Osborne, Matthew

Working Paper

Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination

EAG Discussion Paper, No. EAG 07-10

Provided in Cooperation with:
Economic Analysis Group (EAG), Antitrust Division, United States Department of Justice

ECONOMIC ANALYSIS GROUP

DISCUSSION PAPER

Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination

by

Matthew Osborne*

EAG 07-10 September 2007

---

*Research Economist, Antitrust Division, U.S. Department of Justice, Washington, DC 20530. Email: matthew.osborne@usdoj.gov. The views expressed are not purported to reflect those of the United States Department of Justice. I am indebted to my advisors, Susan Athey, Timothy Bresnahan and Wesley Hartmann for their support and comments. I would also like to thank Liran Einav, Chuck Romeo, and Dan Quint for helpful comments. I would like to thank the Stanford Institute for Economic Policy Research for financial support when I was in graduate school, and the James M. Kilts Center, GSB, University of Chicago, for provision of the data set used in this report.

Abstract

I formulate an econometric model of consumer learning and experimentation about new products in markets for packaged goods that nests alternative sources of dynamics. The model is estimated on household level scanner data of laundry detergent purchases, and the results suggest that consumers have very similar expectations of their match value with new products before consumption experience with the good, but once consumers have learned their true match values they are very heterogeneous. I demonstrate that resolving consumer uncertainty about the new products increases market shares by 24 to 58%. The estimation results also suggest significant switching costs: removing switching costs increases new product market shares by 12 to 23%. Using counterfactual computations derived from the estimates of the structural demand model, I demonstrate that the presence of switching costs with learning changes the implications of the standard empirical learning model: the intermediate run impact of an introductory price cut on a new product's market share is significantly greater when the only source of dynamics is switching costs as opposed to when both learning and switching costs are present, which suggests that firms should combine price cuts with introductory advertising or free samples to increase their impact.

Because my model includes two different types of dynamics, I am able to assess the impact of ignoring learning or switching costs on the model's imputed long run price elasticities by re-estimating the model assuming that one of these dynamics is not present. I find that ignoring learning will i) lead to underestimates of the own price elasticities of new products by 30%, ii) will underestimate the cross-price elasticities between new and established products by up to 90%, iii) will overestimate the cross-price elasticities of established products by up to 15%. Ignoring switching costs will lead to underestimates of own price elasticities of up to 60%, and underestimates of cross-price elasticities of up to 90%.

1 Introduction

An experience good is a product that must be consumed before an individual learns how much she likes it. This makes purchasing the product a dynamic decision, since the consumer's decision to experiment with a new product is an investment that will pay off if the consumer likes the product and purchases it again in the future. Consumer learning in experience goods markets has been an important subject of theoretical research in industrial organization and marketing since the 1970's. Learning can be an especially important factor in the demand for new products, and there is a small empirical literature that quantifies learning in household panel data using structural demand models with forward-looking consumers (for example, Erdem and Keane (1996), Crawford and Shum (2005)). In these papers it is assumed that the only type of dynamics in demand come from learning, and alternative types of dynamics, such as switching costs or consumer taste for variety, are not modeled. Similarly, papers that estimate other forms of dynamics (see Chintagunta, Kyriazidou, and Perktold (2001) for an example) usually only allow for one type of dynamics in demand.

In this paper, I estimate a structural model of learning and experimentation that nests alternative sources of dynamics in demand, such as the cost of switching between different products. In my model, consumers are forward-looking and take into account the effect of learning and other sources of dynamics on their future utility. I also allow a rich distribution of heterogeneity in consumer tastes, price sensitivities, expectations of new product match values, and alternative dynamics. To my knowledge, this paper is the first to estimate a demand model with such a rich dynamic structure.

The model is estimated on household-level longitudinal data on laundry detergent purchases. During the time the data was collected, three new product introductions are observed. Learning can be empirically separated from switching costs through differences in the effect of having made a first purchase of a new product on a consumer's current purchase relative to the effect of having used a product in the previous purchase event. Allowing for switching costs in addition to learning changes the implications of the standard empirical learning model: for example, the presence of switching costs makes consumers less likely to experiment with new products. Additionally, I find that the intermediate run impact of an introductory price cut increases when both learning and switching costs are present, compared to the case where only learning is present, since consumers who purchase the product and find they have a low match value for the product (alternatively, a low permanent taste for the product) may find it too costly to switch away from it.

Another contribution of this paper to the industrial organization literature is to examine the impact of misspecifying dynamics on estimates of long term price elasticities. This issue is economically relevant because cross-price elasticities are used to assess the impact of mergers, and it has been examined in some previous empirical research: for example, Hendel and Nevo (2006a) show that there are large biases from ignoring dynamics. In addition to estimating a model of learning and switching costs, I estimate two restricted structural demand models: one with no switching costs, and one with no learning. I find that in both cases, own and cross price elasticities elasticities are underestimated by leaving out dynamics, often by as much as 90%. An exception to this is that cross-price elasticities between established products are slightly overestimated in the model with no learning. A final contribution of this paper relative to the existing literature is that I use a recently developed technique allowing Bayesian estimation of a dynamic discrete choice model to include a richer heterogeneity structure than has been included in most papers.

Although my empirical results are found in a data set on laundry detergents, they should be applicable to other product categories where similar sorts of behavior have been observed. In consumer packaged goods markets, consumer switching costs or taste for variety has been found to play an important role in many product categories: some examples are nondiet soft drinks and liquid laundry detergents (Chintagunta 1999), ketchup (Roy, Chintagunta, and Haldar 1996), margarine, yogurt and peanut butter (Erdem 1996), breakfast cereals (Shum 2004), and orange juice (Dubé, Hitsch, and Rossi 2006). Evidence of learning has been found in liquid laundry detergents (Erdem and Keane 1996) and yogurts (Ackerberg 2003). Since new product introductions are very frequent in packaged goods markets, the issues that were raised in the first few paragraphs are going to be important. These types of dynamics are present in markets other than packaged goods. For example, learning has been found to play an important role in pharmaceuticals (Crawford and Shum 2005), automobile insurance (Israel 2005), and personal computers (Erdem, Keane, Öncü, and Strebel 2005). Markets where switching costs play an important role are discussed in Farrell and Klemperer (2006); some examples are cellular phones, credit cards, cigarettes, air travel, online brokerage services, automobile insurance, and electricity suppliers. It is certainly conceivable that consumer learning could also play a role in these markets. For example, in cellular phone markets consumers will likely learn about aspects of their provider's service after signing a contract with them. Cellular phone contracts often penalize consumers for switching providers, which creates switching costs and makes the decision to invest in a plan a forward-looking one.

The underlying model of consumer learning that is used in this paper is similar to many of the papers that I have mentioned in the previous paragraphs. In particular, I assume that learning occurs during consumption of the product: laundry detergent is assumed to be an experience good. I assume that consumers have an individual-level match value for each product that does not change over time. For the three new products, I assume that consumers may not know their true match values beforehand, and will have expectations about their match values which are right on average. If the consumers' true match values are equal to their expected match values, then there is no learning. This might be the case if consumers learn their match value beforehand through other means, such as experience with established products or by examining the new product's package. I also assume that consumers are forward-looking, which means that they will recognize that there is value to learning about new products. If there is indeed learning, and there are no other sources of dynamics in demand, one should expect to see consumers purchasing the new products very soon after they are introduced. Also, after experimenting with the new product, consumers who have a high match value for the new product will continue to purchase it after experimenting, and consumers who have a low match value will switch back to the established product. These behaviors will identify the magnitude of the learning.

One problem with this identification approach is the presence of dynamics in demand that are not learning. For example, some consumers may be variety-seeking: holding fixed their intrinsic match values, a previous purchase of the new product will decrease their current marginal utility for the product. These consumers will tend to purchase the new product very soon after its introduction and will switch away from it afterwards. To the researcher, it may look like these consumers experimented with the product and found their match value was low. Alternatively, there could be switching costs for some consumers: holding fixed their intrinsic match values, their marginal utility for products other than the new product will decrease after purchasing the new product. When a consumer with switching costs makes a first purchase of the new product, she will be likely to keep on purchasing it. Switching costs could arise in packaged goods markets for psychological reasons: for example, it may be costly for consumers to reoptimize their utility across purchase events, which could bias them to purchase the same product as in their previous purchase, or consumers may become more brand loyal in response to a previous purchase (Klemperer (1995) notes that brand loyalty can create switching costs).[2] To a researcher it may look like these consumers have a high match value for the new product. Thus, will be important to take into account that these other types of dynamics exist in order to properly isolate learning. I will argue that these types of dynamics can be identified in the long run, during periods where there have been no new product introductions for a long time. Learning can then be identified using the arguments specified above in the first few periods after the new product introductions.

Previous papers that have estimated structural models of consumer learning on household level panel data have not included these other types of dynamics; three well-known papers in this category which I previously discussed are Erdem and Keane (1996), Ackerberg (2003), and Crawford and Shum (2005), which estimate structural models of Bayesian learning in laundry detergents, yogurts, and ulcer medications, respectively. A paper that raised the problem of leaving out alternative dynamics is Israel (2005), which looks for learning in the time-series behavior of departure probabilities from an automobile insurance firm. The paper's model allows consumers to learn about the firm's quality, and also controls for consumer lock-in by allowing the number of time periods spent with the firm to enter utility directly. Although this paper was an important first step in examining the importance of more than one type of dynamics, there are two important aspects of demand which need to be addressed in a unified structural model of dynamics. First, the paper does not distinguish between consumer lock-in and unobserved heterogeneity in preferences. A researcher may observe a

[2] In the empirical marketing literature, these types of switching costs are equivalently referred to as habit persistence or inertia. As discussed above, they have been found to play an important role in many packaged goods markets.

consumer staying with the firm for a long time because she has a strong preference for the firm, or it may be because she becomes locked in to it. This issue may not be critical in automobile insurance markets, but in packaged goods markets it is important not to confuse these two behaviors, because the long run effect of a temporary price cut on a product's future share will be different under switching costs as opposed to taste heterogeneity. Under switching costs, a temporary price cut will increase a product's future market share; under heterogeneity, this will not be the case. Second, the paper does not directly model the forward-looking behavior of consumers by solving for their value function, but instead includes a term in the utility function which is interpreted as the value function. The estimated parameters of this paper are potentially subject to the Lucas critique: they will be functions of policy variables, such as the distribution of future prices.

To assess the impact of ignoring switching costs in a model of consumer learning, I compare the estimates from my model of learning and switching costs to estimates produced by a restricted model, where I assume that there are no switching costs. In the model with no switching costs, the variances in consumer taste distributions are estimated to be much larger. Consistent with the intuition above, the impact of learning is estimated to be smaller: consumers are much less similar in their initial predictions of how much they will like the new products. This difference is economically significant: in the full model, learning increases new product market shares by 24% to 58%; in the model with no switching costs, the impact of learning on market shares is smaller, ranging from 9.9% to 34%.[3] Furthermore, when switching costs are ignored, the impact of introductory price cuts on future market shares is underestimated, and long term price elasticities are significantly underestimated.

An alternative way of misspecifying the model would be to ignore learning, and assume that the only source of dynamics is switching costs or consumer taste for variety. I also estimate a restricted version of my model, where I assume that switching costs and taste for variety are the only source of dynamics. I find if learning is ignored, the distributions of price sensitivities, taste distributions, and switching costs look similar to those implied by the full model. However, the model does a significantly worse job at predicting the initial market shares of both new and established products. The model overpredicts the impact of introductory price cuts on the subsequent market shares of new products, and significantly underestimates long term price elasticities.

As with the literature on learning, there are many papers that estimate models of switching costs. In economics, perhaps the most well-known work about forward-looking switching costs is the work on rational addiction in Becker and Murphy (1988) and Becker, Grossman, and Murphy (1994). In marketing, there are many papers which estimate structural models of switching costs (alternatively, inertia or habit-formation) or variety-seeking in the presence of unobserved taste heterogeneity (for an example see Seetharaman (2004)). Although these papers account for rich sources of dynamics

[3]To quantify the importance of learning, I simulate the new product market shares using my model estimates, and then re-simulate them assuming that consumers have full information.

in demand, they typically do not model consumers as forward-looking. Two examples of papers that do allow for forward-looking behavior in this type of model are Chintagunta, Kyriazidou, and Perktold (2001) and Hartmann (2006). Chintagunta, Kyriazidou, and Perktold (2001) estimates a model of consumer switching costs in panel data on yogurt purchases. Hartmann (2006) examines intertemporal consumption effects in consumer decisions to play golf. In this paper, consumers are forward-looking, and dynamics arise through the fact that a consumer's decision to play golf will affect her future marginal utility for golf.

A third type of misspecification arises from failing to control for unobserved heterogeneity. As I mentioned above, this can bias the predictions of the impact of price cuts on future market shares. Additionally, it can make the identification of learning more difficult. To see why, suppose that when a new product is introduced, its price is initially low and then it is raised, a practice that is common and is observed in the data set used in this paper. Suppose further that there is a group of consumers who are very responsive to price cuts. These consumers will purchase the new product right after its introduction, when it is inexpensive, and will switch away from it as it gets more expensive. To a researcher who does not take into account that they are price sensitive, it may look like they experimented with the product and disliked it.

For computational reasons, most papers that estimate structural models of consumer learning or switching costs have had to be parsimonious in how they specify consumer heterogeneity, if it is modeled at all. For example, Erdem and Keane (1996) do not allow for consumer heterogeneity,[4] while Crawford and Shum (2005) allows for individual level heterogeneity in two dimensions: how serious the patient's sickness is, and how good a match a particular ulcer medication is for the patient. The paper assumes that the distribution of unobserved heterogeneity is discrete: in each of the 2 dimensions, consumers fall into a small number of types. Ackerberg (2003) allows 2 dimensions of individual-level heterogeneity: the intercept of each consumer's utility for a new yogurt, which is assumed to be known and observed by the consumer, and the consumer's intrinsic match value which is being learned. Unlike Crawford and Shum (2005), who assume that the population distribution of unobserved heterogeneity is discrete, this paper assumes the heterogeneity is normally distributed across the population. Although allowing for continuously distributed heterogeneity increases computational burden, the model is kept computationally tractable since consumer choice is binary: there is only one new product introduction, and consumers either purchase the new product or they do not. This approach would be less tractable in markets where there are multiple new product introductions.

These papers estimate their econometric models using classical methods such as the maximum-

[4] In Erdem and Keane (1996), consumers are learning about one unobserved attribute of each brand of laundry detergent, which is interpreted as the detergent's cleaning power. Each time an individual purchases a product she receives a signal of the product's quality, which is her perceived product quality. Under full information, consumer tastes for each product are this attribute level plus an idiosyncratic error term that is i.i.d. across time and consumers.

likelihood estimator. In models where consumers are forward-looking, it is necessary to solve their Bellman equation whenever the parameters of the model are changed, such as when a derivative is evaluated. This makes the model estimation computationally difficult. Allowing for unobserved heterogeneity substantially increases the computational difficulty of the estimation due to the fact that the unobserved heterogeneity must be integrated out by simulation. Because of these issues, researchers who have estimated these types of models have had to be parsimonious in their specification of unobserved heterogeneity. As I have already discussed using my example with consumer price sensitivities, failing to account for unobserved heterogeneity can result in biases. One way to reduce this computational burden is to use an importance sampling method developed by Ackerberg (2001) to reduce the computational burden induced by the heterogeneity. Hartmann (2006) uses this to allow for a richer distribution of heterogeneity than in the learning papers I have previously discussed.

I overcome the problems associated with unobserved heterogeneity by estimating my model using the Bayesian method of Markov Chain Monte Carlo, which is well suited to dealing with high-dimensional unobserved heterogeneity than classical techniques. To reduce the computational burden that is created by solving the consumers' Bellman equations, I apply a new technique by Imai, Jain, and Ching (2005). In contrast to most classical techniques, which require the Bellman equation to be calculated many times, this new technique only requires one full solution of the Bellman equation. The basic idea behind this method is to update the value function once in each step of the Markov Chain Monte Carlo algorithm using information from previous steps, so that by the time the estimation is completed an accurate approximation of the value function is obtained. This paper is the first to apply this new technique to field data. It would also likely be possible to estimate this model using the method of Ackerberg (2001). An interesting topic for future research would be to compare the computational speed and accuracy of these two estimation techniques.

2 Data Set

2.1 Discussion of the Scanner Data

The data set I am using is A.C. Nielsen supermarket scanner data on detergent purchases in the city of Sioux Falls, South Dakota between December 29, 1985 and August 20, 1988. This data is particularly useful for identifying consumer learning for two reasons: first, since this data is a panel of household purchases, it allows one to track individual household behavior over time. Second, during the period that this data was collected, three new brands of liquid laundry detergents were introduced to the market: Cheer in May 1986, Surf in September 1986 and Dash in May 1987. Households that participated in this study were given magnetic swipe cards, and each time the household shopped at a major grocery or drugstore in the city, the swipe card was presented at the checkout counter. Additionally, households that participated in the study filled out a survey containing basic demographic information. The distributions of household demographics are shown in Table 1.

Although a visit to the grocery store will reveal many different brands of laundry detergent, the market is dominated by 3 large companies: Procter and Gamble (Dash, Cheer, Era, Tide), Unilever (Wisk, Surf) and Colgate-Palmolive (Fab, Ajax). During this period, laundry detergents were available in two forms: liquids and powders. Table 2 shows the distribution of sizes for liquid and powder products. For liquids, the most popular size was the 64 ounce size. Table 3 shows the market share for the 7 most popular brands of laundry detergents (the other category covers purchases of smaller brands), in their liquid and powdered forms. As can be seen from the last column of the table, the market share of liquids is about 52%. Well known brands, such as Wisk and Tide, have high market shares.

