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

July 2002

ABSTRACT

Temporary price reductions (sales) are common for many goods and naturally result in large increase in the quantity sold. In previous work we found that the data support the hypothesis that these increases are, at least partly, due to dynamic consumer behavior: at low prices consumers stockpile for future consumption. In this paper we quantify the magnitude of the effect and derive the quantitative 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 overestimate own price elasticities by up to 50-80 percent.

1

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. This behavior 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. Our goal is to structurally estimate the model in order to study the economic implications of stockpiling behavior.

Stockpiling has several economic implications. First, when the good in question is storable, there is a distinction between the short run and long run reactions to a price change. Short run reactions reflect both consumption and stockpiling effects. In contrast, for most demand applications (e.g., merger analysis or computation of welfare gains from introduction of new goods) we want to measure long run responses. A simple back of the envelope calculation presented in Hendel and Nevo (2002) shows that neglecting dynamics may overstate price sensitiveness by a factor of 2 in detergents (and up to 6 times in other products.) Second, even if we are interested in short run price responses, static demand estimation yields biased estimates. Third, stockpiling has implications for how sales should be treated in the consumer price index. If consumers stockpile, then ignoring the fact that they can substitute over time will yield a bias similar to the bias generated by ignoring substitution between goods as relative prices change (Feenstra and Shapiro, 2001). A final motivation to study stockpiling behavior, is to understand sellers' pricing incentives when products are storable.

Before we proceed we quantify the potential gains from dynamic behavior for the type of products we have in mind. To do so we compare the actual amount paid in the data to what would have been paid if the price was drawn at random from the distribution of prices for the same product

at the same location over time. This is a lower bound on the potential gross gains from optimizing behavior since we are keeping the purchased bundle constant. 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.[2] 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). We find this number encouraging, especially since it is a lower bound on the potential gains from optimization.

In a previous paper (Hendel and Nevo, 2002) we documented the following facts, which are consistent with stockpiling behavior. First, aggregate demand is a function of the duration from previous sale, both during sale and non-sale periods.[3] Second, household proneness to buy on sale is correlated with variables that proxy for storage costs. Third, both for a given household over time, and across households, we find a significant difference, between sale and non-sale purchases, in duration to next purchase. Fourth, 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. Fifth, we find that a proxy for inventory is negatively correlated with the quantity purchased and with the probability of buying.

All the finding described in the previous paragraph are consistent with the predictions of a stockpiling model. However, 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 directly address the economic implications detailed above.

We address these implications here by considering a consumer's dynamic problem when she

---

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

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

has an expected stream of future demands, is able to store a consumption good and faces uncertain future prices. In each period the consumer 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 subject to the constraint that the quantity consumed does not exceed purchases plus inventory. Optimal behavior is a function of current prices, the current inventory and stochastic shocks. In this model the consumer will 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.

The structural estimation basically follows the "nested algorithm" proposed by Rust (1987). We have to make two adjustments. First, inventory, one of our state variables, is unobserved (by us) and is endogenously determined. 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 potentially 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 will not depend (in our model) on the dynamic elements. Therefore, we can consistently estimate many of the parameters of the model without solving the full 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 allows us to estimate a very general model, which allows for a large degree of consumer heterogeneity and nests standard static choice models. We discuss below the assumptions necessary to validate this procedure, which we believe are natural for the product in question. The costs of this procedure is a loss in efficiency, and loss of generality in defining the transition probabilities.

Preliminary results suggest that ignoring the dynamics can have strong implications for demand estimates.

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. (2000) 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 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.[4]

The closet paper to ours is Erdem et. al. (2002). 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 take a different approach than ours. They simplify the state space by assuming that all brands are consumed proportionally to the quantity in storage. Which together

---

[4]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.

with the assumption that brand differences in quality enter linearly in the utility function imply that only the total inventory and a quality weighted inventory matter as state variables, instead of the whole vector of brand inventories. Their focus in on the role of price expectations. Hence, they estimate a different price process than we do. They compare consumers responses' to price cuts, both allowing for the price cut to affect future prices expectations, and holding expectations fixed. Interestingly, they are able to separate the price and the expectation effect, of a sale, on total sales.

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.

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 1. The top 11 brands account for roughly 90 percent of the quantity sold.

6

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,[5] 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.[6] 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-

---

[5]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.

