 are offered by several brands. The dynamic problem would have an extremely large state, which includes the inventories of all brands held by the household as well as the price vector of all brands in all sizes; plus all other promotional activities for each product/size combination.

Erdem et. al. (2003) propose a different solution to the problem. Their solution to reduce the complexity of the problem is to assume that all brands are consumed proportionally to the quantity in storage. Together with the assumption that brand differences in quality enter linearly in the utility function this implies that only the total inventory and a quality weighted inventory matter as state variables, instead of the whole vector of brand inventories. To reduce the dimensionality of the price vector they concentrate in 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 like advertising.

As we discussed above our method considerable reduces the computational burden. There are two advantages. First, the dynamic problem, our third step, is considerably simpler and therefore can in practice be more flexible. Second, since most of the parameters of the model are estimated in the first step, which does not require solving the dynamic programming problem, we can allow for a richer model that includes, for example, observed heterogeneity (demographics) and promotional activities. Controlling for the latter seems to be particularly important since there seems to be substantial effect on elasticities (at least in our data). Moreover, advertising and low prices are correlated, hence, as we show in the results below, neglecting advertizing biases demand elasticities upward, by confounding high demand due to advertizing with high demand due to low prices. The cost of the split in the estimation is that we cannot allow for as rich unobserved heterogeneity in brand preferences as Erdem et. al.[19]

In sum, each approach has its strength and each is suitable for different applications. The split is not necessary when the number of brands is small. However, its main advantage is to collapse the utility from each quantity choice to a single number. Hence, the larger the number of brands the larger the computational gain. In the case of detergents, with a large number of brands (see Table 2) the split is necessary.

[19] We allow for unobserved heterogeneity in the dynamic estimation, which is considerably simplified by the split.

5. Results

In order to estimate the model we have to choose functional forms. The results below use $u(c_t + \nu_t) = \alpha \log(c_t + \nu_t)$ and $C(i_t) = \beta_1 i_t + \beta_2 i_t^2$. The distribution of ν_t is assumed to be log normal. 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, 17335 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.

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) conditional logit choice of brand conditional on size. This stage was estimated using choices by all households, where the choice set was restricted to products of the same size as the observed purchase. Different columns 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 (i) and (ii) (as well as (viii) and (ix)). 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. The size of the change is intuitive. It implies that the large effect on quantity sold seemingly associated with price changes are largely driven by the feature/display promotional activity. This suggests that we would want to control for the static effects of these activities as well as potential dynamic effects. Most the elasticities reported in the literature do not control for feature and display. Therefore, one has to be careful in interpreting estimates that do not properly control for promotional activities other than sales. For example, if the elasticities below see somewhat low, these in large part is driven by the effect of feature/display.

More importantly these effects allow us to demonstrate 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, in columns (iii)-(viii) we interact price and the brand dummy variables with two demographic variables: income and a dummy if the family has more than 4 people. These interactions are highly significant. The signs on the coefficients make sense. Larger families are more price sensitive and households with higher income are less price sensitive. The interactions with the brand dummy variables imply, for example, that high income households prefer Tide while larger families prefer a private label brand. Together these two variables generate roughly 20 different "types" of households.[20]

Third, columns (x) and (xi) interact the brand-dummy variables with size (either by multiplying the dummy by the size, in column (x), or by allowing a full interaction). Notice that interacting brand effects with sizes makes the benefit or 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. Notice that the effect on the price, display and feature coefficients is negligible (as can be seen by comparing to column (ii)).

Finally, we note that our most general specifications include dozens of parameters, which we are able to estimate, with essentially zero added complexity, due to our estimation algorithm. Such a large number of parameters would be essentially impossible to estimate using the standard nested algorithm.

Table 5 reports the estimates of the price process. As explained above this process was estimated using the inclusive values (given in equation (4)) computed from the estimates of column (viii) in Table 4. The inclusive values can be considered a quality weighted price: for each household

[20] A typical random effects approach will allow for only three to four types, so in principle we allow for a lot of heterogeneity. From an economic point of view the true test of our specification of heterogeneity is in the patterns of cross-price elasticities. As we discuss below our estimates perform well in this regard.

they combine all the different prices (and promotional activities) of all the products offered in each size into a single index. This index varies by household, since the brand preferences are allowed to vary. The first set of 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. The coefficients vary between 0.24 and 0.6, while the cross-size effects are smaller than 0.1 in absolute value. The process for 32 and 64 oz seems slightly more persistent, which is consistent with these sizes having less sales. Overall the fit is reasonable. Indeed if the fit was much higher one could claim that there is not much uncertainty for the household regarding future prices.