Table 4 shows the market shares of different brands of liquids over different periods of time. It is notable that for all three new products, their market share tends to be significantly higher in the first 12 weeks after introduction than it is for the remainder of the sample period. This fact is consistent with learning, since the option value of learning induces consumers to purchase new products early. However, it is also consistent with consumer response to introductory pricing. The average prices of different brands at different periods of time are shown in the same table, underneath the shares. There are two noteworthy facts in this table. First, prices of the new brands Cheer and Surf tend to be lower in the first 12 weeks after introduction than they are later on in the data. This fact suggests that we should be aware of possible biases due to consumer heterogeneity: for example, price sensitive consumers could purchase the new products initially when they are cheap, and switch away from them as they get more expensive, which could be mistaken for learning. Second, when Cheer is introduced to the market by Procter and Gamble, the price of Wisk, a popular product of Unilever, goes down. Similarly, when Unilever's Surf is new, Procter and Gamble's Tide drops in price. Cheer and Surf have been successful products since their introductions, but Dash was discontinued in the United States in 1992. One possible reason for this is that Dash was more of a niche product: it was intended for front-loading washers, which constituted about 5% of the market at the time.

2.2 An Overview of the Laundry Detergent Market Prior to 1988

The fact that the three new products were liquid detergents was not a coincidence, and to see why it is useful to briefly discuss the evolution of this industry. The first powdered laundry detergent for general usage to be introduced to the United States was Tide, which was introduced in 1946. Liquid laundry detergents were introduced later: the popular brand Wisk was introduced by Unilever in 1956. The market share of liquid laundry detergents was much lower than powders until the early

1980's. The very successful introduction of liquid Tide in 1984 changed this trend, and detergent companies began to introduce more liquid detergents. Product entry in this industry is costly: an industry executive quoted the cost of a new product introduction at 200 million dollars (Cannon et al. 1987). Industry literature suggests a number of reasons for the popularization of liquids during this time: first, low oil and natural gas prices, which made higher concentrations of surfactants[5] more economical; second, a trend towards lower washing temperatures; third, increases in synthetic fabrics; fourth, on the demand side, an increased desire for convenience. In the third and fourth points, liquids had an advantage over powders since they dissolved better in cold water, and did not tend to cake or leave powder on clothes after a wash was done.

The fact that new liquids were being introduced at this time suggests that learning could be an important component of consumer behavior. Many consumers may not have been familiar with the way liquids differed from powders, and they might learn more about liquids from experimenting with the new products. Further, there may be learning across the different brands of liquids. For example, using liquid Tide might not give consumers enough information to know exactly how liquid Cheer or Surf will clean their clothes. Learning about these products could be important for consumers to know how well these products will work for a number of reasons. First, laundry detergents are fairly expensive and the household will use the product for a long period of time, so the cost of making a mistake is not trivial. Second, consumers may have idiosyncratic needs which require different types of detergents. As an example, a consumer whose wardrobe consists of bright colors will likely prefer to wash in cold water, where liquids are more effective.

3 Econometric Model

3.1 Specification of Consumer Flow Utility

In my structural econometric model an observation is an individual consumer's purchase event of a liquid laundry detergent. In the following discussion, I index each consumer with the subscript i, and number the purchase events for consumer i with the subscript t. The dependent variable in this model is the consumer's choice of a given size of one of the 13 different laundry detergents listed in Table 3. I index each product with the variable j, and each size with of the product with s. In a particular purchase event t for consumer i, not all of the 49 choices may be available. I denote the set of products available to consumer i in purchase t as J_{it}. I assume that a consumer's period utility is linear, as in traditional discrete choice models. The period, or flow utility for consumer i for product-size $(j, s) \in J_{it}$ on purchase event t is assumed to be

[5]The most important chemical ingredient to laundry detergents are two-part molecules called synthetic surfactants which loosen and remove soil. Surfactants are manufactured from petrochemicals and/or oleochemicals (which are derived from fats and oils).

$$
\begin{aligned}
&u_{ijst}(S_{it-1}, \alpha_i, p_{ijt}, c_{ijt}, \beta_i, x_{ijt}, \eta_i, y_{ijt-1}, \varepsilon_{ijt}) \\
&= \Gamma_{ij}(S_{ijt-1}, y_{ijt-1}) + \xi_{is} + \alpha_i(p_{ijst} - \alpha_{ic}c_{ijt}) + \beta_i x_{ijt} + \eta_i y_{ijt-1} + \varepsilon_{ijst},
\end{aligned} \tag{1}
$$

where $\Gamma_{ij}(s_{ijt-1}, y_{ijt-1})$ is consumer i's match value, or taste, for product j. A consumer's match value with a product is a function of the two "state variables" s_{ijt-1} and y_{ijt-1}. The variable y_{ijt} is a dummy variable that is 1 if consumer i chooses product j in purchase event t, so y_{ijt-1} keeps track of whether consumer i chose product j in her previous purchase event. The state variable s_{ijt} keeps track of whether consumer i has ever purchased product j prior to purchase event t, and it evolves as follows:

$$
s_{ijt} = s_{ijt-1} + 1\{s_{ijt-1} = 0 \text{ and } y_{ijt-1} = 1\}. \tag{2}
$$

For the 10 established products, I assume that consumer match values do not change over time, so $\Gamma_{ij}(s_{it-1}, y_{it-1}) = \gamma_{ij}$. For identification purposes, I normalize every consumer's match for other liquid (product 1) to 0. For the three new products, I assume that the evolution of the consumer's permanent taste is as follows:

$$
\begin{aligned}
\Gamma_{ij}(s_{ijt-1}, y_{ijt-1}) &= \gamma^0_{ij} \text{ if } s_{ijt-1} = 0, \text{ and } y_{ijt-1} = 0 \\
\Gamma_{ij}(s_{ijt-1}, y_{ijt-1}) &= \gamma_{ij} \text{ if } s_{ijt-1} = 1, \text{ or } y_{ijt-1} = 1.
\end{aligned} \tag{3}
$$

The consumer's match value for the new product is γ^0_{ij} if the consumer has never purchased the product before, and it is γ_{ij} once she has. For the three new products, γ^0_{ij} is consumer i's prediction of how much she will like product j before she has made her first purchase of it. γ_{ij} is her "true" match with the product.

I assume that

$$
\gamma_{ij} \sim N(\gamma^0_{ij}, \sigma^2_{ij}), \tag{4}
$$

where σ^2_{ij} is consumer i's uncertainty about her true taste for product j. I allow σ^2_{ij} to vary with the household i's income and size as follows:

$$
\sigma^2_{ij} = \sigma_{max} \frac{\exp(\sigma_{0ij} + \sigma_{1j} INC_i + \sigma_{2j} SIZE_i)}{1 + \exp(\sigma_{0ij} + \sigma_{1j} INC_i + \sigma_{2j} SIZE_i)}. \tag{5}
$$

Note that there is unobserved heterogeneity in σ^2_{ij} as well as observed heterogeneity: σ_{0ij} varies across individuals and accounts for unobserved heterogeneity. INC_i is a variable that varies from 1 to 4, where the four possible categories correspond to the four income groups in Table 1. Household size, the variable $SIZE_i$, also varies from 1 to 4 and is defined similarly. Note that σ^2_{ij} is always positive and bounded above by σ_{max}, which I assume is equal to 5.[6]

The parameter α_i is consumer i's price sensitivity. I also allow this parameter to vary with household income and size as follows,

$$\alpha_i = \alpha_{max} \frac{\exp(\alpha_{0i} + \alpha_1 INC_i + \alpha_2 SIZE_i)}{1 + \exp(\alpha_{0i} + \alpha_1 INC_i + \alpha_2 SIZE_i)}, \tag{6}$$

where α_{max} is set to -100. α_i is assumed to always be negative and, like σ_{ij}^2, it is bounded. p_{ijst} is the price in dollars per ounce of size s of product j in the store during purchase event t, and the variable c_{ijt} is the value of a manufacturer coupon for product j that consumer i has on hand in purchase event t, also measured in dollars per ounce. The parameter α_{ic} is consumer i's sensitivity to coupons. I assume that α_{ic} lies between 0 and 1, and that

$$\alpha_{ic} = \frac{\exp(\alpha_{0ic})}{1 + \exp(\alpha_{0ic})}, \tag{7}$$

where α_{0ic} lies on the real line.

In Equation (1), β_i is a vector that measures consumer i's sensitivity to other variables, x_{ijt}. The first and second elements of the x_{ijt} vector are dummy variables which are equal to 1 if product j is on feature or display, respectively. The third element is a dummy variable that is 1 if purchase event t occurs in the first week after the introduction of Cheer, and j is Cheer. The fourth is the same thing for the second week of Cheer, the fifth for the third and so on up to the fourteenth week after the Cheer introduction. The next element is a dummy variable that is 1 if purchase event t occurs in the third week after the introduction of Surf, and j is Surf (the first and second week dummy variables were dropped because they were not identified in simple logit estimations of the demand model). The next 11 elements are the same thing for weeks 4 to 14 after the Surf introduction. The next 14 elements of the vector are the same time-product dummy variables for the Dash introduction. These time dummy variables are included to capture the effect of unobserved introductory advertising for the new products.

The consumer's utility in purchase event t is increased by η_i if she purchases the same product that she did in purchase $t-1$.[7] Note that the parameter η_i and the function $\Gamma(s_{ijt-1}, y_{ijt-1})$ allow two different sources of dynamics in consumer behavior: consumer's previous product choices can

[6] The choice of the number 5 is somewhat ad hoc, but the important thing is that when choosing the upper bound for this parameter the number should be high enough to not be binding - there should not be consumers with values of σ_{ij}^2 greater than five. In the model estimates section I will examine the distribution of σ_{ij}^2 across the population - they do not appear to approach the upper bound. Furthermore, in my thesis research I estimate a version of this demand model where there is no learning and consumers are not forward-looking. The estimates of the taste variances for all the three new products are significantly smaller than 5.

[7] Although some purchases were dropped due to data problems as described in Section 4.1, y_{ijt-1} is defined to be the product chosen in the consumer's previous purchase event, even if that purchase event was dropped. When two product choices are made in the same week, I set y_{ijt-1} to be the second (or third, if there were three) recorded purchase. This avoids measurement error in y_{ijt-1}. Also, the fact that I include the consumer's previous purchase event in the model

affect her current utility. One way in which a consumer's past product choices affect her current product choice is through the $\Gamma(s_{ijt-1}, y_{ijt-1})$ function: this is *learning*. If she has never purchased the new product j prior to purchase event t, her taste for this product is her expected taste, γ_{ij}^0, whereas if she has purchased it at some point in the past I assume that she knows her true taste for the product, γ_{ij}. The term η_i accounts for the dynamic behaviors of *switching costs* or *variety-seeking*. If $\eta_i > 0$, consumer i's utility is greater if she consumes the same product twice in a row. Thus, a positive η_i induces a switching cost (Pollack (1970), Spinnewyn (1981)). An alternative way to model switching costs would be to subtract a positive η_i from all products except the one that was previously chosen; since utility functions are ordinal and there is no outside good in this model, these two formulations are equivalent. As discussed in the introduction, switching costs have been found to be an important part of demand for consumer packaged goods, and could arise if there are costs of recalculating utility if a consumer decides to switch products. If $\eta_i < 0$, the consumer will prefer to consume something different than her previous product choice: I label this as variety-seeking (McAlister and Pessemier 1982). Variety-seeking is not likely an important behavior in laundry detergent markets, but I allow it in the model for the sake of generality. As with the price coefficient and consumer uncertainty, I allow both observed and unobserved heterogeneity in η_i:

$$\eta_i = \eta_{i0} + \eta_1 INC_i + \eta_2 SIZE_i \tag{8}$$

Last, the ε_{ijst} is an idiosyncratic taste component that is i.i.d. across i, j, s and t, and has a logistic distribution. I assume this error is observed to the consumer but not the econometrician and is independent of the model's explanatory variables and the individual's utility parameters such as α_i and β_i.

I allow unobserved heterogeneity in most of the individual-level parameters for every consumer: the γ_{ij}'s for all products except for the Powder Other and Powder Tide products, the γ_{ij}^0's, the α_{0i}'s, the the ξ_{is}'s's, the α_{0ic}'s and σ_{0ij}'s for the three new products, the intercept of the switching costs parameter η_{i0}, and the β_i vector. Denote the vector of population-varying individual level parameters for consumer i listed previously as θ_i, and the vector of individual level parameters with the γ_{ij}'s for the three new products removed as $\tilde{\theta}_i$. I assume that $\tilde{\theta}_i \sim N(b, W)$ across the population, where W is diagonal.[8] This assumption means that the household's uncertainties about tastes for

creates an initial conditions problem: I do not observe the consumer's purchase event before the data collection starts. To get around this problem, I estimate the model using data from the period after the introduction of Cheer Liquid, so that the consumer's initial y_{ijt-1} is her latest purchase before the introduction of Cheer. For a very small number of households the first observed purchase event occurs after the introduction of Cheer. For these households I impute the first purchase as the most frequently purchased established product.

[8] A possible worry may be that the assumption of normally distributed heterogeneity is too restrictive. In my thesis research I present the results of the estimation of an extended model where some of the heterogeneity is assumed to be

the new products, σ_{ij}^2's, and the price sensitivities α_i's will be transformations of normals as shown in Equations (5) and (6). Their distribution is Johnson's S_B distribution, which is discussed in Johnson and Kotz (1970), page 23. The parameters which do not vary across the population are the γ_{ij}'s for Other Powder and Tide Powder, the coefficients on household demographics for the learning parameters, the price sensitivities and the switching costs, which are σ_{1j} and σ_{2j}, α_{1j} and α_{2j} and η_1 and η_2 respectively, and a group of parameters which capture consumer expectations of future coupons c_{ijt}. These latter parameters will be discussed further in the next section. I denote the vector of population-fixed parameters as θ.

3.2 Consumer Dynamic Optimization Problem

I assume consumers are forward-looking[9] and in each purchase event they maximize the expected discounted sum of utility from the current purchase into the future. The consumer's expected discounted utility in purchase event t is

$$V(\Sigma_{it};\theta_i,\theta) = \max_{\Pi_i} E\left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} u_{ijs\tau}(S_{i\tau-1}, p_{ij\tau}, c_{ij\tau}, x_{ij\tau}, y_{ij\tau-1}, \varepsilon_{ij\tau}, \theta_i)|\Sigma_{it}, \Pi_i;\theta_i,\theta\right], \quad (9)$$

where Π_i is a set of decision rules that map the state in purchase t, Σ_{it}, into actions, which are the y_{ijt}'s in purchase event t. The parameter δ is a discount factor, which is assumed to equal 0.95.[10] The function $V(\Sigma_{it};\theta_i,\theta)$ is a value function, and is a solution to the Bellman equation

$$V(\Sigma_{it};\theta_i,\theta) = E_{\varepsilon_{ijt}}[\max_{(j,s)\in J_{it}} \{u_{ijst}(S_{it-1}, p_{ijt}, c_{ijt}, x_{ijt}, y_{ijt-1}, \varepsilon_{ijt}, \theta_i) + \delta EV(\Sigma_{it+1};\theta_i,\theta)\}]. \quad (10)$$

The state vector in purchase event t, Σ_{it}, has the following elements: the S_{ijt-1}'s for the new products, the y_{ijt-1}'s for all 13 products, the prices of all products, p_{ijt}, the set of available products, J_{it}, and a new state variable n_t, which will be discussed later.

The expectation in front of the term $V(\Sigma_{it+1};\theta_i,\theta)$ in Equation (10) will be taken over the distributions of future variables, which are

i) the true tastes for new products the consumer has never purchased, as in Equation (4),

ii) future prices,

a two point mixture of normals. Identification of some parameters becomes more difficult in this case, but generally the results do not change, which suggests that the assumption of normality is a good fit.

[9] In my thesis research (Osborne 2006), evidence is provided that consumers are forward-looking in this data set.

[10] The discount factor is usually difficult to identify in forward-looking structural models, so it is common practice to assign it a value. Since the timing between purchase events varies across consumers, it is possible that the discount factors may also vary across consumers. As I will discuss in a few paragraphs, I assume that all consumers have the same expectations about when their next purchase will occur, which removes this problem. Also, the estimation method I will use requires that the discount factor is not an estimated parameter.

iii) future coupons, and

iv) future product availabilities.

For reasons of computational tractability that will be discussed in the next section, I assume that consumers have naive expectations about future x_{ijt}'s, which are the feature, display, and time dummies. By this I mean that consumers expect all these variables to have future levels of zero. A result of this assumption is that these variables do not have to be included in the state space.[11]

I account for consumer expectations about future prices p_{ijst} and product availability J_{it} in the following way. I estimate a Markov transition process for prices and availability from the data on a store-by-store basis, using a method similar to Erdem, Imai, and Keane (2003) which I will briefly summarize. A detailed description of the estimation of this process can be found in Appendix A.2. I assume that consumers' actual expectations about these variables are equal to this estimated process. In my data, prices tend to be clustered at specific values, so the transition process for prices is modeled as discrete/continuous. The probability of a price change for a product conditional on its price in the previous week, last week's prices for other products, and whether a new product was recently introduced is modeled as a binary logit. Conditional on a price change, the probability of a particular value of the new price is assumed to be lognormal given the previous week's prices in the same store and whether a new product introduction recently occured. Note that there are 49 possible brand-size combinations, which makes the state space of prices very large. To reduce the size of the state space, the Markov process for prices is only estimated on the most popular sizes of liquids and powders. The prices of other sizes are assumed to be a function of the prices of the popular sizes.