[6]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.[7] 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 phenomena. 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. 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 given household over time, and across households, we find a significant difference, between sale and non-

---

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

sale purchases, in both duration from previous purchase and duration to next purchase. The duration effects are a consequence of the dependence of the trigger and target inventory levels on current prices. In order to take advantage of the low price, during a sale a household will buy at higher levels of current inventory. Furthermore, during a sale a household will buy more and therefore, on average, it will take more time till the next time 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 model.

The findings suggest that in the presence of stockpiling standard , static, demand estimation may be misleading. The results can give an approximate idea of the relevant order of magnitude of the impact of stockpiling on demand estimates. We find that static demand estimates that neglect dynamics may overestimate price sensitiveness by up to a factor of 2 to 6. Our goal below is to get precise estimates of the magnitude of these effects.

In addition to supporting our theory we have learned the following points that are relevant for the modeling below. First, even for laundry detergents there seems to be a consumption effect. When purchasing on sale the quantity purchased increases more than the duration to next sale. Second, controlling for feature and display (not only current, but also lagged values) is important.

3. The Model

3.1 The Basic Setup

We consider a model in which a consumer obtains the following per period utility

$$u(c_{it}, v_{it}; \theta_i) + \alpha_i m_{it}$$

where c_{it} is the quantity consumed of the good in question, v_{it} is a shock to utility that changes the current marginal utility from consumption, θ_i is a vector of consumer-specific taste parameters, m_{it}

is utility from the outside good, which is multiplied by the marginal utility of income, α_i. The stochastic shock, v_{it}, introduces randomness in the consumer's needs, unobserved to the researcher. For simplicity we assume the shock to utility is additive in consumption, $u(c_{it}, v_{it}; \theta_i) = u(c_{it} + v_{it}; \theta_i)$. Low realizations of v_{it} increase the household's need, increasing demand and making it more inelastic. The product is offered in J different varieties, or brands. The consumer faces random and potentially non-linear prices. We denote by p_{jxt} the total price associated with purchasing x units (or size x) of brand j.

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.[8] Quantity not consumed is stored as inventory. In the estimation we assume that the purchase amount, denoted by x_{it}, 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 $d_{ijxt} = 1$, where $x = 0$ stands for no purchase, and we assume $\sum_{j,x} d_{ijxt} = 1$.[9] Dropping the subscript i, to simply notation, the consumer's problem can be represented as

$$V(I_0) = \max_{\{c_t, d_{jxt}\}} \sum_{t=0}^{\infty} \delta^t E\left[u(c_t, v_t; \theta) - C(i_t) + \sum_j d_{jxt}\left(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt}\right) | I_0\right]$$

$$s.t. \quad 0 \le i_t, \quad 0 \le c_t, \quad 0 \le x_t, \quad \sum_{j,x} d_{jxt} = 1 \tag{1}$$

$$i_t = i_{t-1} + x_t - c_t$$

where " is the marginal utility from income, I_t denotes the information at time t, $\delta > 0$ is the discount factor, $C(i_t)$ is the cost of storing inventory, ξ_{jx} is a fixed taste of brand j that could be a function of brand characteristics, size and could vary by consumer, βa_{jxt}, the effect of advertising variables on the consumer choice, and ϵ_{jxt}, is a random shock that impacts the consumer's choice. Notice, the latter is size specific, namely, different sizes get different draws introducing randomness

---

[8]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) show consumption effects are consistent with the data.

[9]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.

in the size choice.

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 unpredictable future prices. We assume the consumer knows the current shock to utility from consumption, ν_t, which are independently distributed over time. Prices are (exogenously) set according to a first-order Markov process, which we describe in Section 4.[10] 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.

3.2 Model with Inter-visit Consumption

The model presented in the previous section assumes that consumers visit the store every period. In the data we observe variation in the time between store visits. Such observed variability in the data can be exploited in the estimation, hence we want to model it. This variation impacts the previous model in two ways. First, random visits create variability in the evolution of inventories and an extra precautionary incentive to hold inventories. Second, consumption should vary with the duration between visits. In principle this could be handled by allowing the distribution of the consumption shock, ν_t, to depend on the duration for previous visit. This approach does not account for the effect of duration to next visit on the expected distribution of prices in the next visit. Therefore, we propose the following variation of the model.