Considering the supply-side there are good reasons to believe that prices will not follow a first-order process.[21] 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 dynamics of choice of size. 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. Their implications are reasonable. Larger households have higher values of α, 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.

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

[21] 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.

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 hold slightly higher inventory 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. Furthermore, we can create an inventory series by using the assumption of constant consumption and observed purchases. If we set the initial inventory for such a series so that the inventory will be non-negative then the mean inventory is essentially the same as the inventory simulated from the model.

5.2 The Fit of the Model

In order to test the fit of the model we simulated the implications of the estimates and compared them to observed behavior. We simulated the predictions of the model using the observed data. First, we compare predictions regarding quantities and brand choices. Simulated and sample size choice probabilities are presented in Table 7. They aline almost perfectly. A similar match is also present once we look at choice of a brand conditional on size. This is not surprising since the brand choice is estimated by a conditional logit model, which includes brand fixed effects. Generally, the choice probabilities vary with the state variables as expected: the higher the inventory the lower the probability of purchase.

Ideally in order to further test the fit 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. The later represents the best one can do without considering dynamics. 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. The fit of the survival function is very good. The fit of the hazard rate is also reasonable.

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. We then 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. Next, we use the exact solution on the grid to fit continuous value and policy functions and used these to evaluate the data. The results were qualitatively the same.

Second, we explored a variety of functional forms for both the utility from consumption and for the cost of inventory. Once again the results are qualitatively similar.

5.3 Implications

In this section we present the implications of the results, and compare them to static estimates. In Table 8 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. Next, we generated simulated price changes separately for 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. 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 and used them to compute the price elasticities.

The results are presented in Table 8. 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 product 128 oz, the most popular size. The own price elasticities are between -1.5 and -2.5. While at first this might seem low we recall that these are long run elasticities, which as we will see below are lower, in absolute value, than the short run elasticities. Furthermore, their magnitude is driven by the inclusion of feature and display in the first stage. If we were to exclude these the price elasticities would be roughly double in magnitude.

The cross-price elasticities also seem reasonable.[22] There are several patterns worth pointing out. First, we note that the cross-price elasticities to other brands of the same size, 128 oz., are generally higher. This is what one would expect given the dynamic considerations introduced by inventories. The current inventory a household holds is a key determinant of whether to purchase and if so how much. Therefore, if the price of a product changes consumers currently purchasing it are more likely to substitute towards other similar size containers of different brand.

Second, we note that the cross-price elasticities to other sizes of the same brand are generally higher, sometimes over 3 times higher, than the cross-price elasticities to other brands. For example, if the price of a 128 oz container of Tide changes the substitution to other sizes of Tide is roughly 0.15, while it is roughly 0.06 to some of the other brands. This suggests that some brands have a relatively loyal base. This pattern is driven by the heterogeneity estimated in the first step. As we noted, our method does not allow us to introduce unobserved heterogeneity in this step and therefore one might worry that we cannot capture brand loyalty to generate this type of behavior. The real test of the model of whether it allows for sufficient heterogeneity, whether observed or unobserved, is in the substitution patterns it predicts. It is well-known that the fixed parameters Logit would imply that the cross-price elasticities in each column would be the same (at least for all the products of the same size). The way to relax substitution patterns is by allowing for heterogeneity in the utility from the brands. So one should look at the matrix of price elasticities to see if they differ from the

[22]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.

standard Logit prediction, which does not allow for higher substitution between similar products. Since the substitution matrix does not resemble the patterns of one implied by Logit, we think the household demographics are capturing heterogeneity in a successful way.

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 log 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 (namely, 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. It would include 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 overestimate by a factor between 2 and 3 the substitution to the outside option. This highlights the bias from static estimates that our framework overcomes.