An important part of the price process is that we observe introductory pricing for the new products. I assume consumers understand that the prices of new products will rise after their introduction, so I include a dummy variable in both the price transition logit and regression which is 1 for the first 12 weeks after the introduction of Cheer, a separate dummy variable which is 1 for the first 12 weeks after the introduction of Surf, and one for the first 12 weeks after Dash's introduction. Allowing for introductory pricing in this way will complicate the state space. To see why, consider a consumer who purchases a laundry detergent on the week of Cheer's introduction. Suppose further that this person purchases detergent every 10 weeks, and she knows exactly when she will make her future purchases. This person's next purchase will occur in 10 weeks, when the price of Cheer is still low. Her next purchase after that will occur in 20 weeks, when the price process is in its long run state. The number of purchase events before the consumer enters the long

[11] Assuming that consumers do not expect future advertising is probably not that unrealistic in the laundry detergent market. For this product category, it is likely that consumers will care more about future prices and how well the product they purchase will function. Future advertising is likely to be more important with "prestige" products, such as shoes or clothing.

run price state will be a state variable, which I denote as n_t.

A complication this variable n_t creates is that consumers probably do not know exactly when their next purchases of laundry detergents will be. Because the econometrician does not observe consumer expectations, the best we can do is to make an assumption about this. I assume that all households expect to make their next purchase of laundry detergent in exactly 8 weeks. In the sample of households I use to estimate the model on, household interpurchase times are clustered between 6 and 8 weeks, with a median interpurchase time of 8 weeks. This means that n_t will take on 2 values: 1 if the consumer's purchase occurs within the first 4 weeks after the new product introduction, and zero anytime afterwards.

For the state variable J_{it}, I estimate the probability of each detergent being available in a given calendar week for a given store separately using a binary logit. As was the case with prices, the process for availability is only estimated for the most popular sizes of each product, and so the only part of J_{it} that is a true state variable are the availabilities of these products. This means I estimate 13 logits, one for each product, where one of the regressors is whether the product was available in the previous week. The availabilities of less popular sizes are assumed to be a function of the availability of the popular sizes. I assume that the introductions of new products are a surprise to consumers, so this aspect of the state space is not taken into account by my availability estimation. A result of this assumption is that consumers will recalculate their value functions after each new product introduction: there will be a value function for after the Cheer introduction, a new one after the Surf introduction, and another one after the Dash introduction. Hence, there will be three times where it will be possible for n_t to be equal to 1, right after the introduction of each new product.

I treat consumer expectations about future coupons, which are the c_{ijt}'s, differently than future prices. As I will discuss further in the Section 4.2, I specify a process for the distribution of coupons and estimate the parameters of this process along with the other model parameters. I assume that the future c_{ijt}'s are composed of two random variables: a binary random variable $\bar{c}_{ijt}$ which is 1 if consumer i receives a coupon for product j in purchase t, and a random variable v_{ijt} which is the value of the coupon received. Denote probability of a consumer having a coupon on hand and available for use for product j when $n_t = 0$ as p^0_{cj}. Because consumers may expect more coupons to be available for new products when they are new, I allow the probability of receiving a coupon for a given product j to be different when $n_t = 1$. In particular, for the new products $j =$ Cheer, Surf and Dash I assume the probability of having a coupon is $p^0_{cj} + p^1_{cj}$. For established products, I assume the probability of receiving a coupon when $n_t = 1$ after the Cheer introduction to be $p^0_{cj} + p^{Cheer,1}_c$, after the Surf introduction to be $p^0_{cj} + p^{Surf,1}_c$, and after the Dash introduction to be $p^0_{cj} + p^{Dash,1}_c$. Note that the parameters $p^{Cheer,1}_c$, $p^{Surf,1}_c$ and $p^{Dash,1}_c$ do not vary by product. If a consumer receives a coupon for product j, the value of that coupon, which I denote as v_{ijt}, is multinomial and drawn from the empirical density of coupon values. Coupon values are clustered at certain numbers (such as 50 cents, 60 cents, or 1 dollar), so I calculate the probability of getting a particular coupon value for a particular brand in a period[12] by tabulating the number of redeemed coupons of that value for that brand in that period, and dividing by the total number of redeemed coupons for that product in that period.

The last part of the state space is the process on the state variables summarizing purchase history, S_{ijt-1} and y_{ijt-1}. Because these state variables are influenced by consumer choices, it is instructive to examine how we compute the value functions as these parts of the state space change. Suppose first that $S_{ijt-1} = 0$ for some product j. If the consumer decides to purchase product j for the first time, then S_{ijt} will be zero and y_{ijt} will be 1. When we construct the next period value function we will integrate out the consumer's true taste for product j, conditional on γ_{ij}^0 and σ_{ij}^2. Let γ be a random variable with the distribution of true tastes for product j, where $f(\gamma|\gamma_{ij}^0, \sigma_{ij}^2)$ is $N(\gamma_{ij}^o, \sigma_{ij}^2)$, and denote $\theta_i(\gamma)$ as the vector of individual level parameters for consumer i with her true taste draw for product j replaced by γ. Denote $v_{ikst+1}(\gamma)$ as consumer i's utility for size s of product k in purchase event $t+1$ as a function of γ, minus the logit error ε_{ijst+1}:

$$\begin{aligned}
\text{Product } k = j \quad &: v_{ikst+1}(\gamma) = && \gamma + \xi_{is} - \alpha_i(p_{ikst+1} - c_{ikt+1}) + \eta_i y_{ikt} + \delta EV(\Sigma_{it+2}; \theta_i(\gamma), \theta) \\
\text{Product } k \neq j \quad &: v_{ikst+1}(\gamma) = && \Gamma_{ik}(S_{ikt}, y_{ikt}) + \xi_{is} - \alpha_i(p_{ikst+1} - c_{ikt+1}) + \eta_i y_{ikt} \\
& && + \delta EV(\Sigma_{it+2}; \theta_i(\gamma), \theta).
\end{aligned} \tag{11}$$

Consumer i's expected value function in purchase event $t+1$, at her first purchase of product j ($S_{ijt} = 0$ and $y_{ijt} = 1$) will be

$$\begin{aligned}
& EV(\Sigma_{it+1}; \theta_i, \theta) \\
& = E_{c_{it+1}} E_{p_{it+1}|p_{it}} E_{J_{it+1}|J_{it}} \left[\int_{\gamma_{ij}} \ln \left(\sum_{(k,s) \in J_{it+1}} \exp(v_{ikst+1}(\gamma_{ij})) \right) f(\gamma_{ij}|\gamma_{ij}^0, \sigma_{ij}^2)\, d\gamma_{ij} \right].
\end{aligned} \tag{12}$$

When the consumer has purchased product j in the past, such as at state space points $S_{ijt} = 1$ and $y_{ijt} = 1$ or $S_{ijt} = 1$ and $y_{ijt} = 0$, the value function will be defined similarly, but will be simpler: the consumer's utility for all products given in Equation (11) will be a function of the true taste γ_{ij} rather than γ and the value function in (12) will not include the integral over γ. Note that even if consumer i knows her true taste for all 3 new products ($S_{ijt} = 1$ for all these products), there will still be dynamics in demand arising from the η_i. The consumer will take into account the fact that her purchase today will change y_{ijt}, and affect her utility in period $t+1$.

[12] There are six periods in all - when $n_t = 1$ after Cheer's introduction, when $n_t = 0$ after Cheer's introduction, when $n_t = 1$ and $n_t = 0$ after Surf's introduction, and when $n_t = 1$ and $n_t = 0$ after Dash's introduction.

3.3 Model Identification

I will explain the identification of the model in two steps. For simplicity, assume that we are examining a market with one new product introduction, similar to the market analyzed with the simple model in section 3. Assume further that we see each consumer for a long period of time. Although the estimation procedure I am using is likelihood-based, for brevity I will discuss it in the context of method of moments estimation. Thus, I will consider which moments in the data will be necessary to solve for the model's parameters. This is sufficient to show that the likelihood-based estimates are identified since the likelihood-based estimator, if it is correctly specified, is consistent and will converge to the same value as the method of moments estimator.

First, consider the period after most or all of the learning has occurred. In the long run, there will be no learning: since the distribution of the idiosyncratic error, ε_{ijst}, has infinite support, at some point in time everyone in the market will purchase the new product at least once. After every consumer has experimented with the new product, the only dynamics left in demand will be the switching costs or variety-seeking captured by the η_i's. At this point we are left with separately identifying the distribution of η_i's and the distribution of the "non-dynamic" coefficients in the consumer's flow utility: consumer tastes for established products, consumer price sensitivities, and the distribution of the coefficients for the x_{ijt}'s, the β_i's.

Consider first the task of identifying η_i for an individual consumer. The η_i causes state dependence in her demand: a consumer's choice in purchase event $t-1$ will affect her choice today. Chamberlain (1985) has argued that state dependence can be identified through the effect of previous exogenous variables on today's purchase probabilities. As an example, consider the effect of a price cut for Tide in purchase event $t-1$ on the probability of consumer i purchasing Tide in purchase event t. If the price cut has no effect of this probability, then $\eta_i = 0$. If the price cut increases the probability that the consumer purchases Tide in purchase event t, then $\eta_i > 0$ and the consumer has a switching cost. If the price cut decreases the probability of the consumer purchasing Tide in purchase event t, then $\eta_i < 0$ and the consumer is a variety-seeker. If we observe consumer i for a long period of time, and there is variation in the time series path of prices the consumer observes, then it should be possible to infer the size of the consumer's η_i. In the data, for many consumers we will not observe them long enough to be able to accurately estimate a consumer's individual η_i; identification is made easier by the fact that η_i is assumed to only be a function of household demographics.

Once the η_i distribution has been identified, we are left with identifying the heterogeneity of the non-dynamic coefficients in the consumer's flow utility. Identification of this part of consumer heterogeneity is straightforward and will come through the effect of variation in purchase event t exogenous variables on purchase event t purchase probabilities.

Now consider the periods right after the new product introduction, when we will need to identify

σ^2_{ij} and γ^0_{ij} for the new product j. In my model I allow these parameters to vary across the population, but to get a feel for identification it is easier to start with the case where there is no heterogeneity. Hence, for the next few paragraphs I will drop the i subscript. First, consider the identification of σ^2_j. In my thesis research (Osborne 2006), I solve a simplified version of the econometric model and simulate it numerically under different values of σ^2_j. In the simplified model there is one new product introduction, and one established product, and I assume that η_i is fixed across the population. Prices for each product are constant over time and there is no advertising, so product switching will be induced by variation in the error term. I simulate the model when σ^2_j=0, when it is positive, and for different values of η_i. Doing this allows me to find the model's testable implications, where the null hypothesis is σ^2_j=0 and the alternative is $\sigma^2_j > 0$.

There are two testable implications to this model which are examined in Osborne (2006), who finds support for them in the same laundry detergent scanner data which is used in this paper. The test statistics associated with them are shares of consumers who take actions at certain times, controlling for any time-series variation in prices. The first implication is that, under the maintained hypothesis that δ is high and $\eta_i = 0\ \forall i$, in the first two periods after the new product's introduction, the share of consumers who purchase the new product and then do not is greater than the share who do not and then do. This is due to the fact that when there is learning and $\eta_i = 0$, there will be a positive option value of learning which induces consumers to purchase the new product sooner rather than later.[13] When there is no learning, the test statistic will be zero since the order of purchase does not matter. This share difference is an increasing function of σ^2_j, because the option value of learning induces consumers to purchase the new product sooner rather than later, and the option value of learning is increasing in σ^2_j. If this share difference is greater in the data than the model would predict at $\sigma^2_j = 0$, then σ^2_j will pick up that difference.

The second testable implication is that for any value of the discount factor and for any value of η_i, among consumers whose previous purchase was the new product, the share of consumers who repurchase the product increases over time if $\sigma^2 > 0$. This is because initially the consumers whose previous purchase was the new product consist mostly of consumers who are experimenting; later it consists mostly of consumers who like the new product. The share of consumers who repurchase the new product is an increasing function of the population variance in tastes for the new product. Immediately following the new product introduction, this share will reflect the population variance in expected tastes, the γ^0_{ij}'s (which for the moment we have assumed to have zero variance). As consumers learn, the population variance in tastes will be increased by σ^2_j. Since consumers' taste draws will be taken from more extreme ends of the taste distribution, those who purchase the new product will tend to have higher taste draws after the learning has occurred and will be more likely to repurchase it. An increase in σ^2_j will increase the share of consumers who repurchase the

[13]Since evidence in favor of this implication is found in the data set I use (Osborne 2006), it is reasonable to conclude that for some new products the option value of learning is positive, and that consumers are forward-looking.

new product in periods after all learning has occurred. Hence, σ_j^2 can also be identified from the difference between the share of consumers who repurchase the new product immediately following the new product introduction and the share of consumers who repurchase the new product after all learning has occurred: the greater this difference, the greater is σ_j^2.

The identification of the mean of γ_j^0 and its variance is straightforward when σ_j^2 is constant across the population. First, note that in the period after the learning occurs, we can identify the distribution of true tastes for the new product. The mean of the population distribution of tastes for the new product will be the same as the mean of the γ_i^0 distribution. The variance in the distribution of true tastes for the new product is the variance of γ_j^0 plus σ_j^2. We can identify σ_j^2 using the share difference moment or from the change in the share of consumers who repurchase the new product. The variance of γ_j^0 will simply be the variance in the population distribution of true tastes minus the σ_j^2.

Last, I will relax the assumption that σ_j^2 is constant across the population. Separating out heterogeneity in σ_{ij}^2 and γ_{i0} is a harder task, because we only observe consumers making one first purchase of each new product. To see why this could be a problem, let us reconsider the simplified model which was just discussed, and assume that there are only 3 periods in it, the distribution of the error term for product 2 is standard normal, η_i is 0 for all consumers, and consumers discount the future at a rate of 0.95. Figure 1 shows the probability of experimenting with the new product in period 1 for different values of γ_{i0} and σ^2. I assume that the price of product 2 is zero in the first two periods, and 1 in the 3rd period, that consumer price sensitivities are equal to 1, and that consumers know the path of future prices. The lines on the bottom of Figure 1 are the level curves of the probability of experimenting. The set of consumers who are on a level curve are the set of consumers who have the same probability of experimenting for certain values of γ_i^0 and σ^2. We can see that if we take a consumer with some value of γ_{i0} and a low σ^2, a consumer with a higher σ^2 and a somewhat lower γ_i^0 will have the same probability of experimenting. There needs to be some way to tell apart these consumers in order to separately identify the population distributions of γ_i^0 and σ_{ij}^2.

The intuition behind how this is done is that σ_{ij}^2 will affect how consumers respond to changes in future prices. In the case where the $\eta_i = 0$, consumers with values of σ_{ij}^2 close to zero will be very unresponsive to future price changes compared with consumers who have higher values of σ_{ij}^2. In Figure 2, I show the probability of experimenting when the price path is 0,0, and 1 minus the probability of experimenting when the price path is 0, 1 and 1. These probabilities are very close when σ_{ij}^2 is close to zero, but are large when it is greater. The direction of the response depends on the value of γ_{i0}: it is positive for consumers with low γ_{i0}'s and negative for those with high γ_{i0}'s.[14]

Figure 3 shows an overlay of the level curves from Figures 1 and 2. The level curves from Figure

[14] In my thesis work, I demonstrate that the option value of learning can be increasing or decreasing in future prices.

1 are shown in solid lines, and the probability associated with the given level is labeled. The level curves from Figure 2 are shown in the dotted lines. They are not labeled, but the important thing to notice about them is them is that the level curve for 0 is the curve that starts in between $\gamma_i^0 = 0.5$ and $\gamma_i^0 = 1$ on the γ_i^0 axis. In the area below that line, the level curves increase as σ^2 increases, and above that they decrease. In general, it appears that for a given value of the probability of experimentation, the response to future price changes increases as we increase σ^2. As an example, consider the set of consumers for which the probability of experimenting with the new product is 0.4, which is shown by the line 0.4 in Figure 3. We can see that as σ^2 increases, the response of these consumers to future price changes also increases. Thus, even though all these consumers will have the same probability of experimenting, we can tell apart consumers with high σ^2's and low σ^2's by their response to future price changes. Unfortunately, there are some consumers we cannot tell apart by their response to future prices. For example, take the set of consumers whose probability of experimenting with the new product is 0.8. The level curve for which the response to future prices is 0 crosses this line twice, and we cannot tell apart the consumers at the points where it crosses. In my thesis research (Osborne 2006) I argue that the set of consumers that we cannot tell apart by this method is of zero measure.

In the data set we observe variation in future prices that is qualitatively similar to this example; the reason for this is that there is introductory pricing for the three new products. Some consumers will arrive to the market right after the new product introduction, when future prices will be low, and some will arrive near the end of the introductory pricing period, when future prices will be high. Increasing the variance in σ_{ij}^2 will make some consumers more responsive to future price changes, and some consumers less responsive to future price changes. The population response to a change in price will be a function of the moments of the distributions of γ_{i0} and σ_{ij}^2. Unfortunately, I have not been able to find a population moment that is a monotonic function of the variance of σ_{ij}^2. However, I believe that the arguments presented so far, which demonstrate that households with different γ_{i0}'s and σ_{ij}^2's behave differently, suggest that identification of these parameters is possible. Moreover, as I will discuss in Section 5.1, my model estimation procedure produces an empirical density for the variance of σ_{ij}^2 for the new products, conditional on the data. I show in that section that the density has a small variance and thin tails, which suggests that there is enough information in the data to identify the parameters in question.