In periods when the consumer visits the store his behavior is described by the above model. In each period there is a probability, q, that he will visit the store next period. If he does not visit the store he only chooses consumption and does so as to maximize the current utility, minus

---

[10]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.

inventory cost, plus future gains, subject to the same constraints as before. As before let the value function in periods of store visits be $V(I_t)$, and the value function during non-visit periods be $W(I_t)$.[11] The optimal behavior can be characterized by the following Bellman equations

$$V(I_t) = \max_{\{c_t, d_{jxt}\}} u(c_t + v_t) - C(i_t) + \sum_{j,x} d_{jxt}\left(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \epsilon_{jxt}\right) + \delta E\left[qV(I_{t+1}) + (1-q)W(I_{t+1})|I_t\right]$$

$$W(I_t) = \max_{\{c_t\}} u(c_t + v_t) - C(i_t) + \delta E\left[qV(I_{t+1}) + (1-q)W(I_{t+1})|I_t\right].$$

We can, in principle, allow the probability of a visit in the next period, q, to depend on consumer characteristics and to let it vary between visit and non-visit periods.

4. Econometrics

The structural estimation is based on the nested algorithm proposed by Rust (1987), but has to deal with issues unique 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

We base our estimation on the nested algorithm proposed by Rust (1987). The procedure is 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 . However, since we do not observe the random shocks, which are part of the 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

---

[11] We abuse notation here since the information in each period is different. In store visit periods it includes the random shocks, ϵ_{jxt}, while in non-visit periods it does not.

inventory since both the initial inventory and consumption decisions are unknown. We deal with the unknown inventories using the model to derive the optimal consumption[12] in the following way. Assume for a second that the initial inventory is observed. Therefore, we can use the procedure described in the previous paragraph to obtain not only the likelihood of the observed purchases, but also the probability of different consumption levels, and therefore the likelihood of different 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 observed purchase, but now we account for the distribution of the inventory level when computing the likelihood. We can continue this procedure to obtain the likelihood of observing the sequence of purchases for each household. In order to start this procedure we need a value for the initial inventory. We experiment with drawing this value from different distributions and with using part of the data in order to simulate the initial distribution of inventory.

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

$$Pr(d_1...d_T|p_1...p_T) = \int \prod_{t=1}^{T} Pr\Big(d_t|p_T, i_{t-1}(d_{t-1},...,d_1, v_{t-1},...v_1, i_0), v_t\Big) dF(v_1,...v_T) dF(i_0). \qquad (2)$$

Note that the beginning-of-period inventory is a function of previous decisions, the previous consumption shocks and the initial inventory. The above implicitly incorporates the first order Markov assumption on prices and the independence (over time) assumptions on v and ϵ. Given the assumption that ϵ_{jxt} follows an i.i.d. extreme value distribution,

$$Pr(d_j|p_t, i_{t-1}, v_t) = \frac{\exp\Big(\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + \max_{c_t}\big\{u(c_t + v_t) - C(i_t) + \delta EV(I_{t+1}; d_t, c_t, p_t)\big\}\Big)}{\sum_{k,x} \exp\Big(\alpha p_{kxt} + \xi_{kx} + \beta a_{kxt} + \max_{c_t}\big\{u(c_t + v_t) - C(i_t) + \delta EV(I_{t+1}; d_t, c_t, p_t)\big\}\Big)} \qquad (3)$$

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-

---

[12] 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).

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 <u>conditional</u> on the size (quantity) bought in order to recover the marginal utility of income, $"$, and the parameters that measure the effect of advertising, $\$$ and $>$'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 the a simplified problem in order to estimate the rest of the parameters. Rather than having the state space include prices of all available brand-size combinations, it includes only a single "price" for each size. For our data set 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 to estimate the rest of the 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.[13]

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 this is justified by the structure 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 computational simpler approach. We can write

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

In general, this does not help us since we need to solve the consumer's dynamic programming problem in order to compute $Pr(d_{jt} = 1 | p_t, x_t, i_{t-1}, v_t)$. However, given the primitives of our model, 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., $EV(I_{t+1}; d_{jt} = 1, x_t, c_t) = EV(I_{t+1}; x_t, c_t)$, the term $\max_{c_t} \{ u(c_t + v_t) - c(i_t) + \delta EV(I_{t+1}; d_{jt} = 1, x_t, c_t) \}$ is independent of brand choice, thus

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

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.

---

[13]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.

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 for 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

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

can be thought of as a quality adjusted price index for all brands in that size category. All the information needed, in the second step, 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 these values will differ by consumer.