We next turn to comparing the long run elasticities to those computed from static demand models. Table 9 present that 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. This ratio appears to be constant across brands and also across the main sizes, 64 and 128 oz. This highlight 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 but the following is the one that seems to dominate. Consider a reduction in the price of Tide. The static elasticities capture, for example, the reduction in the quantity of Cheer sold today. However, there will also be a reduction in the quantity of Cheer sold next period, and the period after, etc. These effects are captured in the dynamic elasticity and therefore the static elasticity, which account only for the customers substituting today, will under estimate the substitution to the other products, especially those of the same size.

The results in Table 9 display precisely these patterns. There are several patterns in these ratios. The ratio of the elasticities towards other brands of the same size is roughly 0.25. In contrast, the ratio of elasticities towards other sizes is in the 0.7 to 0.8 range. Finally, we note that the bias in the substitution towards the outside option is larger for 128 oz than for 64 oz.

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).

PCM computed in this way are used to test among different supply model, 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 both cases 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 stockpile 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.

Although our model has limitations and might not be correctly specified it is important to point out that our main interest resides in the ratio of static to dynamic estimates (Table 9). These ratios might not be affected even if our model is slightly misspecified ; since both components of the ratio will be affected. For example, if we were to neglect promotional activities, like feature and display, we would get higher elasticities (as we saw from the first stage estimates) in both the static and dynamic models. Yet, the ratio is largely unaffected.

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 stockpiling effects and bias the estimates.

Our approach is an alternative to the one proposed by Erdem, Imai and Keane (2003). They were the first to structurally estimate a consumer inventory model. They base their estimation on alternative simplifying assumptions which render their method better suited for markets with a smaller number of brands. In turn, their method accommodates unobserved heterogeneity while our can accommodate observed heterogeneity at a much lower computational cost. We find their results supportive of ours. Their reported results cannot be used directly to address our main focus: the difference between elasticities computed from a static model and long run elasticities computed from a dynamic model. Nevertheless, despite taking a different modeling approach and using different data they too find that stockpiling and dynamics are important.

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. We are in the process of exploring ways to exploit this potential shortcut in order to suggest approximations one might use to at least get a sense of the importance of dynamics. We believe these will be useful in making the ideas easier to use in applied and policy work.

We are in the process of extending our theoretical analysis to include the supply side. We aim to theoretically characterize optimal firm behavior in the presence of stockpiling behavior by consumers. We have several goals. First, given our estimates, we could ask what are the optimal patterns of sales. For example, at what frequency should a sale be held? Or, what is the optimal discount? Second, we could ask what proportion of the variation in prices over time can be explained by firms' attempts to exploit heterogeneity in storage costs, as apposed to other reasons for conducting a sale. Finally, in the analysis above we focused on the effects of stockpiling on demand elasticities and the implications these have on policy assuming a static Bertrand pricing game. With a better specified supply model we could address questions like what effects would mergers have on the *distribution* of prices and not just the average price.

Appendix

We provide the proof of the claim made in Section 4, 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

$$\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt} + u(c_j^* + \nu_t) - C(i_{t-1} + x_t - c_j^*) + \delta EV(\Omega_t; d_{jt} = 1, x_t, c_j^*) >$$

$$\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt} + u(c_k^* + \nu_t) - C(i_{t-1} + x_t - c_k^*) + \delta EV(\Omega_t; d_{jt} = 1, x_t, c_k^*)$$

and therefore

$$u(c_j^* + \nu_t) - u(c_k^* + \nu_t) > \delta EV(\Omega_t; d_{jt} = 1, x_t, c_k^*) - \delta EV(\Omega_t; d_{jt} = 1, x_t, c_j^*) + C(i_{t-1} + x_t - c_k^*) - C(i_{t-1} + x_t - c_j^*)$$

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

$$u(c_j^* + \nu_t) - u(c_k^* + \nu_t) < \delta EV(\Omega_t; d_{jt} = 1, x_t, c_k^*) - \delta EV(\Omega_t; d_{jt} = 1, x_t, c_j^*) + C(i_{t-1} + x_t - c_k^*) - C(i_{t-1} + x_t - c_j^*),$$

which is a contradiction.Q

References

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

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. 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.