4 Estimation Procedure

4.1 Selection of Household Sample

Although there are 1693 households in the total sample, I remove roughly two thirds of them from the sample before estimation, leaving a final sample of 519 households. There are two main reasons for reducing the sample so much: first, and most importantly, for computational reasons and second, to address data issues. The full structural model takes more than 1 week to estimate, so many of the households I remove were randomly dropped from the sample. Some households were dropped due to data issues, such as lack of observations. More than 80 percent of the dropped households were removed for these reasons. The other households who were removed from the sample were taken out for a third reason: they were households who were thought to be prone to inventory behavior. There are two reasons for removing these types of households: first, modeling inventory behavior structurally is computationally difficult (see Erdem, Imai, and Keane (2003) for an example), and adding this element to my model of learning and switching costs would make the model computationally intractable. Second, stockpiling behavior can potentially lead to a source of bias that is similar to that caused by ignoring heterogeneity in price sensitivities. Since new products are introduced at low initial prices, some consumers may be induced to purchase them simply in order to stockpile. These consumers will likely purchase something else when the new products are more expensive and they need to buy detergent again.

Household sample selection is performed as follows. The full sample includes 1693 households. First, I remove households who are likely to be extremely price sensitive, and who will stockpile the products. Hendel and Nevo (2006a) examine stockpiling behavior in laundry detergents, and find that suburban families, non-white families and households with 5 or more members are more price sensitive.[15] In my data set, it was not possible to determine whether a household was suburban or not. Data was collected on the type of residence where the household lived, and about 85% of households live in a single family home. 94% of these households owned the home rather than rented it. If all these families were suburban, it would not be feasible to cut them out. Household racial characteristics were collected in the data set. I constructed a household race variable which was defined to be the race of the male head of household if there was a male head of household, and the female head of household otherwise. I found that over 99% of households were white. For household size, there were 211 households with 5 or more people in them. I remove all these households from the sample.

In my thesis research (Osborne 2006), I examine whether 5 person households were indeed more likely to stockpile than other households. First, I examine time series patterns in the quantities purchased by households. Households that do not stockpile will likely be less responsive to sales in stores, and will be more likely to repeatedly purchase the quantity that suits the household's needs. To examine this, for each household in the sample I calculate the standard deviation of the quantity purchased by the household over the household's purchase history. I find that the standard deviation of quantity purchased by 5 person households is larger than for smaller households, and this difference is statistically significant.

[15] The data set used in that paper is different from the one I use.

Another way to tell if these households are stockpiling is to look at how much they buy in response to sales in the detergent product category. We would expect households who stockpile to purchase larger quantities in response to store price drops. To examine this, I regress the quantity purchased by a household during a given week on the minimum store price in the laundry detergent category that week, the price interacted with a dummy variable for whether the household has 5 or more persons in it, the next week's price, feature, display and household inventory. Household fixed effects were also included in the regression. The coefficient on the interaction between price and household size of 5 is negative and significant.

A third way to tell if 5 person households are stockpiling they should be more likely to make a purchase in the laundry detergent category in response to a store price drop. In the data set, household shopping trips were recorded, so it was possible to tell if a household visited a store in a given week but did not buy anything. To estimate a household's sensitivity to store price drops, for each household in my data set I estimate a binary logit model where the dependent variable is 1 if the household makes a purchase in a given week, and the independent variables are the minimum price observed in the store during the current week, the minimum price in the next week, a measure of the household's inventory, and two dummy variables for whether any products were on feature and display during the current week. I find that households with 5 or more people in them have lower price coefficients, but the difference is not large or statistically significant.

I also examine a number of different statistics to verify that the households I removed actually had the potential to impart the positive bias described above. I examine this using the two testable implications which were described in Section 3.3. Recall that the first testable implication was that the share of consumers who purchase the new product and then do not minus the share who do not and then do is increasing in in the amount of learning. If 5 person households are imparting positive bias to the sample, then including them should make this test statistic larger in magnitude. I computed the test statistic for both the full sample and the sample with less than 5 households in it and find that this is the case. The second testable implication can also be used to examine this. The share of consumers who repurchase the new product should increase more if there is more learning. The results of this test are not sensitive to the choice of sample.

A potential problem with this procedure is that there may still be significant stockpiling in the sample that is left. All households likely stockpile to some degree, so it may be difficult to remove this source of bias entirely. To examine whether this bias still exists in the sample I have left, I examine whether there is less learning among households who are less likely to stockpile. In particular, I compute the share difference test statistic for households that have only 1 person, or are in the highest income bracket. I find that the share difference test statistic for smaller and higher income households does not vary systematically across products. Another problem with selecting out 5 person households is that this demographic variable may be correlated with consumer tastes, leading to sample selection bias. In my estimates of reduced form demand models, I find that taste coefficients, coefficients on the previous product purchase, and coefficients on prices do not change very much when I remove 5 person households from the sample. Thus, the impact of removing these households on these quantities is likely to be small.

Removing 5 person households might not be the only method of selecting a sample of households that are not prone to stockpiling. Another way is to remove households who are sensitive to price drops in the laundry detergent category. Recall that I estimated binary logit models for each household in the sample to examine whether 5 person households were more sensitive to category price drops than smaller households. An alternative sample selection criterion I examine is to remove households who have estimated price coefficients less than 1 standard deviation below the population mean price coefficient. There are 148 such households. I find that the share difference test statistic decreases when these households are removed, which suggests they may also be subject to the problem of bias. The tenure test is not impacted by changing the sample. A disadvantage of this selection criterion is that it is less transparent than removing 5 person households. The estimates of reduced form demand models produced using this selection criterion look similar to those produced by removing 5 person households, so it will probably not matter which one is used.

Another possible selection criterion is to examine household interpurchase timing behavior. Households who do not stockpile likely wait until they run out of detergent to purchase more of it. To examine this, for each household I compute a consumption rate, which is the total number of ounces the household purchased in the three year period, divided by the number of weeks. Then, at each purchase event I predict the week when the household will use up the amount they purchased on that occasion, assuming a constant consumption rate. I compute a measure of the error of this prediction on a household by household basis by taking the average of the absolute difference between the predicted week of the next purchase and the actual week of the next purchase. I remove households whose error is larger than the population mean of the error, which was 7.43. There are 425 such households. I find that when I use this sample selection criterion to estimate the share difference test, the share difference increases. This is the opposite of what should happen if these households are indeed stockpilers. The tenure test is not affected by this sample. I also find that households with more predictable purchase behavior are more sensitive to category price drops, and that they tend to be larger households. This suggests that error in predicted interpurchase timing is not a good proxy for how likely households are to stockpile.

After removing 5 person households, I remove households where there appeared to be issues with data. Households that made more than 50 purchases in the three year period were removed. Also, any households who made less than 5 purchases were removed from the sample. It is probable that these households were purchasing at stores which were not included in the survey, and I felt it was best to remove them since we might not be observing many of their purchases. 188 households were removed for these reasons.

In addition to removing households from the sample, some individual purchase events needed to be removed from the sample. Recall that in my model, an observation is a household purchase of one of the 49 possible brand-size combinations. Any purchase events of products outside these categories were removed from the sample. Purchase events where the household purchased different products in the same week were also removed. Purchase events where the only product available in the store was the product purchased were also removed from the sample. Any households whose first purchase of one of the new products was removed for these reasons were dropped from the sample. Also, for the share differences test to work (and for the identification arguments laid out later in the paper to apply), we need to observe a household's first and second purchase events after the new product introduction. Any households for whom we do not see at least 2 purchase events in either the window of time after Cheer's introduction and before Surf, or the time after Surf and before Dash, or the time after Dash, are removed from the sample. 238 households were dropped for these reasons. Of the households who were kept, 12.6% of purchases were removed, about half of those being multiple brand purchases. Of the 1066 households remaining, one half of these households were dropped randomly to ease computational burden. The sample of households remaining was 519.

4.2 Coupon Parameters

Before I discuss in detail the estimation procedure, I wish to discuss an issue that arises in estimation due to the inclusion of coupons. In my model, I assume that the price of a product j to a consumer is the shelf price, p_{ijst}, minus the value of a coupon c_{ijt}. Coupons present an estimation difficulty: in my data set, I only observe whether a consumer has a coupon for the particular product that she purchases in a given purchase event. We do not observe whether the consumer has a coupon for any other products at that time. I overcome this problem by treating any coupons for products that the consumer did not choose as unobservables.

I assume that for each purchase event every coupon c_{ijt} for a non-purchased product (one for which $y_{ijt} = 0$) received by the consumer is drawn from the same distribution as consumer expectations about future coupons that is described in Section 3.2; hence, consumer expectations about future coupons are rational. To summarize the notation developed in that section, recall that the c_{ijt} for a non-purchased product is composed of two random variables, the binary random variable $\overline{c}_{ijt}$ which is 1 if the consumer receives a coupon for product j, and v_{ijt}, which is the value of the coupon received. Then the variable c_{ijt} is equal to $\overline{c}_{ijt}v_{ijt}$, and the vector of population-fixed parameters, θ, contains the parameters p^0_{cj}, p^1_{cj} for Cheer, Surf and Dash, and $p_c^{Cheer,1}$, $p_c^{Surf,1}$, and $p_c^{Dash,1}$.

This specification is a first approximation to solving the problem of unobserved coupons and represents a step forward from most papers that estimate discrete choice dynamic programming problems. The procedure I use is similar to Erdem, Keane, and Sun (1999), who also propose a discrete distribution for the probability a consumer has a coupon on hand for a non-purchased product, and estimate the parameters of the distribution. Note that there is more than one explanation for why a consumer might have or not have a coupon on hand for a non-purchased product. It could be that no coupon was available for the product, or it could be that a coupon was available but the consumer found it too costly to search for it and cut it out. The scanner data does not contain information on coupon availability and how likely a consumer was to search for coupons, so there is no way to separate these explanations. There is also a subtle endogeneity issue that could arise with coupon use: consumers could be more likely to search for coupons for products for which they have high tastes. I do not take this source of endogeneity into account, and to my knowledge this problem has not been addressed in scanner data research.

A more difficult issue with estimating the coupon parameters is that it may be difficult to separately identify p_{cj}^1, which is the amount that the probability of getting a coupon for the new products differs in their introductory periods, from the learning parameters. To see why, recall that introductory pricing can cause patterns in purchase behavior which look like learning. Introductory couponing may also have the same effect: if a lot of coupons for one of the new products are available right after its introduction, consumers will be induced to purchase the new product sooner rather than later, which will look like learning. Obviously, if we observe the entire distribution of coupon availability then there will be no identification problem - we can treat coupons just like prices. Since we are estimating the probability a consumer gets a coupon for a new product, it may be difficult to tell whether or not consumers are likely to make an initial purchase of the new product because the option value of learning is high, or because the likelihood they have a coupon for it is high.[16]

There are three things that help the identification. First, for some consumers the first purchase events after the new product introduction will occur when $n_t = 0$. Given that the coupon probabilities when $n_t = 0$ can be estimated from the period when most consumers have learned, if the probability of making a first purchase of the new product when $n_t = 0$ is higher than it should be, then that difference will pin down σ_{ij}^2. Second, some consumers will experiment with the new product when $n_t = 1$, and will make a second purchase when $n_t = 1$. For these consumers, their purchases will be pinned down by parameters we have already estimated - the state dependence and taste parameters. Hence, if the likelihood of them purchasing the new product is higher than it should be, this will raise the probability that they got coupons for the new product. Third, since we observe coupon use for a product when consumers purchase it the probability of receiving a coupon for the product will be bounded. As an example, suppose that during Cheer's introductory period 10 percent of all purchases involve a Cheer coupon, and 50 percent of Cheer purchases involve a coupon. The probability of receiving a coupon for Cheer will not likely be lower than ten percent, and not likely be higher than 50 percent, since 50 percent of the consumers who purchased Cheer

[16]Further, if $n_t = 1$, raising the probability a consumer gets future coupons will raise the value of purchasing the new product when there is no learning and only switching costs.

did not have (or use) a coupon for it.

4.3 The Markov Chain Monte Carlo Estimator

I estimate the structural model described in the previous section using Markov Chain Monte Carlo, which is abbreviated as MCMC. MCMC methods are Bayesian methods, which differ from classical methods in that they do not involve maximizing or minimizing a function. In models with high dimensional unobserved heterogeneity, like the one I have specified, maximization of a likelihood function can be numerically difficult. Bayesian procedures proceed differently: the researcher must specify a prior on the model parameters and then repeatedly draw new parameters from their posterior distribution conditional on the observed data.

Drawing from the posterior is made easier using an MCMC procedure called Gibbs sampling, which involves breaking the model's parameter vector into different blocks, where each block's posterior distribution, conditional on the other blocks and the observed data, has a form that is convenient to draw from. Gibbs sampling proceeds by successively drawing from each parameter block's conditional posterior. This procedure results in a sequence of draws which converge to draws from the joint distribution of all the model parameters. The initial draws in the sequence are discarded, and remaining draws from the converged distribution are used to calculate statistics of model parameters, such as mean or variance.[17] My underlying demand model is the random coefficients logit model, with two differences: the coupon parameters and the value function solution. Thus, the setup for my Gibbs sampler is very similar to that used to estimate the random coefficients logit model. This estimator is well understood and is described in Train (2003), pgs xxx - xxx.

To form the conditional posterior distributions for the blocks of parameters it is necessary to impose a prior distribution on some of the model parameters. I assume flat priors on θ, a normal prior on b which I denote $k(b)$, and inverse gamma priors on the elements of the diagonal matrix W, which I denote as $IG(W)$. The posterior distribution of the model parameters will depend on the parameters' prior distribution and the probability of the data given the parameters. The priors on the p^0_{cj}'s are uniform on [0,1], the priors on p^1_{cj} are uniform on [-p^0_{cj},1-p^0_{cj}] for Cheer, Surf and Dash, and the priors on $p_c^{Cheer,1}$, $p_c^{Surf,1}$ and $p_c^{Dash,1}$ are uniform on [-$\min_j\{p^0_{cj}\}$,1-$\max_j\{p^0_{cj}\}$]. The priors on b and W are assumed to be non-informative, so that $k(b)$ has zero mean and infinite variance. The prior on W is also chosen to be non-informative, so that the scale is set to 1 and the degrees of freedom approaches 1. The posterior distribution of the model parameters will depend on the

[17] Determining when the sequence of draws produced by the Gibbs sampler has converged to draws from the joint posterior distribution is difficult, which is a tradeoff of Bayesian methods relative to classical methods. The simplest approach is for the researcher to observe the sequence and to see the draws trending towards the posterior. After convergence the draws will traverse the posterior. A more formal method of testing for convergence is suggested in Gelman and Rubin (1992), who propose running the Gibbs sampler from several different starting points and testing whether the posterior means calculated from the converged sequences are equal across runs.

parameters' prior distribution and the probability of the data given the parameters.

The probability a consumer chooses a particular product in purchase event t, given her preferences and the values of observables, can be expressed using a simple logit formula. Denote d_{ijst} as the variable that is 1 if consumer i chooses size s of product j in purchase event t. Denote d_{it} as the vector of observed d_{ijst}'s, c_{it} as the vector of c_{ijt}'s, x_{it} as the vector of x_{ijt}'s and v_{ijst} as the consumer's flow utility minus the logit error. The probability of the consumer's choice in purchase event t will be

$$Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) = \sum_{(j,s)\in J_{it}} d_{ijst} \frac{\exp(v_{ijst} + \delta EV(\Sigma_{it+1}; \theta_i, \theta))}{\sum_{(k,l)\in J_{it}} \exp(v_{iklt} + \delta EV(\Sigma_{it+1}; \theta_i, \theta))}. \tag{13}$$

Denote $g(\theta_i|b, W)$ as the density of an individual level θ_i and $Pr(c_{it}|\theta)$ as the probability of a particular c_{it}. Then the posterior density of the parameters is proportional to

$$\begin{aligned}\Lambda(\theta_i \forall i, b, W, c_{it} \forall i \text{ and } t, \theta) \propto & \prod_{i=1}^{I} \left[\prod_{t=1}^{T_i} \left\{ Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) Pr(c_{it}|\theta) \right\} g(\theta_i|b, W) \right] \\ & \cdot k(b) IG(W)\end{aligned} \tag{14}$$

I draw from this posterior in 5 different blocks, where each block has a functional form that is convenient to draw from. I will describe these formulas briefly in the next few paragraphs. More details on the specifics of the Gibbs steps are given in detail in the Appendix.

The first block draws θ_i for each household conditional on the d_{it}'s, the c_{it}'s, b and W. Because of the assumption that the error term is logit, the conditional posterior likelihood of a particular vector of θ_i is proportional to $\prod_{t=1}^{T_i} \left\{ Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) \right\} g(\theta_i|b, W)$. This distribution is not conjugate, which means that the Metropolis-Hastings algorithm (see Appendix A.1 for the steps I use to implement this) must be used in this step.[18]

In the second step, I draw a new vector of fixed parameters, θ. The posterior distribution of θ conditional on θ_i, the $\bar{c}_{ijt}$'s, v_{it} and the d_{it}'s is

$$\prod_{i=1}^{I} \prod_{t=1}^{T_i} \left\{ Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) Pr(c_{it}|\theta) \right\}. \tag{15}$$

This distribution is also not conjugate and the Metropolis-Hastings algorithm must be used to draw from it.

The third step draws a new b vector conditional on $\tilde{\theta}_i$ for $i = 1, ...I$ and W. The conditional posterior distribution for b is normal, so this step is straightforward. Similarly, the conditional posterior of the elements of W given $\tilde{\theta}_i$ for $i = 1, ...I$ and b are inverse Gamma, which is straightforward to draw from. For unobserved coupons, each $\bar{c}_{ijt}$ is drawn separately across households, products and purchase events, and has a Bernoulli posterior distribution conditional on v_{it}, θ_i, θ and d_{it}.