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 nine brands times five sizes (roughly the dimensions in our data), we only have to follow five 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 $N(@@$ 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 discussed above 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

$$V\left(i_{t-1}, p_t, \epsilon_t, v_t\right) =$$

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

Note that since both ϵ and v are identically and independently distributed, and the choice of brand j does not affect future utility, then we can write $E\left[V\left(i_t, p_{t+1}, \epsilon_{t+1}, v_{t+1}\right) | i_{t-1}, p_t, c_t, x_t\right]$. Assuming consumption, c_t, and purchase, x_t, are chosen optimally, we can be write this expectation as a function of only current prices and beginning of period inventory, denoted by $EV(i_{t-1}, p_t)$.

Using the independence of ϵ, v and p

$$EV\left(i_{t-1}, p_t\right) =$$

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

By Lemma 1, provided in the Appendix, optimal consumption is the same for every size, x, (regardless of the brand chosen) then $Max_{c_t} u(c_t + v_t) - C(i_t) + \delta EV\left(i_t, p_{t+1}\right)$ depends on size, x, but is independent of choice of brand j. Therefore,

$$EV\left(i_{t-1}, p_t\right) =$$

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

17

which is equal to (McFadden, 1981)

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

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

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

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-1} instead of (p_t and i_{t-1}).

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

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

It is this probability that we use to construct a likelihood function in order to consistently 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. We solve it in three different ways. First, by mapping each point in the continuous state space into a point in a smaller finite element state, and then solving the problem (using value function iteration) on this smaller space. The likelihood is then computed by mapping each observed point into our grid. Second, we again solve the dynamic programming problem on the discrete grid. However, we now approximate the value function in other points using the values computed on the grid. Third, we use approximate policy function iteration. We follow the algorithm described in Benitez-Silva et al (2000).[14] Our approximation base consists of polynomials in the logarithm of inventory and in the values of the other state variables. We also hope in the future, once we allow for more heterogeneity in the

---

[14]See also Keane and Wolpin (1994), Rust (1996, 1997) and Bertsekas and Tsitsiklis (1996).

18

dynamic parameters, to use the methods proposed by Ackerberg (2000) to reduce the number of times needed to solve the dynamic programming problem.

4.3 Discussion

The way the above estimation procedure separates the estimation of quantity from 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 are also experimenting with higher order Markov processes, allowing second lags. It is worth mentioning the process can be 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. 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_{jx} 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

consumption are linear (in consumption),[15] 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 R_j is a taste parameter (e.g., Erdem et al., 2002). 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,[16] 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, 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. Consequently the framework is more appropriate to study purchases of detergents, where consumption interaction less important. For other products, where we believe consumption interactions are more important, the state space can be expanded, to include all inventories, as long as either the choice set is small or we can aggregate the products into segments, each of which serve an independent 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, this specification is not more restrictive than standard models of product differentiation. The assumption of no interaction between brands is standard in static discrete choice models.

---

[15]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 to enter linearly.

[16]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.

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.

5. Results

In order to estimate the model we have to choose functional forms. The results below use $u(c_t + v_t) = \alpha \log(c_t + v_t)$ and $C(i_t) = \beta_1 i_t + \beta_2 i_t^2$. 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.

5.1 Parameter Estimates

The parameter estimates are presented in Table 4. The first set of columns 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. It also includes brand fixed effects, which

are not reported. As expected, the price coefficient is negative. Below we present the implied price elasticities. Also as expected, the coefficients on *feature* and *display* are positive and significant.

The second set of columns report 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 the first stage. The price process was estimated using store-level data. The estimated process fits the data reasonably well for sizes 1 and 3. However, the fit is much worse for the two most popular sizes. This could be driven by several factors. First, the first-order Markov assumption might not be reasonable for these sizes. Indeed additional lags of the inclusive values are statistically significant and somewhat improve the fit. Second, while a first-order Markov process might fit the brand-level prices reasonably well, it could be that once we create the inclusive values we need a more elaborate process. We are in the midst of exploring this possibility.

The final set of columns reports the results from the third stage dynamics of choice of size. This stage was estimated (for now) using a sub sample of 64 households, 5487 observations, where an observation is a visit to the store. It also includes size fixed effects, which are not reported. The households were selected based on two criteria: (i) they made at least 16 purchases and (ii) purchased at no more than a third of their visits to the store. The parameters are statistically significant. To get an idea of their magnitude consider the following. If the beginning of period inventory is 150 ounces then buying a 128 ounce bottle increases the storage cost, relative to buying a 32 ounce bottle, by roughly $0.26. This is approximately a quarter of the saving from non-linear pricing.