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..

Gonul, F., and K. Srinivasan (1996), "Estimating the Impact of Consumer Expectations of Coupons on Purchase Behavior: A Dynamic Structural Model." *Marketing Science*, 15(3), 262-79.

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.

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)	(xi)
price	-0.95	-0.45	-0.40	-0.42	-0.61	-0.57	-0.55	-0.53	-1.08	-0.45	-0.45
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
* large family			-0.43	-0.25			-0.42	-0.21	-0.27		
			(0.04)	(0.04)			(0.05)	(0.05)	(0.05)		
*income					0.36	0.26	0.35	0.24	0.37		
					(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
feature		0.91	0.91	0.92	0.91	0.90	0.91	0.91		0.91	0.89
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		(0.05)	(0.05)
display		1.24	1.24	1.24	1.23	1.24	1.23	1.25		1.25	1.24
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		(0.04)	(0.04)
brand dummy var	✓	✓	✓	✓	✓	✓	✓	✓	✓		
* large family				✓				✓	✓		
*income						✓		✓	✓		
*size										✓	
brand-size dummy var											✓

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

	ω_{1t}	ω_{2t}	ω_{3t}	ω_{4t}	ω_{1t}	ω_{2t}	ω_{3t}	ω_{4t}
$\omega_{1,t-1}$	.60 (.004)	-.02 (.011)	.08 (.014)	-.04 (.015)	.47 (.004)	-.05 (.012)	.05 (.02)	.02 (.02)
$\omega_{2,t-1}$	-.02 (.001)	.31 (.004)	.004 (.005)	.07 (.006)	-.007 (.001)	.28 (.004)	-02 (.005)	.04 (.006)
$\omega_{3,t-1}$	.02 (.001)	-.02 (.004)	0.47 (.02)	-.03 (.006)	.01 (.002)	.01 (.004)	0.35 (.005)	.01 (.006)
$\omega_{4,t-1}$	-.01 (.001)	.02 (.003)	-.03 (.004)	.24 (.004)	.001 (.001)	.01 (.003)	-.01 (.003)	.22 (.04)
$\sum_{\tau=2}^{5} \omega_{1,t-\tau}$					.06 (.001)	-.03 (.003)	-.01 (.004)	-.01 (.005)
$\sum_{\tau=2}^{5} \omega_{2,t-\tau}$					-.02 (.001)	.03 (.001)	-.02 (.002)	.03 (.002)
$\sum_{\tau=2}^{5} \omega_{3,t-\tau}$					.001 (.001)	-.01 (.002)	.07 (.002)	-.01 (.002)
$\sum_{\tau=2}^{5} \omega_{4,t-\tau}$					-.004 (.001)	.01 (.001)	-.02 (.001)	.02 (.002)
R-squared	.84	.73	.65	.66	.85	.74	.67	.67

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 (viii) of Table 4.

Table 6
Third Step: Estimates from the Nested DP Problem

household type:	1	2	3	4	5	6
coefficient on:						
Cost of inv - linear	3.78	2.90	1.69	3.02	1.59	0.88
	(.49)	(.57)	(.40)	(.59)	(.69)	(.10)
Cost of inv - quadratic	5.07	4.37	4.19	4.95	3.61	2.19
	(.64)	(1.15)	(2.84)	(.64)	(2.52)	(1.35)
Utility from consumption	0.59	1.13	1.26	0.65	1.10	1.54
	(.76)	(.60)	(.91)	(.43)	(1.15)	(.74)
Log likelihood	-10900.1					

Asymptotic standard errors in parentheses. Also included 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
Sample and Simulated Choice Probabilities
(percent)

	Data	Simulation
No purchase	78.24	77.77
32 oz.	0.46	0.52
64 oz.	10.74	10.80
96 oz.	1.64	1.84
128 oz.	8.92	9.07