[18]Note that when we perform this step, we will need to evaluate the consumer's expected value function in Equation (13), $EV(\Sigma_{it+1}; \theta_i, \theta)$. The procedure I use to do this is described in Section 4.4.

4.4 Value Function Solution

The method of Imai, Jain, and Ching (2005) works in conjunction with the Gibbs sampler to obtain a solution of the value function. In this section I will broadly describe how I solve for the value function in Equation (13) and how the method works. The innovation of this new method is that discrete choice dynamic programming problem is solved only once, along with the estimation of the model parameters.

Recall that in the Gibbs sampling algorithm described in the previous section, we draw a sequence of model parameters that converges to draws from the parameters' joint distribution. The basic idea of the value function solution method can then be broken up into two steps. First, at a particular point g in sequence, draw small number of values of the unobservable and calculate expected utility at all state space points. The expected utility and the current parameter value are then retained for use in later iterations of the MCMC sequence. In order to calculate expected utility at some point g in the sequence, it is necessary to have an approximation of the value function at the current parameter value. In the second step, the value function is calculated as a weighted average of previously retained expected utilities, where the weights are kernel densities of the difference between the current parameter and the previous saved parameters. In actual implementation these steps are performed in reverse order: first the value function is interpolated at the current parameter draw, and then the expected utilities are calculated. However, I believe it is easier to understand the algorithm by looking at the steps in the order I have laid them out, rather than the order in which they are executed. In the following paragraphs I will describe these two steps in greater detail.

Consider the first step, which is to draw some values of the model's unobservables and calculate expected utility. This calculation is done at points in the state space, $\Sigma = (s, p, J, y, n)$, and the expected utilities and current parameter value are retained. There are two different sets of unobservables which are unobserved to the consumer at the time she makes her purchase decision, and must be integrated out when the value function is formed: the ε_{ijst}'s, and the consumer's future tastes for products she has not yet purchased, the γ_{ij}'s. Integrating out the ε_{ijst}'s does not require numerical approximation: because of the assumption that they are logit errors, the consumer's expected utility has a closed form solution, conditional on θ_i, θ, and future coupons. This is not true when we integrate out the future γ_{ij}'s and c_{ijt}'s, so these must be approximated numerically. As an example, let us consider constructing an analogue to the consumer's expected value function in Equation (12), which is the value at state space point $s_j = 0$, $y_j = 1$ for some new product j. First I draw $L = 10$ draws from the true taste distribution for product j, which is $N(\gamma_{ij}^0, \sigma_{ij}^2)$, and from the coupon distribution implied by θ. To calculate the expected utility, we need to calculate first each consumer's exact utility (ignoring the logit error) at each product at simulation l. Denote the lth taste draw as γ_{ij}^l and the lth coupon draw as c_{ij}^l, and denote θ_i^l as the vector of θ_i with the consumers true taste for product j (γ_{ij}) taken out and replaced with the simulated tastes (γ_{ij}^l). Assume that we have an approximation of the expected value function at point n of the sequence for next period's state space point, $\Sigma' = (s', p', J', y', n')$, which I will denote as $E_{(p',J')|(p,J)}V_n(s', p', J', y', n'; \theta_i^l, \theta)$.[19] Then the consumer's utility for product j at simulation l, v_{ij}^l, will be

$$\begin{aligned} \text{Product } k = j \quad &: v_{iks}^l = && \gamma_{ik}^l + \xi_{is} - \alpha_i(p_{ks} - c_{ik}^l) + \eta_i y_k \\ & && +\delta E_{(p',J')|(p,J)}V_n(S', p', J', y', n'; \theta_i^l, \theta) \\ \text{Product } k \neq j \quad &: v_{iks}^l = && \gamma_{ik}(S_k) + \xi_{is} - \alpha_i(p_{ks} - c_{ik}^l) + \eta_i y_k \\ & && +\delta E_{(p',J')|(p,J)}V_n(S', p', J', y', n'; \theta_i^l, \theta), \end{aligned} \tag{16}$$

which corresponds to Equation (11).

Her expected utility for purchasing product j for the first time (state space point $y_j = 1, s_j = 0$) at the individual i's θ_i is then calculated as

$$\hat{EV}_g(S, p, J, y, n; \theta_i, \theta) = \frac{1}{L}\sum_{l=1}^{L} \ln\left(\sum_{(k,s)\in J} \exp(v_{iks}^l)\right). \tag{17}$$

The second step of the algorithm is to calculate the approximation of the value function at the parameter draw for the current point in the sequence, g. Denote consumer i's individual level parameters at this iteration as $\theta_{i,g}$, the population-fixed parameters as θ_g, and the vector of $\theta_{i,g}$ stacked on θ_g as $\overline{\theta}_{i,g}$. Recall that at each point in the sequence, the expected utilities calculated in the first step are retained along with the parameter draws. Assume that at iteration g we have retained $N(g)$ previous parameter draws and expected utilities, and we want to calculate the expected value function at $\theta_{i,g}$. This is then calculated as

$$E_{(p',J')|(p,J)}V_g(s, p, J, y, n, \theta_{i,g}, \theta_g) = \frac{\sum_{r=1}^{N(g)} \left[\hat{EV}_r(s, p, J, y, n; \theta_{i,r}, \theta_g)\right] k((\overline{\theta}_{i,g} - \overline{\theta}_i^r)/h_k)}{\sum_{i=1}^{N(g)} k((\overline{\theta}_{i,g} - \overline{\theta}_{i,r})/h_k)}, \tag{18}$$

where $k(\cdot)$ is a kernel density function and h_k is a bandwidth parameter, and $\hat{EV}_r(s, p, J, y, n; \theta_{i,r}, \theta)$ is the rth retained expected utility. The approximated value function is used to calculate the utilities in Equation (16).

5 Estimation Results

The main estimation results are shown in Table 5. Recall that in my model, the coefficients of consumer i's flow utility are broken up into two groups: those that vary across the population, denoted θ_i, and those that are fixed across the population, denoted θ. The population-varying

[19]Since the state space is quite large, and computer memory is limited, I only evaluate the value function at a subset of the state space points, and interpolate it everywhere else. The details of this procedure, as well as other computational details associated with the value function solution, are described in the Appendix.

coefficients are normally distributed across the population with mean b and diagonal variance matrix W. The Markov Chain Monte Carlo estimator produces a simulated posterior distribution of b, W, and the fixed parameters, θ. The two columns of estimates under the headings "Mean" show the means and standard deviations (shown in parentheses) of the simulated posterior density for each element of b; similarly, the two columns of estimates under the headings "Variance" show the mean and standard deviation of the simulated posterior for W. Estimates of parameters that are fixed across the population, θ, are shown under the "Mean" heading; the corresponding entries under the "Variance" heading are dashed for these parameters. Although the numbers in the table are means and standard deviations of parameter posterior densities, they can be interpreted in the same way as the estimated coefficients and standard errors that are produced by classical methods.

Consider the first block of estimates, labeled "Taste parameters". The first 9 rows show the estimated tastes for each established product. The liquid Other product is normalized to 0, and the Other Powder, Tide Powder and parameters associated with switching costs are fixed across the population. The first element of the first row shows the population average of consumer tastes for liquid Era, which is -0.908. It may look like people like Era less than the Other product, but this is not the whole story. The fourth column shows the variance in tastes for Era across the population, which is 3.258. This variance is large, which indicates that consumers are very heterogeneous in their taste for Era: some consumers like it a lot, and some do not like it very much at all. The results are very similar for almost all the established products: the mean tastes are negative, and most of the variances are high, so there is a lot of heterogeneity in tastes. Consumer heterogeneity in tastes is very important in this market, which is consistent with these products being experience goods. It is also consistent with important heterogeneity in factors such as the types of fabrics in a household's wardrobe, the types of soils and stains that need to be cleaned, the water temperature used, the household's washing machine quality, and the types of scents the household prefers. The next six rows show the taste parameters for the different size categories.

Skipping the last three rows of the taste parameters section, which will be discussed later, consider the second block of estimates in the table, under the heading "Learning Parameters". The first row of this section shows the estimated population mean and variance of consumer's expected tastes for Cheer, γ_{ij}^0. The population average predicted taste for Cheer is -0.518, and this estimate is statistically different from zero. The population variance of predicted tastes is statistically significant, but small relative to the mean at 0.356. This means that there is not a lot of heterogeneity in how much consumers expect to like Cheer: most of them don't expect to like it very much, and most consumers do not have a very good idea of how much they will like the product in advance.

Consider the next three parameters, which correspond to the consumer's uncertainty about her true taste for Cheer. The mean of the parameter σ_{0ji}^2, the intercept, is precisely estimated at -0.58, while the parameters on household size and income are negative and statistically significant. The positive coefficients suggest that the amount of variance in true tastes is higher among larger and higher income households. Recall that the actual consumer uncertainty in tastes is a transformation of these parameters (as specified in Equation (5)). As an example, for a household of income 3 and size 3 that has population average value of σ^2_{0ji}, the variance in her true taste for Cheer is $5\frac{\exp(-0.58+0.148\cdot 3-0.068\cdot 3)}{1+\exp(-0.58+0.148\cdot 3-0.068\cdot 3)}$, which is about 2.08. If the consumer's expected taste for Cheer is -0.58, the population average, then her true taste will be drawn from a $N(-0.58, 2.44)$. Her true taste distribution looks very similar to the taste distributions for the established products. The results for Surf and Dash follow a similar pattern to those of Cheer.

In summary, there are two important facts about the learning parameters: first, the variance across consumers in γ_i^0 is low. Before consumers make their first purchases of the new product, their expectations are similar. Second, the variance in their true tastes is large, which indicates that after consumers make their first purchases of the new products, they are very different in how much they like it. These facts are consistent with these products being experience goods: consumers need to purchase and consume the product in order to find out how much they like it.

Let us return to the last three rows of the first block of parameter estimates. This shows the estimates for the coefficient on y_{ijt-1}, which is η_i. The intercept for η_i, η_{i0}, is allowed to vary across the population. Its mean is 1.311, and its variance is large at 2.002. The distribution of η_i across the population will depend on two things: the distribution of unobserved heterogeneity, which is normal, and the distribution of demographics. Taking both of these into account, the expected value of η_i in the population is 1.32, and its variance is 1.98. This means that most households have switching costs, but a small portion of them are variety-seeking (about 84% of them have values of η_i that are positive). The coefficient on household size, η_1, is negative and the one on household income, η_2, is positive. The fact that switching costs is increasing in income is consistent with the idea that the switching cost is generated by a cost of recalculating utility: for high income consumers, time is likely more valuable and the cost of recalculating utility may be higher. There are a few explanations for the negative coefficient on household size. One is that larger households may switch brands more easily since detergent is a smaller part of household consumption. A second explanation for the negative coefficient on household size could be due to within household heterogeneity in tastes. Different members of a household may have different tastes, and since the data does not record consumption by different members, brand switching among households with more members may be due to different members purchasing different products rather than variety-seeking. Similar findings are documented in Che, Sudhir, and Seetharaman (2006).

The fact that most consumers have switching costs has interesting implications for pricing policy. As an example, suppose that it has been a long time since the introduction of Dash, so that most consumers have experimented with all the three new products. Suppose that Unilever decides to temporarily drop the price of Wisk. Procter and Gamble might worry that this price drop could decrease the market share of Tide in the intermediate run. The price drop will draw consumers away from Tide who will have a cost of switching back to Wisk. It would be optimal for Procter and Gamble to respond with a subsequent price drop in order to get them back.

The last block of parameters shows consumer responses to the exogenous variables prices, features and displays. The parameter for consumer price sensitivities is constructed in the same way as for the learning parameters (Equation (6)). The population average value of the price coefficient is about -24.9. The parameter on household income is positive while that of household size is negative, which means that higher income and larger households are more price sensitive. The fact that the coefficient on income is positive is surprising, but the magnitude of the coefficient is very small. The variance in the intercept of the price coefficient is quite large, which indicates that there is substantial heterogeneity in price sensitivity. The population distribution of the price coefficient is shown in Figure 4. The distribution of price coefficients is right skewed, and somewhat less than half of all households have price coefficients that are less than 5 (about one quarter have price coefficients that are less than 1). The estimates of the coupon sensitivity parameter, α_{0ic}, show that its mean is -0.673 and its variance is 0.229. Recall that the coupon sensitivity coefficient that enters the consumer flow utility, α_{ic}, is a transformation of α_{0ic}, $\frac{\exp(\alpha_{0ic})}{1+\exp(\alpha_{0ic})}$ (Equation (7)). The population mean of α_{ic} is 0.35, and its variance is 0.01, so there is very little heterogeneity in consumers' sensitivities to coupons. The feature and display variables are both positive on average in the population, which is to be expected.

Table 6 shows the parameters of the coupon distribution. The first column of the table shows the mean of the posterior draws of the p_{cj}^0's, the p_{cj}^1's, and the $p_c^{Cheer,1}$, the $p_c^{Surf,1}$, $p_c^{Dash,1}$; the second column shows their standard deviation. Almost all the mean parameters are precisely estimated. To see how to interpret the parameters, recall that the p_{cj}^0's are the probability that a consumer receives a coupon for product j after the "introductory pricing" period. So the probability a consumer gets a coupon for Tide Liquid is 0.263. The parameters under $n_t = 1$ are added to the $n_t = 0$ parameters during introductory pricing periods. So the probability of a consumer getting a coupon during the introductory period for Surf Liquid is $p_{cj}^0 + p_{cj}^1 = 0.237 - 0.078 = 0.159$. The probability a consumer gets a coupon for Tide Liquid during the introductory period for Surf Liquid is $p_{cj}^0 + p_{cj}^{Surf,1} = 0.263 - 0.038 = 0.225$.

5.1 An Examination of Consumer Uncertainty About the New Products

In this section I will examine two aspects of consumers' uncertainty about their true tastes for the three new products. First, I will examine how consumer uncertainty varies across the population. Recall from the previous discussion that consumer i's uncertainty about her true taste for a new product j, σ_{ij}^2, is a transformation of the three parameters in the second block of Table 5, σ_{0ij}^2, σ_{1j}^2 and σ_{2j}^2, and the consumer's household income and size. Heterogeneity in consumer uncertainty about product j will come from two sources: unobserved heterogeneity in the random coefficient σ_{0ij}^2, and observed heterogeneity in household demographics. I will demonstrate that across the population as a whole, there is not a lot of variance in the σ_{ij}^2's. I will also show there is not a consistent pattern in the σ_{ij}^2's across demographics. Second, I will examine the effect of removing consumer uncertainty on the market shares for new products. I will demonstrate that removing consumer uncertainty substantially increases the overall market share for a new product, and the impact is largest for niche products.

The first column of Table 7 shows the average value of σ_{ij}^2 in the population for each of the three new products, and the second shows the standard deviation across the population.[20] There are two important patterns to notice. First, we can see from the table that the average amount of uncertainty is greater for Cheer than for Surf, the first two liquid introductions observed in my data set. This may be due to the fact that these products are liquid detergents, and consumers' experience with Cheer helped them resolve some uncertainty about liquids as a product category. The amount of learning about Dash, the last liquid introduction in this data set, is a little bit lower than Surf, which may also be a result of consumers learning more from Surf's introduction. Second, we can see that the variance of the learning parameters is small, which indicates that the amount of learning does not vary a lot across the population. Recall that in the previous section, I showed that consumer's expected tastes for the new products also did not vary significantly across the population. These two facts together indicate that consumer expectations about their true tastes for the new product did not vary across the population by very much.

Table 8 shows the average consumer uncertainty broken down by household income and size. For Surf and Dash, the amount of learning is decreasing in income and household size. For Cheer, the amount of learning increases in household size until it reaches 4, where it drops. The amount of learning for Cheer is increasing in household income.

To examine the effect of learning on the market shares of the new products, I conduct the following simulation experiment. First, using the retained draws on θ_i and θ in each step g of the Gibbs sampler I simulate each consumer's product choice in each purchase event. The error terms and unobserved coupons observed by the consumer in each purchase event are drawn from their underlying distribution. I then calculate the overall market share for each product from the simulated choices, averaged over the g draws. The first column of Table 9 shows the average of this simulated market share over all the weeks that the product was available, and over the first 12

[20] When I compute the population distribution of σ_{ij}^2, I use the estimated individual level parameters, the θ_i's, rather than the estimated b and W, which are respectively the population mean and variance of the θ_i's. Recall that in a given step g of the Gibbs sampler, I draw the population-varying coefficients θ_i for each consumer i, and the population-fixed coefficients θ. In step g (assuming step g is retained), I calculate each consumer's uncertainty, $\sigma_{ij,g}^2$, using $\theta_{i,g}$, θ_g, and demographics for i (Equation (5)). I then calculate the population mean and variance of $\sigma_{ij,g}^2$. The numbers in the table are the average over draws of the mean and variance calculated in each step g.

weeks that each new product was available.

Then I run the same simulation setting $s_{ijt-1} = 1$ for all three new products: in this case consumer tastes for the new products are assumed to always be γ_{ij}. These simulated market shares are shown in the second column of Table 9, and are substantially larger than the shares in the first column: the market share of Cheer rises by 34%, Surf by 24%, and Dash by 58%. In the period right after the new products are introduced, the impact of removing learning is smaller: for Cheer, the market share rises by only 9%, 10% for Surf, and 33% for Dash. Why does this happen? The answer to this question is twofold. First, consider the short run (the first 3 months after the introduction), and assume that $\delta = 0$.