For this sample the estimated mean inventory held is 229 oz. (with a 5th percentile of 116 oz. and a 95th percentile of 290 oz.). The mean consumption is 24 oz. (with a 5th percentile of 8.7 oz. and a 95th percentile of 81 oz.). This suggests that these households basically keep on reserve an extra bottle of 128 oz. Once they open this bottle they buy another. 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 compare the simulated probabilities to the observed purchases. Simulated and sample choice probabilities are presented in Table 5. They aline almost perfectly. A similar match is also present once we look at choice of a brand conditional on size. Generally, the choice probabilities vary with the state variables as expected: the higher the inventory the higher the probability of no purchase.

To further present the fit of the data we examine, in Figures 1 through 3, how the model predicts the inter-purchase duration. Figure 1 displays the distribution of the duration between the purchases. In addition to the simulation from the model and the empirical distribution, we also present the distribution predicted by a model of constant probability of no purchase set to 0.75. Overall our model traces the empirical distribution quite closely. The same is generally true also in Figures 2 and 3, where we display the survival function and hazard rate.

We tested the robustness of the results in several ways. First, we explored a variety of methods to solve the dynamic programming problem. We list these in Section 4. 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, relative to static models. In Table 6 we present the own price elasticity of a 64 oz bottle of liquid Tide. The long run own-price elasticity computed from the dynamic model, and averaged across the observed sample, is -1.66. This is the model's prediction of the change in the quantity of 64 oz Tide sold if we change the price of this product forever. We changed all prices by a fixed percent, i.e., keeping the same promotion schedule. Alternatively, in the future we plan to experiment with changing the timing of sales.

For comparison Table 6 also presents elasticities from several static models. By the nature of these models these are short run elasticities. We examine four models. All are static conditional Logit models. The first three are estimated using the above household data. They differ only in the way price enters. In Model 1 we enter the price of the product directly. This will bias the price coefficient since the model does not know that some options are larger than others. Model 2 includes price per ounce instead of price, while model 3 includes price and a size dummy variable. Model 4 is estimated with the aggregate data, looking at each store in each week as a "market", aggregating across sizes and using the price per ounce. The elasticities are evaluated for purchases at the purchase price.

The results from Table 6 display the significance of the dynamics. The average (short run) own-price elasticities from the static models are significantly higher, in absolute value, than the (long run) elasticity from the dynamic model. The bottom rows of the table present the distribution of the bias, computed as (static model elasticity)/dynamic model elasticity. They show a wide range.

In order to get an idea of the importance of cross price elasticities we repeat the above exercise, but instead of focusing on a single product we examine all the Procter and Gamble products. Instead of presenting a (firm wide) price elasticity we compute the price cost margins (PCM) implied by static Bertrand pricing.[17] We compare this figure to what is implied by the static models. We present results only for Models 2-4. As before the differences are substantial.

6. Conclusions and Extensions

In this paper we structurally estimate a model of household inventory holding. Our estimation procedure allows us to introduce features into the model which are needed in order to fit the data.. The estimates suggest that ignoring consumer inventory can have strong implications on demand estimates and any policy based on demand estimates.

We are currently in the process of estimating more flexible versions of the model. In

---

[17]We are by no means claiming that this type of pricing is optimal from the firm's perspective in the dynamic setup. We use it here for descriptive purposes only.

particular our current focus is on two elements. First, we would like to allow for more heterogeneity across households. The results in Hendel and Nevo (2002) suggest that not only is there a lot of heterogeneity across households, but that it is important to control for these differences by more than just random effects models. Our econometric method allows us to do this. We plan to exploit the richness of the data in order to include heterogeneity in all three stages of the estimation. Second, we are in te process of further refining the price process we use.

We also plan to further study the implications of dynamic behavior on demand estimation. Our ultimate goal is not just to measure the bias from using static models, but hopefully also suggest simple ways that one could approximate the effects of the full dynamic model.

We are also currently in the process of extending our theoretical analysis to include the supply side. This, jointly with the structural estimates, will allow us to examine questions like what proportion of the variation in sales can be explained by our estimates, and given our estimates what are the optimal patterns of sales.

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 $c_k^*(x_t, v_t)$ be the optimal consumption conditional on a realization of v_t and purchase of size x_t of brand k..

Lemma 1: $c_j^*(x_t, v_t) = c_k^*(x_t, v_t)$.

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

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

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

and therefore

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

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

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

which is a contradiction.□

References

Ackerberg, D. (2000), "Importance Sampling and the Simulated Method of Moments," Boston University, mimeo.