Table 8
Long Run Own and Cross-Price Elasticities

#	Brand	Size (oz.)	All*	Wisk	Surf	Cheer	Tide	Solo	Era
1	All*	32	0.063	0.031	0.071	0.119	0.050	0.045	0.080
2		64	0.102	0.070	0.078	0.065	0.050	0.082	0.048
3		96	0.084	0.086	0.084	0.083	0.061	0.107	0.053
4		128	-1.425	0.219	0.204	0.223	0.147	0.205	0.162
5	Wisk	32	0.041	0.136	0.097	0.029	0.059	0.116	0.066
6		64	0.045	0.186	0.086	0.037	0.062	0.098	0.068
7		96	0.045	0.202	0.108	0.052	0.068	0.128	0.087
8		128	0.106	-2.274	0.230	0.108	0.160	0.214	0.195
9	Surf	32	0.035	0.034	0.120	0.033	0.065	0.128	0.062
10		64	0.046	0.069	0.252	0.043	0.058	0.096	0.053
11		96	0.045	0.064	0.220	0.050	0.067	0.125	0.064
12		128	0.081	0.183	-2.153	0.115	0.157	0.243	0.157
13	Cheer	64	0.055	0.064	0.099	0.132	0.080	0.113	0.074
14		96	0.062	0.034	0.094	0.088	0.095	0.079	0.036
15		128	0.127	0.197	0.248	-2.387	0.214	0.231	0.213
16	Tide	32	0.033	0.056	0.093	0.044	0.115	0.098	0.088
17		64	0.033	0.063	0.069	0.046	0.154	0.090	0.078
18		96	0.035	0.079	0.095	0.052	0.152	0.142	0.127
19		128	0.080	0.173	0.190	0.133	-2.316	0.204	0.244
20	Solo	64	0.029	0.058	0.068	0.037	0.048	0.220	0.046
21		96	0.037	0.060	0.064	0.027	0.059	0.212	0.052
22		128	0.082	0.162	0.191	0.101	0.141	-1.501	0.179
23	Era	32	0.024	0.057	0.045	0.026	0.076	0.076	0.096
24		64	0.027	0.066	0.057	0.035	0.076	0.080	0.199
25		96	0.034	0.076	0.059	0.053	0.086	0.087	0.162
26		128	0.070	0.189	0.153	0.101	0.209	0.235	-1.862
27	No purchase		0.027	0.066	0.085	0.042	0.056	0.079	0.076

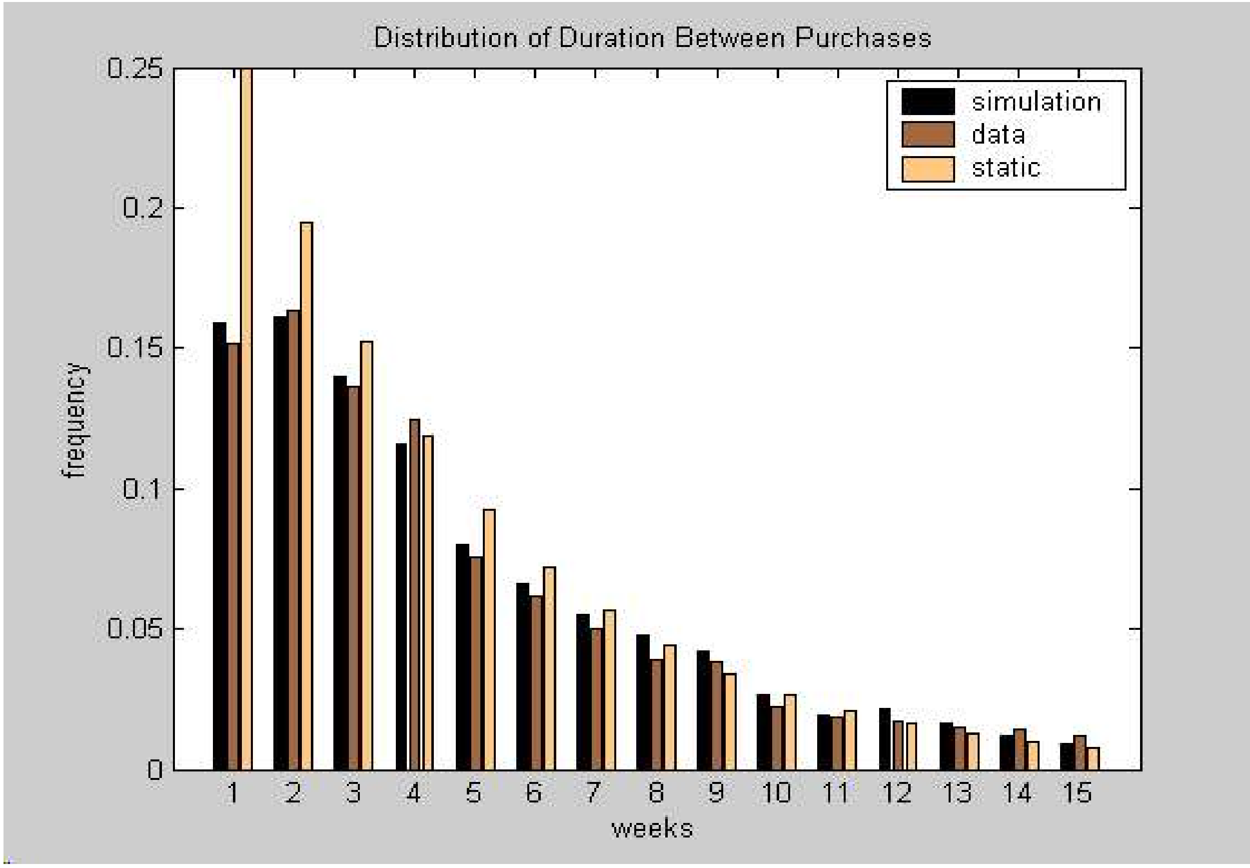
Cell entries *i*, *j*, where *i* indexes row and *j* column, give the percent change in market share of brand *i* with a one percent change in price of *j*. All columns are for a product 128 oz, the most popular size. Based on the results of Tables 4-6.
(*) Note that "All" is a name of a detergent produced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