I refer the reader to Figure 8, which shows the estimated population distribution of tastes for Cheer before and after all learning has occurred. The thinner distribution is the population distribution of predicted means for Cheer (the γ_i^0's), or the tastes for consumers who have not yet learned about Cheer. This distribution is normal with mean of -0.518 and variance of 0.356 (Table 5). The flatter one is the population distribution of true tastes for Cheer, tastes after learning has occurred. This distribution is normal, and has mean of -0.518, and a variance of 2.59. The number 2.59 is the variance in γ_{ij}^0, 0.36, plus the average of σ_{ij}^2 across the population, which is 2.03.

A myopic consumer will experiment with Cheer when her prior draw is greater than her maximum utility for other products. In the figure, the line labeled $\delta = 0$ shows the cutoff for a consumer with average values of tastes for all products, assuming that there is no state dependence, prices for all products are the same, and the error terms are set to zero. The share of consumers who will experiment will be those whose prior is to the right of this line. We can see that the share will increase when consumers know their true tastes, since the area under the posterior curve is larger than under the prior.

Since I assume consumers are forward-looking, there will be an option value of learning, which will shift the cutoff to the left and result in more experimentation. I compute this option value of learning at the given parameter values (assuming the consumer has average tastes, and that there is no switching costs), assuming consumers expect prices to stay the same over time. This new cutoff is shown by the line $\delta = 0.95$; when the option value of learning is taken into account, significantly more consumers experiment. Although it is difficult to see from the picture, the shaded area to the right of $\delta = 0.95$ line on the expected tastes distribution is a little bit smaller than that to the right of the $\delta = 0$ line on the true taste distribution. This means that informing consumers of their true match values will cause an increase in the product's short run market share, even in when consumers are forward-looking. During the first three months after Cheer's introduction, the increase in Cheer's market share from removing learning increases from 9.1% to 10.0%, which is a 9% increase. In the intermediate run, the effect of giving consumers their true taste draws will be even greater. The consumers who will be affected by this will be those who have not yet experimented. The consumers who have experimented will tend to be those who have a high option value of learning, so the consumers who will be left will have a low option value of learning. Their behavior will be closer to consumers who are myopic. This explains why the increase in market shares over the intermediate run is 34%, which is much larger than the increase in the short run, which was only 9%.

Note that it is possible to do a similar exercise where I set the switching costs parameter, η_i to be 0 for all consumers to see the impact of switching costs on the new product market shares. The results of this exercise show that removing switching costs increases Cheer's market share by 23%, Surf's market share by 13%, and Dash's market share by 23%. This result is intuitive: when there are switching costs, consumers will find it costly to switch away from one of the established products at the time the new products are introduced, which will make them less likely to experiment with them.

One final issue to discuss is the identification of the population variances of the learning parameters. This was discussed to some degree in Section 3.3, where I argue that individuals with different values of σ_{ij}^2 can theoretically be distinguished. This does not imply that they can easily be distinguished in the population, which might mean that the variance of σ_{0ij}^2 is not identified. One method of assessing this is to examine the posterior densities of the variances of σ_{0ij}^2. If they look very similar to the prior densities, then we have not obtained identification. Figures 5, 6, and 7 show plots of the kernel density of the saved draws on the variances of σ_{0ij}^2 for each of the three new products. Recall that the priors on these parameters were assumed to be noninformative; in contrast, the posterior densities have easily visible global maxima and thin tails. The posterior density for Cheer, and to a lesser extent Dash, have second modes; these modes occur with less probability than the model's highest peak. This suggests that overall, there is enough information in the data to identify the variance of σ_{0ij}^2; an inverse gamma prior might be too restrictive, though, since it is unimodal.

5.2 Estimates from the No Switching Costs and No Learning Models

One of the important contributions of this research is that I estimate a model with two types of dynamics in it: learning and switching costs. In this section I will discuss how the model estimates differ when one of the two types of dynamics is left out. I will also discuss how the model fit changes when different dynamics are ignored.

First, I re estimated the model with the parameter on alternative dynamics, η_i, restricted to be zero for all consumers, so that the only dynamics in demand were learning.[21] The taste variances produced by the model were larger, which is to be expected. The price coefficient estimate is similar

[21]The tables of estimation results produced by this restricted model and the next one are not shown, but are available upon request from the author.

to that of the full model; the average of the price coefficient across the population is -26.9, which is slightly higher than the estimate produced by the full model. Consumers are estimated to be less sensitive to coupons than in the full model, and the feature and display variables look the same. The most interesting change is in the estimates of the learning parameters. The means of the γ_i^0's drop, but the pattern is the same - one average consumers expect to like Surf more than Cheer, and Cheer more than Dash. The variances in the γ_i^0's are significantly higher, which indicates that consumers are heterogeneous in how much they expect to like the new products. The estimates of the σ^2's, the amount of uncertainty consumers have about their true tastes for the new products, also change. On average, the σ^2 for Cheer is 0.909, for Surf it is 2.239, and for Dash it is 1.473. Thus, the model predicts that consumers are significantly more certain about their true tastes for Cheer and Dash, and less certain for Surf (although they are more heterogeneous *ex-ante*). Thus, the estimates do not give as strong support to the hypothesis that these products are experience goods. This result is intuitive when we consider the results associated with the share difference implication, which is what will identify the amount of learning (Section 3.3). Recall that the estimates of the full model suggest that switching costs plays an important role, and that the share difference test statistic is negatively biased in the presence of switching costs. Thus, it is not surprising that the estimates of this model suggest a less important role for learning.

The importance of learning in this restricted model can also be assessed by simulating the market shares for the new products when there is learning and when there is no learning, as was done for the full model. The results of this simulation are shown in Table 10. As expected, the impact of learning on the new product's market share is smaller in the no switching costs model than the no learning model. Hence, researchers who ignore switching costs will underestimate the importance of learning.

The second restricted model I estimate has the learning parameters, the σ^2's, restricted to zero for all consumers. Compared to the estimates from the full model, the no learning model generally tends to underestimate the amount of variance in consumer tastes. The population average of the price coefficient is -24.8, which is a little bit smaller than the full model; there is less estimated variance in the heterogeneity parameter for price sensitivities as well. Consumers are slightly more sensitive to coupons, and slightly less sensitive to features and displays. The estimates of the switching cost parameters also look very similar to those produced by the full model. The population mean of η_i is 1.37, and its variance is 2.04; both of these numbers are slightly higher than those produced by the full model. Overall, the results for this model look similar to the full model results, which is not entirely surprising. To see why, recall my model identification argument (Section 3.3). I argued that consumer taste distributions and switching cost coefficients will be identified from longer run behavior. In the no learning model, the first few periods after the new product introduction will have some impact on the estimates, but most of the model's identification will come from longer run behavior since we observe consumers for a long period of time after the product introductions.

In order to assess how each type of misspecification affects the model's predictive power, I compute the simulated market shares for each of the three models. Table 11 shows the actual market shares compared to the simulated market shares produced by each model. The final row of the table shows the average of the absolute difference between the predicted market share and the actual market share, where the average is taken across the 13 products. Overall, the full model does the best job at predicting market shares - on average, it will mispredict market shares by 0.26 percent. The model with no switching costs is almost as good as the full model, with an error of 0.33 percent, while the model with no learning is significantly worse, with an error of 1.8 percent. Much of this difference is due to the no learning model poorly predicting consumer response to new product introductions. To see this, consider the final four lines of the table, which shows the absolute average prediction error for market shares in the periods directly following the new product introductions, and for the final 63 weeks of data. During the periods after the new product introduction, the full model and the no switching costs model have similar prediction errors. There is almost twice as much error in the no learning model during these periods. During the final 63 weeks of the sample, the prediction error is very similar across all 3 models.

In summary, it appears that ignoring learning is a more serious misspecification error than ignoring switching costs, if the aim of the researcher is to simply fit market shares. Learning has a large impact on the market shares of products right after the introduction of a new product, so leaving it out will significantly reduce the model's predictive power.

5.3 Counterfactuals

In this section I will examine two important counterfactuals that I have computed: the effect of an introductory price cut for a new product on its intermediate run market share, and the effect of informative advertising on the new product's market share. There are two different ways I will compute each counterfactual. First, using the estimated results from the full model, I will compute the impact of the introductory price cut and introductory advertising when there is learning and switching costs, no switching costs, and no learning. This exercise allows me to explore the impact of dynamics on the impact of introductory pricing and introductory informative advertising. I will also compute these counterfactuals for the restricted models I discussed earlier. This will show how model misspecification affects the counterfactuals. These results will be of interest to brand managers who wish to better understand the effects of their pricing and advertising policies. They will also be of interest to government agencies that are interested in the impact of information provision about new products, and to industrial organization economists who wish to better understand the role for introductory pricing.

First let us consider the effect of an introductory price cut for each of the new products. I compute this counterfactual as follows. First, I set $x_{ijt} = 0$ and $c_{ijt} = 0$ for all i, j and t. For each product j, I set p_{ijt} to its average across all purchase events where the product is available. Thus, aside from the introductory price cuts, prices are the same across time and consumers. If there are any new product introductions after the new product for which I am calculating the price cut, I do not introduce them. I also assume that all other products are always available, so J_{it} does not vary across i and t, except for the introduction of the new product I am interested in. When I compute the counterfactuals, all 519 consumers make exactly one purchase in each period. Since the modal interpurchase time is 8 weeks, each period can be thought of occurring every two months. I compute the counterfactuals for ten periods, so the total length of time covered by these counterfactuals can be thought of as occurring over a twenty month period.

I then solve for every consumer's value function, assuming that they know the path of future prices, and simulate each persons's choice in each period.[22] This requires that I draw 49 new ε_{ijt}'s for each period. To reduce simulation error, I simulate each consumer's sequence of choices ten times and take the average of these choices. I simulate choices for each retained draw on θ_i and θ from the Gibbs sampler (a total of 750 times) under three different assumptions on the type of dynamics in demand: when there is both switching costs and learning, which is at the estimated parameters, when there is no learning, which means every consumer knows γ_{ij} from the beginning, and when there are no switching costs, which means $\eta_i = 0$ for all i. I also assume that there is no learning for any product other than the one for which I am examining the effect of the price cut; for example, if the price cut is for Surf, then I assume consumers know their true taste for Cheer.

I repeat this exercise for an introductory price cut for each new product, which means I cut the price of the new product by 10 percent in period 1, period 3, and period 5. The ten percent price cut is applied to all 4 sizes of the new product. This is a partial equilibrium counterfactual: all other prices are held fixed. I assume that consumers correctly view the price cut as temporary and expect the new product's price to rise to its average in period 2. The first column of Table 12 shows the percentage impact on period t marketshares of the period 1 price cut. For all three products, a ten percent price cut results in a greater than 10 percent increase in period 1 market shares, so own price elasticities are elastic. The intertemporal price elasticities are small. For instance, a ten percent price cut for Cheer only leads to a 0.87% increase in Cheer's period 2 market share. The impact on Cheer's market share drops as time increases. The second column show the response to a price cut that occurs in period 3, and the third in period 5. As time increases, the own price elasticity of each new product drops: in other words, when consumers learn about the new products their demand becomes more inelastic. This provides an alternative explanation for why we observe introductory pricing of the new products to the idea that firms recognize dynamics and are pricing with them in mind. It may be that firms are myopic and simply observe that short run demand

[22]The fact that there is learning about one of the new products means that I need to integrate over the distribution of future tastes for that product. I compute this integral using Gauss-Hermite quadrature. Also, it is necessary to pick a value for y_{it-1} in period 1. I assume that the period 0 choice is whichever brand gives the highest static utility.

elasticity is falling, and raise their prices in response.

The fourth, fifth and sixth columns show the impact of the price cut under the model parameters estimated for the full model when consumers are assigned their true tastes for the product. The own price elasticities are smaller when there is no learning. Generally, the intertemporal price response is larger when the learning is removed. The only place where this does not hold is for Dash in periods after the second, where the intertemporal price elasticity is slightly smaller. The intuition behind this result is that when there is no learning, consumers who respond to the price cut will continue to purchase the new product due to the cost of switching products; when learning is added to the price cut, some consumers who experiment with the product will receive a low match value for it and will switch away to something else. This can also be seen in the first three columns, where the intertemporal elasticity rises as time increases.

This result suggests that firms who wish to make their price cuts more effective at building future market share should combine them with informative advertising. This does not take into account the impact of advertising itself, which could increase or decrease revenues. The increase in revenues for the 10 simulated periods from the price cut alone was 13.14 dollars for Cheer, 18.41 dollars for Surf, and 9.64 for Dash. The increase in revenues from removing learning was 338.45 for Cheer, 122.68 for Surf, and 305.07 for Dash. Finally, the increase in revenues from the price cut combined with removing learning was 349.32 for Cheer, 139.16 for Surf, and 314.69 for Dash. Since removing learning increases overall revenues, price cuts combined with advertising increase revenues more than price cuts alone (this does not take costs of advertising into account).

When switching costs are removed from the model, price cuts become much less effective at building future market shares. This is because when there is learning and switching costs, some of the consumers who experiment with the new product and find they dislike it will continue to purchase it due to the cost of switching products. For introductory price cuts in period 1, the own price elasticity of demand drops when switching costs is removed; it rises for periods 3 and 5.

The tenth to twelfth columns show the impact of these price cuts under the results produced by the no learning model. The impact of a period t price cut on the new product's period t market share can be either higher or lower than in the full model. For Cheer, these elasticities drop, but for Surf and Dash, all but one of them rise. The intertemporal price elasticities are overestimated, especially in the earlier periods. The last three columns show the impact of these price cuts produced by the model with no switching costs. The own price, current period elasticities appear to be underestimated, except for Dash in period 5. This result seems somewhat counterintuitive: one would expect that, all else equal, increasing switching costs should reduce a product's own price elasticity. However, recall that in the no switching costs model, estimated taste variances were much larger, which will tend to raise price elasticities. As expected, the intertemporal elasticities are much lower in the no switching costs model. Thus, the impact of misspecifying dynamics on the model's implied intertemporal price elasticities is significant.

The second counterfactual, shown in Table 13, demonstrates the effect of informative advertising on the short run and intermediate run market shares for the new products. The market shares are simulated in the same way as the price cut counterfactuals. The informative advertising is modeled as follows: when the new product is introduced, I assume that every consumer receives a signal a_{ij} about their true match value for the new product which is normally distributed with mean γ_{ij} and variance σ_{aj}^2. I assume that consumers update their expected true taste, γ_{uij}^0, and the variance of their true taste distribution, σ_{uij}^2, using a Bayesian updating rule (see DeGroot (1970), pg. 166-167):

$$
\begin{aligned}
\gamma_{uij}^0 &= \frac{\frac{\gamma_{ij}^0}{\sigma_{ij}^2} + \frac{a_{ij}}{\sigma_{aj}^2}}{\frac{1}{\sigma_{ij}^2} + \frac{1}{\sigma_{aj}^2}} \\
\sigma_{uij}^2 &= \frac{1}{\frac{1}{\sigma_{ij}^2} + \frac{1}{\sigma_{aj}^2}}
\end{aligned} \tag{19}
$$

For each product, I assume that the signal variance σ_{aj}^2 is one half of the population variance in Table 7, so that for the Cheer counterfactual σ_{aj}^2 is 1.015, for Surf it is 0.925, and for Dash it is 0.905). This counterfactual is simulated using the results of the full model, the full model with switching costs removed, and the no switching costs model.

The simulated market shares are shown in Table 13. For Cheer, informative advertising increases the product's market share in all periods. For Surf, informative advertising reduces the market share in the short run, but increases it in the longer run. One explanation for this is that when consumers have a better signal of how much they will like the new product, their option value of learning is reduced. Also, the informative advertising informs consumers who have low match values with the product that they should not switch into it, which would be very bad when there is a switching cost; when there is no advertising, consumers expect to like Surf (recall that the distribution of expected tastes for Surf has the highest mean out of all three new products) and those who have lower match values for it will continue to purchase it due to the switching cost.[23] For Dash, advertising has the largest impact of all the three products. The reason for this is that consumers' expected taste for Dash is lower than Cheer or Surf, which means that the option value of learning about Dash will be lower than for Cheer or Surf and the fact that advertising reduces it will not matter that much. What the advertising does is it gives consumers a better idea of their true match value for Dash. Since the population variance of true match values for Dash is high, those who have high match values will become more likely to experiment. This makes the advertising have a stronger effect on

[23]The fact that advertising reduced Surf's market share in the presence of switching costs may seem counterintuitive in light of the discussion in Section 5.1. Recall that in the counterfactual experiment discussed there, removing learning increased Surf's market share in the presence of switching costs. A reason for this is that the simulation experiment performed in Section 5.1 was done at the actual data, where there is significant price variation, whereas these counterfactuals are computed at constant prices. Price variation will reduce the impact of the switching costs, making the results look more like the no switching costs case.

the market share for Dash than for Cheer or Surf. This result is interesting, since it suggests that informative advertising is more effective for niche products. The second column shows the predicted revenue increase resulting from the advertising.

The third and fourth columns show the predicted market shares and revenues under the full model with the switching costs parameter, η_i, set to 0 for all agents. For all three new products, informative advertising increases the new product's market share more when there is no switching costs than when it is present, for the first one or two periods. For Cheer and Dash, the impact of the advertising dies out more quickly when there are no switching costs, which is not surprising. For Surf, the advertising increases the product's market share rather than decreasing it. This suggests that the decrease in Surf's market share in the full model was driven by the switching costs. The results of the no switching costs model are shown in the third and fourth columns. The impact of the advertising is smaller for Cheer, but larger for Surf and Dash. Part of this result may be due to the fact that the learning parameter for Dash is estimated to be smaller, and for Surf it is estimated to be larger. For Dash, the learning parameter is slightly smaller, but the lack of switching costs will tend to increase the impact of the advertising.