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 Continous-State Dynamic Programming Problems," Yale University, mimeo.

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.

Chevalier, J., A. Kashyap and P. Rossi (2000), "Why Don't Prices Rise During Peak Demand Periods? Evidence from Scanner Data," NBER Working Paper no. 7981.

Erdem, T., M. Keane and S. Imai (2002), "Consumer Price and Promotion Expectations: Capturing Consumer Brand and Quantity Choice Dynamics under Price Uncertainty," University of California at Berkeley, mimeo.

Feenstra, R. and M. Shapiro (2001), "High Frequency Substitution and the Measurement of Price Indexes," NBER Working Paper no. 8176.

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.

Keane, M. and K. Wolpin (1994), "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76(4), 648-72.

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

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.

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	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	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	21.4	32.5	Tide	P & G	40.0	40.0	25.1
2	All	Unilever	15.0	36.4	47.4	Cheer	P & G	14.7	54.7	9.2
3	Wisk	Unilever	11.5	47.9	50.2	A & H	C & D	10.5	65.2	28.0
4	Solo	P & G	10.1	58.0	7.2	Dutch	Dial	5.3	70.5	37.6
5	Purex	Dial	9.0	67.0	63.1	Wisk	Unilever	3.7	74.2	41.2
6	Cheer	P & G	4.6	71.6	23.6	Oxydol	P & G	3.6	77.8	59.3
7	A & H	C & D	4.5	76.1	21.5	Surf	Unilever	3.2	81.0	11.6
8	Ajax	Colgate	4.4	80.5	59.4	All	Unilever	2.3	83.3	
9	Yes	Dow Chemical	4.1	84.6	33.1	Dreft	P & G	2.2	85.5	15.2
10	Surf	Unilever	4.0	88.6	42.5	Gain	P & G	1.9	87.4	16.7
11	Era	P & G	3.7	92.3	40.5	Bold	P & G	1.6	89.0	1.1
12	Generic	–	3.5	95.8	7.8	Generic	–	3.6	92.6	23.8
13	Other	–	4.1	99.9	9.5	Other	–	7.4	100.0	16.0

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
Parameter Estimates

1st Stage: Brand Choice		2nd Stage: Price Process						3rd Stage: Dynamics	
	Pr(j|x)			T_1	T_2	T_3	T_4		Pr(x)
price	-.45	size 1		.80	-.13	.30	-.08	cost of inv -	1.05
	(.02)			(.02)	(.06)	(.08)	(.09)	linear	(0.29)
feature	.91	size 2		-.03	.29	-.05	.04	cost of inv -	.48
	(.05)			(.02)	(.04)	(.05)	(.06)	quadratic	(0.27)
display	1.24	size 3		.02	-.04	0.81	-.04	utility from	1.06
	(.04)			(.01)	(.02)	(.02)	(.03)	consumption	(0.34)
		size 4		.01	.03	-.01	.17		
				(.01)	(.02)	(.03)	(.04)		
		constant		0.43	2.60	-.08	1.44		
				(.08)	(.19)	(.27)	(.30)		
(pseudo) R^2	.17			.87	.22	.75	.16		
log likelihood	-16394								-4229

Asymptotic standard errors in parentheses. The 1st stage was estimated using all households, 122081 observations. It also includes brand fixed effects, which are not reported. The price process was estimated using store-level data. The 3rd stage was estimated (for now) using a sub sample of 64 households, 5487 observations. It also includes size fixed effects, which are not reported.

Table 5
Sample and Simulated Choice Probabilities
(percent)

	Data	Simulation
No purchase	75.31	74.61
32 oz.	0.26	0.26
64 oz.	13.30	13.79
96 oz.	1.88	1.81
128 oz.	9.26	9.54

Table 6
SR vs. LR elasticities: The Own-Price Elasticity of 64 oz. Tide

	Household data			Aggregate data
	Model 1	Model 2	Model 3	Model 4
Average Own-Price Elasticity (LR, dynamic model == -1.66)	-0.44	-2.53	-2.14	-2.57
Percent Bias: [a]				
Average	26.7	151.7	128.3	160.1
Median	25.9	147.2	124.4	150.2
5 percentile	25.1	142.9	120.9	114.7
95 percentile	31.2	177.8	150.2	232.7

All static models are conditional logit models. Models 1-3 are estimated using the household data, where model 1 includes the price, model 2 includes price per ounce, model 3 includes price and a size dummy variable. Model 4 is estimated with the aggregate data, looking at each store in each week as a "market", aggregating across sizes and using the price per ounce. The elasticities are evaluated for purchases at the purchase price.
[a] Computed as (static model elasticity)/dynamic model elasticity.