Table 9
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	Solo	Era	All*	Wisk	Surf	Cheer	Tide	Solo	Era
1	All*	64	1.29	0.26	0.25	0.25	0.24	0.27	0.24	0.79	0.77	0.76	0.71	0.72	0.85	0.72
2		128	0.67	0.71	0.69	0.67	0.66	0.75	0.66	1.32	0.22	0.22	0.21	0.21	0.25	0.21
3	Wisk	64	0.24	1.30	0.25	0.25	0.24	0.27	0.24	0.79	0.77	0.76	0.71	0.72	0.85	0.72
4		128	0.68	0.72	0.70	0.67	0.66	0.75	0.66	0.23	1.31	0.22	0.21	0.20	0.25	0.21
5	Surf	64	0.24	0.25	1.29	0.24	0.23	0.27	0.23	0.78	0.75	0.74	0.70	0.70	0.83	0.70
6		128	0.66	0.70	0.68	0.66	0.64	0.73	0.64	0.23	0.22	1.30	0.20	0.20	0.24	0.20
7	Cheer	64	0.25	0.26	0.26	1.28	0.24	0.28	0.24	0.80	0.78	0.76	0.72	0.73	0.87	0.73
8		128	0.67	0.72	0.70	0.69	0.66	0.75	0.66	0.24	0.22	0.22	1.29	0.21	0.25	0.22
9	Tide	64	0.25	0.27	0.26	0.26	1.29	0.28	0.25	0.79	0.77	0.75	0.72	0.73	0.86	0.73
10		128	0.68	0.72	0.70	0.68	0.67	0.76	0.67	0.23	0.22	0.22	0.21	1.31	0.25	0.21
11	Solo	64	0.24	0.26	0.25	0.24	0.23	1.27	0.24	0.78	0.77	0.75	0.71	0.72	0.85	0.72
12		128	0.68	0.72	0.70	0.67	0.67	0.76	0.67	0.23	0.22	0.21	0.21	0.20	1.29	0.21
13	Era	64	0.25	0.26	0.26	0.26	0.24	0.28	1.28	0.79	0.77	0.76	0.72	0.72	0.86	0.73
14		128	0.67	0.71	0.69	0.68	0.65	0.74	0.65	0.23	0.22	0.22	0.21	0.21	0.25	1.29
15	No purchase		1.45	1.56	1.49	1.42	1.42	1.59	1.42	2.77	2.50	2.41	2.32	2.41	2.75	2.38

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