6 Empirical Price Elasticities

In Table 14 I compute the empirical price elasticities implied by the full model, the no learning model, and the no switching costs model. The price elasticities are computed in a method similar to how I computed the impacts of price cuts in the counterfactual exercises from the previous section. First, I compute the market shares for ten periods, holding the prices of all products fixed over time at their averages. Then, I cut the price of all sizes of one of the products by ten percent for all ten periods, and compute the percentage change in the product's market share, divided by ten. Note that this type of price elasticity is not simply the impact of a 1 period price cut on the product's market share. When I compute the market shares at the lower price, consumers expect the lower price to last forever, and adjust their value functions accordingly.

The first column of the table shows the impact of a long term price cut for Cheer Liquid on Cheer Liquid, Wisk Liquid, Tide Liquid, and Tide Powder. I show the simulated elasticity for 3 of the ten periods I simulate, periods 1, 3 and 5. Thus, the own price elasticity of Cheer for a permanent price cut in Cheer is -1.85 in period 1, -1.73 in period 3, and -1.64 in period 5. The cross-price elasticity of a price cut on Cheer for Wisk's quantity in period 1 is 0.11. Note that when I computed the elasticities in the first column, I assumed that Surf and Dash were not available, and that there was learning in Cheer. The second, third, and fourth columns show the impacts of long term price cuts in Surf Liquid, Dash Liquid, and Tide Liquid respectively. For the Tide Liquid price cut, I assume that all products are available, and that there is learning in Dash. The own price elasticity of all products are estimated to be significantly larger than in the case of temporary price cuts: they range from -1.45 to -1.87. This is likely due to the fact that the price cuts are affecting consumers' future utility through the switching costs and the learning, in addition to their current utility. The cross-elasticities of demand are roughly between 5% and 30%, with the highest cross elasticity between Dash and Tide Liquid.

One reason that computing these elasticities is important is that own and cross-price elasticities are used by the Federal Trade Commission and the US Department of Justice to evaluate the competitive impact of mergers. Cross-price elasticities are an input into "unilateral effects" analysis, where agency personnel examine whether a producer of several products could profitably raise the price of one (or some) of them post-merger. As an example, suppose that all the laundry detergents analyzed in this paper were produced by different firms, and a merger between Cheer and Wisk was announced. A unilateral effects analysis would ask whether the producer of both products would find it profitable to raise the price of Cheer, for example. If Cheer and Wisk were very similar products, then many consumers of Cheer would switch to Wisk, which would mitigate the impact of the price increase on the merged firm's profits. Clearly, the higher the cross-price elasticity between the two products is, the larger will be the unilateral effect.

Own price elasticities are used in merger analysis to define "antitrust markets." An antitrust market is a group of products such that a hypothetical monopolist who produced those products would find it profit maximizing to raise price above current levels by at least some percentage (in practice this percentage is often taken to be 5%), assuming that the prices of other products did not change. Antitrust markets are used to address the question of market definition in merger analysis: when computing the market shares of merging firms, it is necessary to determine which products should be included in the denominator of the share calculation. A convention that is used by federal agencies is to define the relevant market to be the smallest antitrust market. For example, again suppose that all the laundry detergents analyzed in this paper were produced by different firms. If one was evaluating a merger between Cheer and Wisk, one might want to know whether these two products were in a separate antitrust market from other detergents. One would then ask whether a hypothetical monopolist who produced Cheer and Wisk would find it profit-maximizing to raise the prices of these products by at least 5%. An important input into this exercise will be the own price elasticities of demand for these products (the own price elasticities will themselves be functions of the cross-price elasticities with other laundry detergents). To complete the hypothetical monopolist exercise, an understanding of the production costs of the products is also necessary.

In order to carry out these exercises, it is necessary to have correct estimates of own and cross-price elasticities. If certain types of dynamics are ignored, such as learning or switching costs, then our estimated elasticities will be wrong. Because I have estimated the model without learning and without habit formation, it is possible to assess the magnitude of ignoring these types of dynamics. The results for the no learning model are shown in the fifth through eighth columns. The own price elasticities for the new products drop significantly. One reason for this is that the learning makes consumers more likely to switch into the new products; the option value of learning makes them more attractive, and a permanent price drop will generally raise this option value. It is notable that the own price elasticity for the established product, Tide Liquid, is not affected much by ignoring learning. The cross price elasticities between new products and established products drops significantly when learning is removed, for the same reason. In contrast, the cross price elasticity between established products rises. Thus, by ignoring learning we would tend to underestimate the unilateral impact of mergers between new and established products, and overestimate the unilateral impact of mergers between established products. Cross price elasticities can differ across models by a factor of as much as 90% (for example, Surf Liquid and Wisk Liquid).

7 Conclusions and Extensions

In this paper I propose a structural model of learning and experimentation that nests alternative sources of dynamics in demand, such as switching costs or consumer taste for variety. In this model, consumers are forward-looking, and I allow a rich distribution of heterogeneity in consumer tastes, price sensitivities, consumer expectations of true match values, and the type of alternative dynamics.

I estimate the model on laundry detergent scanner data and find evidence for switching costs and significant learning. The model is estimated using a Markov Chain Monte Carlo and I employ a new method for solving for consumers' value functions that substantially reduces the estimation procedure's computational burden. The results show strong support for learning and suggest that new products are experience goods. Before consumers make their first purchases of the new product, they have very similar expectations of what their true tastes will be. Those who make first purchases end up being very heterogeneous in their true tastes. The results also suggest most consumers have switching costs in addition to learning. Both learning and switching costs have a significant impact on the market shares of new products.

I re estimated the model twice to examine the impact of ignoring dynamics on the model predictions: once without learning, and once without switching costs. I find that if learning is ignored, model parameter estimates for other parameters look similar, but the model does a much worse job at predicting market shares during the periods after new product introductions. If switching costs are left out, the model underpredicts the importance of learning. I also examine the impact of misspecification on If learning is left out, the model underpredicts own price elasticities for new products, and underpredicts cross price elasticities between new and established products. If switching costs are left out, the model underpredicts own and cross price elasticities for all products.

I also examine the effect of two "what-if" experiments. In the first experiment I drop the price of the new products and simulate the products' intermediate run market share in a partial equilibrium setting, under different assumptions about dynamic demand. The results of this counterfactual exercise suggest that the impact of the price cut is greater when consumers both learn and have costs of switching products, as opposed to when there are no switching costs and they only learn. The impact of the price cut is also greater when there are only switching costs than when consumers learn and have switching costs, which suggests that price cuts may be more effective when they are combined with informative advertising or free samples. In my second "what-if" experiment, I give consumers informative advertisements which reduce their uncertainty about their true match value for the new products in the same partial equilibrium setting. The results suggest that for the two mainstream new products, informative advertising reduces the product's market share in the presence of switching costs. For a niche product, informative advertising is beneficial.

There are a number of extensions for this research that would be useful. First, this paper abstracts from the supply side: for example, the counterfactuals I compute are partial equilibrium counterfactuals and do not account for competitor responses. It would be interesting to examine the model's supply side implications: for example, we might be interested in knowing the impact of learning on the ease of new product entry, or on equilibrium pricing behavior. One way to perform this kind of exercise would be to take the demand system as given, solve for the market equilibrium, and compute comparative statics. This sort of exercise has been performed in markets with switching costs in Dubé, Hitsch, and Rossi (2006) and Che, Sudhir, and Seetharaman (2006). In these papers, the computation of the market equilibrium is tractable because switching costs are the only source of dynamics, and consumers are not forward-looking (Dubé, Hitsch, and Rossi (2006) argue that the problem with forward-looking consumers is similar to the problem with myopic ones, when firms prices follow a Markov process.). Solving the model with forward-looking consumers who learn their match values for new products is a more difficult task.

Second, the modeling technique I have used in this paper could be used to examine other problems where consumers are forward-looking and are heterogeneous in dimensions such as price sensitivities. One empirical question would be to examine the patterns of price promotions in supermarkets. For many products, supermarkets have periodic price promotions and quantity purchased spikes sharply during those promotions. Hendel and Nevo (2006b) find that consumer behavior is consistent with stockpiling behavior; Pesendorfer (2002) also finds that sales are consistent with demand accumulation. Villas-Boas and Villas-Boas (2006) provide an alternative explanation for these promotions: they could be due to learning and forgetting. It may be possible to disentangle these stories with a structural demand model that is similar to the one that was estimated in this paper. It would be more complicated to estimate than the one provided in this paper, since there is an additional source of dynamics, which is consumer stockpiling behavior.

Another potential area of application would be consumer demand for durable goods.Two examples of papers that examine the relationship between price declines and consumer dynamics for durables are Gowrisankaran and Rysman (2006) and Nair (2004). These papers have assumed that the only dynamic decision on the consumer side is the decision of when to make a purchase in the presence of declining prices. There may be additional sources of dynamics which will impact price declines, such as the existence of a used market. Adding another source of dynamics to these models will make them more computationally burdensome, but the new technique I have used could ameliorate this problem.

A Appendices

A.1 Markov Chain Monte Carlo Algorithm

Essentially, there are 2 levels to the MCMC algorithm: a level in which population-varying individual parameters on unobserved heterogeneity are drawn, and a level in which the population-fixed parameters are drawn (which includes the parameters that generate unobserved coupons and govern consumer expectations about future unobserved coupons).

1. Update value function at chosen state space points (see Section 4.4 and Appendix A.3 for more details on this process).

2. For each household, draw a new θ_i. The posterior of θ_i is proportional to

$$\left(\prod_{t=1}^{T^i} Pr(d_{it}|\theta_i, \theta, c_{it}, p_{it}, x_{it})\right) \phi(\theta_i|b, W)k(b, W),$$

   where $\phi(\theta_i|b, W)$ is the joint normal density and $k(b, W)$ is the prior on b and W. It is difficult to draw from this posterior directly since $Pr(d_{it}|\theta_i, c_{it}, p_{it}, x_{it})$ is multinomial logit. Hence, I use the Metropolis-Hastings algorithm. This means that for each household i I draw a trial θ_i^1, where $\theta_i^1 \sim N(\theta_i^0, \rho\tilde{W})$, and θ_i^0 is the previous iteration's θ_i. $\tilde{W}$) is the variance matrix W with three extra variances added in to correspond to the posterior draws. In my program, I draw the difference between γ_{ij} and γ_{ij}^0. For a particular person, this difference has variance σ_{ij}^2. We might be tempted to use this value in W, but it would violate the reversibility condition for the proposal distribution. Hence, I put in the average population mean of the σ_{ij}^2's.

   I accept the new draw θ_i^1 with likelihood

$$\frac{\left(\prod_{t=1}^{T^i} Pr(d_{it}|\theta_i^1, \theta, c_{it}, p_{it}, x_{it})\right) \phi(\theta_i^1|\tilde{b}, \tilde{W})}{\left(\prod_{t=1}^{T^i} Pr(d_{it}|\theta_i^0, \theta, c_{it}, p_{it}, x_{it})\right) \phi(\theta_i^0|\tilde{b}, \tilde{W})}.$$

   The scalar ρ is automatically selected so the acceptance rate is about 0.3.

3. Then I draw b conditional on $\tilde{\theta}_i, W$ and W conditional on $\tilde{\theta}_i, b$. The formulas for the posteriors of these parameters are the usual ones. Note that in the posterior distributions for b and W, the individual level posterior draws will drop out since they only directly depend on σ_{ij}^2.

4. Population-fixed parameter layer: at the beginning of this layer, I draw a new set of unobserved coupons, which means drawing the $\bar{c}_{ijt}$'s and the v_{ijt}'s. As described in the body of the paper,

the v_{ijt}'s are drawn from the empirical distribution of coupon values in the data. Denote p_{cjt} as the probability a consumer gets a coupon for product j in period t. This probability will be a function of parameters in θ, as described in Section 3.2. The $\overline{c}_{ijt}$'s are binary, and their distribution is:

$$Pr(\overline{c}_{ijt} = 1)$$
$$= \frac{Pr(d_{it}|c_{it}, \overline{c}_{ijt} = 1, v_{it}, \theta_i, \theta)p_{cjt}}{Pr(d_{it}|c_{it}, \overline{c}_{ijt} = 1, v_{it}, \theta_i, \theta)p_{cjt} + Pr(d_{it}|c_{it}, \overline{c}_{ijt} = 0, v_{it}, \theta_i, \theta)(1 - p_{cjt})}.$$

The more difficult task is drawing the θ, which is performed next. The posterior distribution of θ is proportional to

$$\prod_{i=1}^{I}\prod_{t=1}^{T_i}\left\{Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it})Pr(c_{it}|\theta)\right\}.$$

As with the θ_i, the Metropolis-Hastings algorithm is also used here. I draw a trial θ^1 from a $N(\theta^0, \rho_2)$ distribution. Any trial draw where the coupon probabilities, like p^0_{cj} or $p^0_{cj} + p^1_{cj}$, are outside of the $[0, 1]$ interval are automatically rejected. For cases where the draws are inside this interval, the new draw is accepted with likelihood

$$\frac{\prod_{i=1}^{I}\prod_{t=1}^{T_i}\left\{Pr(d_{it}|\theta_i, \theta^1, \Sigma_{it}, c_{it}, x_{it})Pr(c_{it}|\theta)\right\}}{\prod_{i=1}^{I}\prod_{t=1}^{T_i}\left\{Pr(d_{it}|\theta_i, \theta^0, \Sigma_{it}, c_{it}, x_{it})Pr(c_{it}|\theta)\right\}}.$$

This procedure for drawing fixed coefficients is similar to what is suggested by Train (2003), pgs 311-313, for drawing fixed coefficients in static mixed logit models. I adjust the parameter ρ_2 so that the acceptance rate is about 0.3.

These steps are iterated 15,000 times, with the first 7,500 parameter draws discarded for burn-in.

A.2 Estimation of the Price Process

When I construct consumer price expectations, I estimate a price and product availability process for a number of brand-size combinations in the market. In my data set, prices are only recorded when a consumer makes a purchase of a product. Before we can construct a process for prices, we will need a set of prices and availabilities for all products in all the stores in the data. The data set includes a set of "price files" which contain prices imputed from the household purchase data by A.C. Nielsen; one possibility would be to use this file. A drawback to using these files is that some brand-size combinations were not included. In order to calculate the average price per ounce of every brand in my estimation, I would like to keep track of the prices of the most popular brand-sizes. I therefore use a simple algorithm that is somewhat similar to Nielsen's to impute prices and availability of products in a store during a given calendar week.[24]

[24]A possible problem is that the imputed prices may not be exactly the same as the actual prices. Pesendorfer (2002) examines retail pricing of ketchups, another product included in this data set, and notes no problems with the constructed

In the data set, when a consumer makes a purchase I observe an identification number for the store where the purchase was made. The actual identity of the store was not recorded. When constructing the price series, first I run through all household purchases and store the price of the product purchased in that purchase event. If two different prices are observed at the same store during the same week, I assume that the weekly price is the price that is observed earlier. A "product" denotes a brand-size combination, as is used in estimation - the brand is one of the 13 brands and types described in Table 3, and the size is in one of the categories shown in Table 2. If no consumer purchases a particular product from a given store for an interval greater than 4 weeks, I assume that product is unavailable in that store for that period. No imputed prices are filled in for these periods. Some stores had very few observed purchases, and these stores were not included in the estimation. In all, 15 stores were used for the estimation.

Most of the stores in the sample are identified by Nielsen as being in 1 of five different chains. When Nielsen constructed their price file, they assumed that for two of the chains, pricing patterns were the same across stores in the same chain. I hold the same assumption as they do. Thus, if in a given store in one of these chains, no price is observed for a certain week and the product is assumed to be available during the week, then the price of the product is imputed to be the modal observed price at all the other stores in the same chain (or the lowest price if there are multiple modes). If there is no observed price in the other stores, then the price of the product is imputed to be the same as the previous week's price. For other stores, the price is calculated on a store by store basis. That is, if the price is not observed in a given week, it is imputed to be the same as the previous week's price. Also, during the first few weeks of the data we may not see prices for some products. These prices are imputed backwards from first observed price. Periodically products are marked below their shelf price, which is recorded by a variable in the model. I assume that these discounts only last during the week they are recorded.

Once I have constructed an array of prices and availability for each product, I estimate a discrete/continuous Markov process on prices and availability, for all products, similar to Erdem, Imai, and Keane (2003). As noted earlier, there are 48 possible brand-size combinations. Even though the space of prices is discrete, it would be difficult to accurately interpolate the value function over a 48 dimensional state space. In order to reduce the size of the state space, I estimate the Markov price process for the most popular sizes of liquids and powders, which are the 64 oz bottle for liquids, and the second category for powders. I will denote these sizes as the reference sizes for these products. For the other sizes, I assume that the price and availability is a function of the price and availability of the reference size's price and availability.