Table 7
SR vs. LR elasticities: P&G brands implied PCM
(assuming multi-product <u>static</u> Bertrand pricing)

	Model 2	Model 3	Aggregate data Model 4
Average PCM (LR, dynamic model == 70.6%)	48.0	43.3	50.5
Percent Bias: [a]			
Average	53.4	80.9	53.6
Median	34.1	54.4	36.6
5 percentile	-18.8	-23.1	-32.6
95 percentile	188.2	253.8	186.8

All static models are conditional logit models. Models 2 and 3 are estimated using the household data, where model 2 includes price per ounce, model 3 includes price and a size dummy variable. Model 4 is estimated with the aggregate data, looking at each store in each week as a "market", aggregating across sizes and using the price per ounce. The PCM are evaluated for purchases at the purchase price.
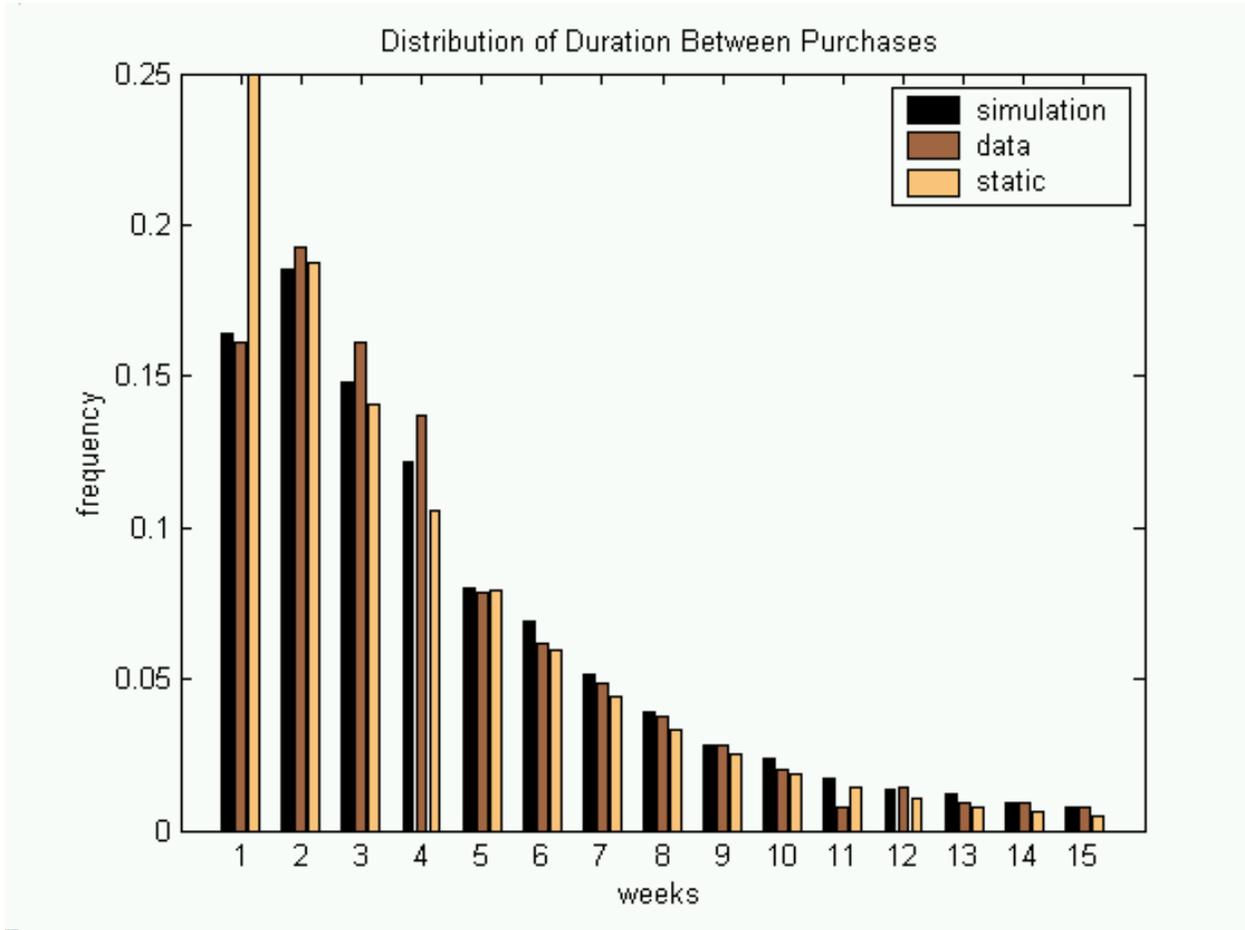[a] Computed as (dynamic PCM - static PCM)/static PCM.

Figure 1

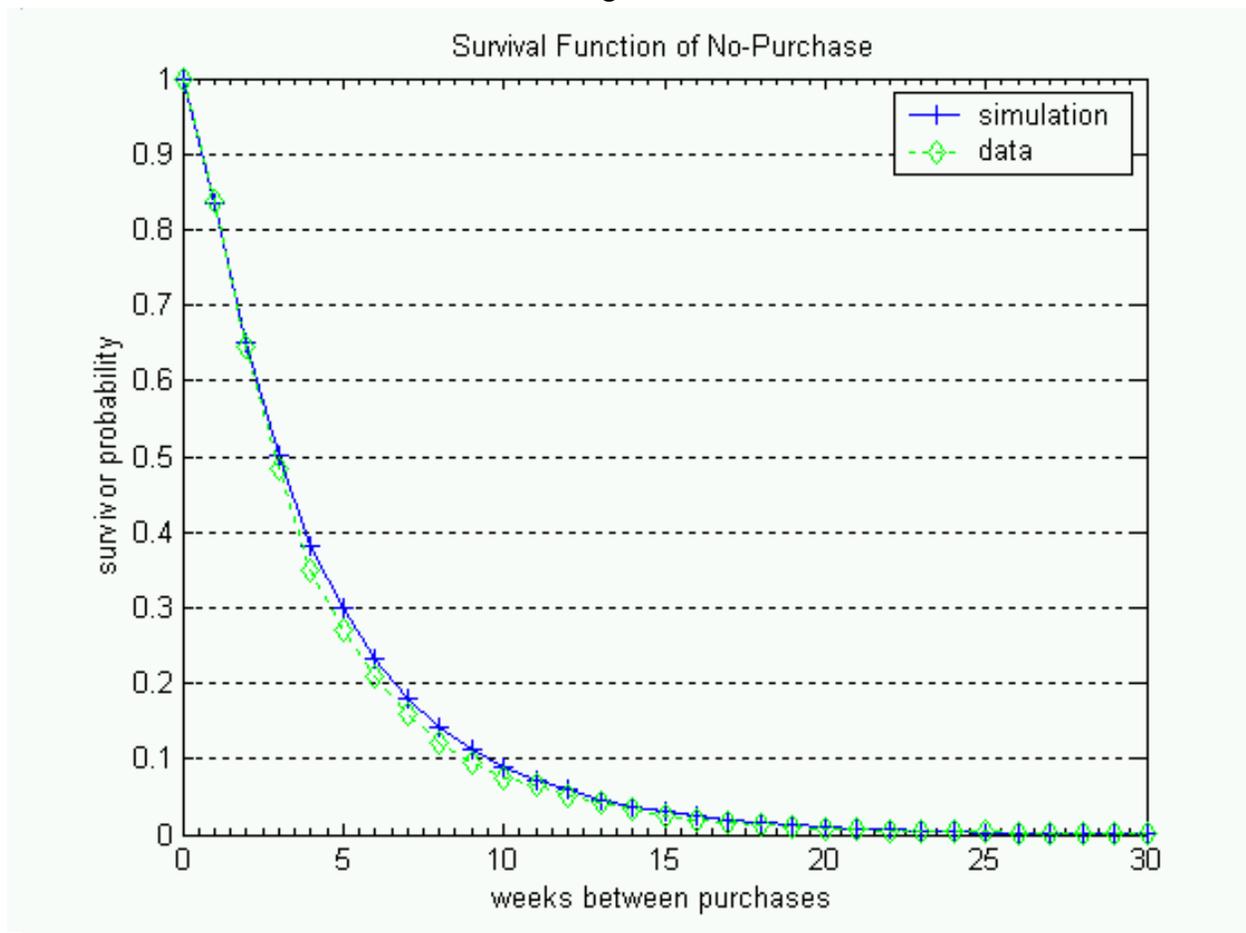Distribution of Duration Between Purchases

Figure 2

Survival Function of No-Purchase

Figure 3

Hazard Function of No-Purchase