An observation in the price process estimation is the price/availability of a product in a given store during a given week. The price process is estimated only for weeks after the introduction

price series. It would also be possible to estimate a price distribution along with the model parameters, treating prices for nonpurchased brands as latent unobservables like I did for coupons.

of Cheer. First, I will describe the estimation of the Markov process for the reference sizes. If a particular product was available in the store I assume the probability of a product j's price staying the same in weeks t and $t-1$ is

$$\frac{\exp(\kappa_j' X_{jt})}{1+\exp(\kappa_j' X_{jt})},$$

where

The X_{jt}'s include a constant term, three dummy variables that are 1 for the first three months after each new product introduction, a dummy variable that is 1 during the period after the introduction of Cheer and before the introduction of Surf, which I will denote as D_1, a dummy variable that is 1 during the period after the introduction of Surf and before the introduction of Dash. To allow previous prices to affect the probability of a price change, I also include the difference between the price of product j in week $t-1$ and the average of all other brands available in week $t-1$ (that is, $p_{jt-1} - \overline{p}_{jt-1}$, where $\overline{p}_{jt-1} = (1/J)\sum_{k=1}^{J} p_{kt-1}$), and this difference interacted with D_1 and D_2. The probability of a price change varies with the difference between the brand's previous price and the average price of other products in order to allow for competitive responses. The price difference regressors are interacted with D_1 and D_2 to allow the price process to be different when different sets of products are available on the market. If the price changes in period t then I assume the density of the price change is

$$\ln(p_{jt-1}) = \lambda_j' X_{2jt} + \varepsilon_{itj},$$

where X_{2jt} includes a constant, three dummy variables for the first three months after each new product introduction, the dummy variables D_1 and D_2, the log of the price of product j in week $t-1$, the average of the logarithm of the prices of all other products in week $t-1$ (that is, $(1/J)\sum_{k=1}^{J}\ln(p_{kt-1})$), and these previous two variables interacted with D_1 and D_2. I assume $\varepsilon_{itj} \sim N(0, \sigma_j^2)$. If a product is not available in week $t-1$ but is available in week t then I estimate a similar regression to the one above but I leave out the previous price of product j. Last, I estimate a logit to model product stockouts from week to week. Letting a_{jt-1} be a dummy variable that is 1 if product j is not available in period $t-1$, I assume the probability of a store stockout in week t is

$$\frac{\exp(\zeta_j' X_{3jt})}{1+\exp(\zeta_j' X_{3jt})},$$

where X_{3jt} includes a constant, the dummy variables D_1 and D_2, a_{jt-1}, D_1 and D_2 interacted with a_{jt-1}, $(1-a_{jt-1})(p_{jt-1} - \overline{p}_{t-1})$, the interaction of the previous term with D_1 and D_2, $a_{jt-1}\overline{p}_{t-1}$, and this term interacted with D_1 and D_2. I run these estimations in Stata and keep the results in data files my fortran programs can access. Parameter estimates are not shown in this paper, but are available upon request from the author.

For the non reference sizes, I assume that the price of the product is

$$\ln(p_{jst}) = \psi_{0js} + \psi_{1js}D_1 + \psi_{2js}D_2 + \psi_{3js}D_1(1-a_{jt})\ln(p_{jt}) + \psi_{4js}D_2(1-a_{jt})\ln(p_{jt})$$
$$+ \psi_{5js}(1-a_{jt})\ln(p_{jt}) + \psi_{6js}D_1a_{jt} + \psi_{7js}D_2a_{jt} + \psi_{8js}a_{jt} + \varepsilon_{jst},$$

where the ε_{jst} is normally distributed, and the probability of the product stocking out is

$$\frac{\exp(\varphi_{0js} + \varphi_{1js}D_1 + \varphi_{2js}D_2 + \varphi_{3js}a_{jt})}{1 + \exp(\varphi_{0js} + \varphi_{1js}D_1 + \varphi_{2js}D_2 + \varphi_{3js}a_{jt})}.$$

As before, p_{jt} and a_{jt} are the price and availability of the reference size of product j in week t. Tables of parameter estimates are available upon request from the author.

For some products, it was difficult to identify parameters of the price process due to lack of observations. For example, there were only 2 observations for the log price regression for Other Liquid when it was not available in the previous period, and only 7 for the Other Powder product. This is not surprising since these products were popular and almost always available. For these products, I tabulate the empirical distribution of prices when the product is not available and draw from it when I form consumer's price expectations rather than using the regression results. For Dash Powder, there were also only 7 observations for the price change regression when the product was available. I also tabulate these 7 observations and draw from their empirical distribution to form consumer's price expectations.

As described in the paper, I solve the value function on a grid of $M = 100$ prices. Each time a household makes a purchase, it is necessary to calculate the probability of each price point p^m conditional on the observed price vector at the time of purchase. A complication is that the price process is weekly, but households do not make purchases every week. As I describe in the paper, I assume that every household expects their next purchase to take place in 8 weeks, the median interpurchase time.[25] When I calculate the probability of a particular grid point p^m given today's price, I simulate the transition probability 100 times in the 7 intervening weeks.

A.3 Details of the Value Function Solution

In this section I will describe some of the details about the computation of the value function that were left out of Section 4.4. The first detail is about dealing with the large size of the state space, which is the vector of (S, p, J, y, n). One important part of the state space is the vector of prices p_{ijst} and the set of available products, J_{it}, in a given purchase event. Recall that I only include the prices of the most popular size of each product in the price state space. Even with this simplification, because there are 13 products, this portion of the state space is still high-dimensional. Recall that the expected utility which is calculated in (17) must be retained for future use. During

[25] A less restrictive assumption would be to allow the household's expected next purchase time to be the average interpurchase time for that particular household. Doing this will mean calculating a separate value function for each household, increasing memory requirements substantially.

the estimation, these expected utilities must be stored in computer memory, which is limited in size. Because of this, I do not evaluate the value function at all possible price/availability states, but I instead do it only on a grid of M points, following Rust (1987). Although the estimated price process treats prices as a continuous variable, prices in the data are clustered at certain points. I choose the grid points as follows: for each product, I find the five most frequently occurring prices, and randomly choose each product's price from these points. This ensures that the approximated value function will be more accurate at frequently visited state space points. At any other point, I interpolate the value function as follows. Suppose that the estimated transition density of a price/availability grid point (p^m, J^m), where $m = 1, ..., M$, given a price/availability vector (p, J), is $f(p^m, J^m|p, J)$ (details of the estimation of this density are described in Appendix A.2). Assume that at the current point in the MCMC sequence we have an approximation to the value function for individual i, who is represented by the parameter vector θ_i, at all the price/availability grid points, (p^m, J^m), the learning state S, the previous product purchase y and the time state n, which I denote $\hat{EV}_i(S, p^m, J^m, y, n; \theta_i)$. Then the expected value function for some other price/availability vector (p, J) at θ_i is approximated as

$$E_{(p',J')|(p,J)}V_i(S, p, a, y, n; \theta_i, \theta) \approx \frac{\sum_{m=1}^{M} \hat{EV}_i(S, p^m, J^m, y, n; \theta_i, \theta) f(p^m, J^m|p, J)}{\sum_{m=1}^{M} f(p^m, J^m|p, J)}. \tag{20}$$

This equation is plugged into Equation (18) in the second step of the value function calculation, so the version of Equation (18) that is used in practice is

$$\begin{aligned} &E_{(p',J')|(p,J)}V_g(S, p, J, y, n, \theta_{i,g}, \theta_g) \\ &= \frac{\sum_{r=1}^{N(g)} \left[\frac{\sum_{m=1}^{M} \hat{EV}_r(S, p^m, J^m, y, n; \theta_{i,r}, \theta_r) f(p^m, J^m|p, J)}{\sum_{m=1}^{M} f(p^m, J^m|p, J)} \right] k((\overline{\theta}_{i,g} - \overline{\theta_{i,r}})/h_k)}{\sum_{i=1}^{N(g)} k((\overline{\theta_{i,g}} - \overline{\theta_{i,r}})/h_k)}. \end{aligned} \tag{21}$$

I choose to save $N(g) = 800$ previous value functions. When I estimate the model, I make a simplification to steps 1 and 2. Saving 800 previous value functions at all the state space points for all 519 households will still require a large amount of computer memory, and will be computationally intensive. I overcome this problem by recognizing that the value function only depends on the θ_i's and θ, and not any individual specific characteristics. Demographics enter utility in linear combinations with the θ_i's, so in practice I store $\alpha_{0i} + \alpha_1 INC_i + \alpha_2 SIZE_i$ rather than storing α_{0i}, α_1 and α_2 separately and treating demographics as state space variables. The same is done for the learning parameters. At the end of step 1 I randomly select a household whose parameter draw is accepted in the first Metropolis-Hastings step (the one for the population-varying coefficients) and I store only that θ_i. The $\theta_{i,r}$ that is used in (18) will in practice not depend on i.

For the kernel function $h_k(\cdot)$, I use the Epanechnikov kernel for computational efficiency. Optimal bandwidth selection requires that the bandwidth parameter of the kernel, h_k, is a function of N and that as $N \to \infty$, $h_k(N) \to 0$ and $Nh_k(N)^{2k} \to \infty$, where k is the dimension of the vector in the kernel function. In my model, there are 46 parameters that enter the kernel function. 20 of these parameters are population-fixed parameters (the 18 coupon parameters and the 2 fixed taste coefficients), and the rest are population-varying coefficients which include tastes, the underlying learning parameters, price coefficient, switching costs and coupon sensitivity parameter. For the full model, $h_k(N)$ is set to be $2/N^{(1/124)}$. There are less parameters in the kernel function for the restricted models, so the bandwidth is slightly different in them. For the no switching costs model, it is $2/N^{(1/124)}$, and for the no learning model, it is $2/N^{(1/112)}$.

References

Ackerberg, D. (2001). A new use of importance sampling to reduce computational burden in simulation estimation. Working Paper.

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

Becker, G., M. Grossman, and K. Murphy (1994). An empirical analysis of cigarette addiction. *The American Economic Review 84*(3), 396–418.

Becker, G. and K. Murphy (1988). A theory of rational addiction. *The Journal of Political Economy 96*(4), 675–700.

Cannon, D., A. Watzman, L. Lazorko, J. Bishop, T. Shafers, P. Kemezis, A. Slakter, P. Savage, and J. Winton (1987, January). Soaps and detergents: A basketful of high-tech products. *Chemical Week 140*, 22–33.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, pp. 3–38. Cambridge, New York and Sidney: Cambridge University Press.

Che, H., K. Sudhir, and P. Seetharaman (2006). Bounded rationality in pricing under state dependent demand: Do firms look ahead? how far ahead? Working Paper, Forthcoming in Journal of Marketing Research.

Chintagunta, P. (1999). Variety seeking, purchase timing, and the lightning bolt brand choice model. *Management Science 45*(4), 486–498.

Chintagunta, P., E. Kyriazidou, and J. Perktold (2001). Panel data analysis of household brand choices. *Journal of Econometrics 103*(1-2), 111–153.

Crawford, G. and M. Shum (2005, July). Uncertainty and learning in pharmaceutical demand. *Econometrica 37*, 1137–1174.

DeGroot, M. (1970). *Optimal Statistical Decisions*. McGraw-Hill, Inc.

Dubé, J., G. Hitsch, and P. Rossi (2006). Do switching costs make markets less competitive? Working Paper.

Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing Science 15*(4), 359–378.

Erdem, T., S. Imai, and M. Keane (2003). A model of consumer brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics 1*(1), 5–64.

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

Erdem, T., M. Keane, T. Öncü, and J. Strebel (2005). Learning about computers: An analysis of information search and technology. *Quantitative Marketing and Economics 3*, 207–247.

Erdem, T., M. Keane, and B. Sun (1999). Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters. *Journal of Econometrics 89*, 177–196.

Farrell, J. and P. Klemperer (2006). Coordination and lock-in: Competition with switching costs and network effects. Preliminary Draft Chapter for the Handbook of Industrial Organization, Vol 3. M. Armstrong and R. Porter (eds.), North-Holland (forthcoming).

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–472.

Gowrisankaran, G. and M. Rysman (2006). Dynamics of consumer demand for new durable goods. Working Paper.

Hartmann, W. (2006). Intertemporal effects of consumption and their implications for demand elasticity estimates. Working Paper.

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

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

Imai, S., N. Jain, and A. Ching (2005). Bayesian estimation of dynamic discrete choice models. Working Paper.

Israel, M. (2005). Services as experience goods: An empirical examination of consumer learning in automobile insurance. *American Economic Review 95*(5), 1444–1463.

Johnson, N. and S. Kotz (1970). *Continuous Multivariate Distributions I*. John Wiley, New York.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies 62*, 515–539.

McAlister, L. and E. Pessemier (1982). Variety-seeking behavior: An interdisciplinary review. *The Journal of Consumer Research 9*(3), 311–322.

Nair, H. (2004). Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. Working Paper.

Osborne, M. (2006, August). *Consumer Learning and Habit Formation in Packaged Goods Markets*. Ph. D. thesis.

Pesendorfer, M. (2002). Retail sales: A study of pricing behavior in supermarkets. *Journal of Business 75*(1), 33–66.

Pollack, R. (1970). Habit formation and dynamic demand functions. *Journal of Political Economy 78*(4), 745–763.

Roy, R., P. Chintagunta, and S. Haldar (1996). A framework for investigating habits, the hand of the past, and heterogeneity in dynamic brand choice. *Marketing Science 15*(3), 280–289.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurchner. *Econometrica 55*, 993–1033.

Seetharaman, P. (2004). Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science 23*(2), 263–271.

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

Spinnewyn, F. (1981). Rational habit formation. *European Economic Review 15*, 91–109.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Villas-Boas, S. and M. Villas-Boas (2006). Learning, forgetting and sales. Working Paper.

Table 1: Distributions of Household Demographics

Income Bracket:	Less than 20,000	20,000 - 40,000	40,000 - 60,000	60,000+
Percent:	11.5	21.9	29.1	37.6

Household Size:	1	2	3	4+
Percent:	16.9	33.7	17.1	32.4

Income and size distributions are calculated as the fraction of households observed of a particular income/size in the Sioux Falls, SD sample. Household demographics were collected in a survey that was given to all households who participated in the study.

Table 2: Shares of Sizes

Size Distribution of Liquids:

32 oz	64 oz	96 oz	128 oz	Other
15.5	52.9	11.4	17.5	2.7

Size Distribution of Powders:

17 to 20 oz	34 to 49 oz	65 to 84 oz	144 to 157 oz	Other
8.8	33.5	32.3	20.5	4.9

Size distributions were calculated by taking the number of observed purchases of liquids (powders) in a certain size category and dividing them by the number of liquid (powder) purchases observed in the entire sample.

Table 3: Market Shares for all Products

Type	Other	Era	Wisk	Tide	Solo	Cheer	Surf	Dash	Total
Liquid	0.14	0.06	0.10	0.09	0.03	0.03	0.06	0.02	0.53
Powder	0.21	-	-	0.16	-	0.07	0.03	0.01	0.47

Market share is calculated as the total number of observed purchases of a specific brand divided by the total number of observed purchases. The sample is all observed purchases in Sioux Falls over the sample time period, which starts on December 29, 1985 and ends on August 20, 1988.

Table 4: Market Shares, Average Prices: Liquids Only at Different Periods

Period	Actual Time YYYY/MM	Other	Era	Wisk	Tide	Solo	Cheer	Surf	Dash
Entire	1985/12 -	0.26	0.12	0.19	0.17	0.06	0.06	0.11	0.03
Sample	1988/08	2.80	4.21	2.90	3.97	4.12	3.57	2.67	3.12
Before Any	1985/12 -	0.41	0.14	0.19	0.16	0.10	0.00	0.00	0.00
Product Intro	1986/05	2.56	4.12	3.03	4.41	3.26	.	.	.
First Quarter	1986/05 -	0.24	0.11	0.27	0.11	0.07	0.20	0.00	0.00
After Cheer	1986/08	2.69	3.55	2.79	3.98	4.10	3.13	.	.
First Quarter	1986/09 -	0.24	0.13	0.15	0.17	0.06	0.05	0.19	0.00
After Surf	1986/11	2.91	3.87	3.05	3.10	3.85	3.76	2.01	.
First Quarter	1987/03 -	0.24	0.10	0.18	0.10	0.05	0.07	0.15	0.12
After Dash	1987/06	2.80	4.15	2.88	3.96	4.42	2.90	2.70	3.15
Remaining	1987/06 -	0.24	0.11	0.18	0.21	0.04	0.05	0.12	0.04
Time	1988/08	2.91	4.42	2.88	4.01	4.83	4.07	2.95	3.11

Market share is calculated as the total number of observed purchases of a specific brand divided by the total number of observed purchases in a given time period. Brand introduction is defined as the first time a purchase is observed of a new brand. The actual introduction dates were verified by telephone conversation with representatives of the companies; these dates coincide closely with my definition of the introduction date. According to my definition, Cheer was introduced in the last week of May, 1986, Surf in the first week of September, 1986, and Dash in the third week of March, 1987. Average prices in dollars are shown under the market share. Prices are calculated using observed purchase data. If there are I purchases in a given period, the average price for a specific brand in the particular period is calculated as $(1/I)\sum_{i=1}^{I}(p_i - c_i)$, where p_i is the shelf price at the time of purchase, and c_i is the total value of coupons used at the time of purchase.

Table 5: Parameter Estimates of b, θ, and W (Utility Function)

Coefficient	Mean, b, θ		Variance, W		Coefficient	Mean, b, θ		Variance, W	
Taste Parameters					Learning Parameters				
Era L	-0.908	(0.161)	3.258	(0.462)	Cheer, γ_i^0	-0.518	(0.089)	0.356	(0.167)
Wisk L	-0.456	(0.104)	2.249	(0.326)	Cheer, σ_{i0}^2	-0.58	(0.072)	0.368	(0.157)
Tide L	-0.049	(0.106)	1.511	(0.28)	Cheer, Size (σ_{j1}^2)	0.148	(0.005)	-	
Solo L	-2.208	(0.677)	5.632	(2.061)	Cheer, Inc (σ_{j2}^2)	-0.068	(0.004)	-	
Other P	0.218	(0.004)	-		Surf, γ_i^0	-0.161	(0.07)	0.432	(0.123)
Tide P	0.031	(0.005)	-		Surf, σ_{i0}^2	-0.324	(0.08)	0.425	(0.072)
Cheer P	-1.417	(0.254)	3.6	(0.849)	Surf Size (σ_{j1}^2)	-0.127	(0.004)	-	
Surf P	-0.461	(0.184)	1.139	(0.448)	Surf Inc (σ_{j2}^2)	0.019	(0.003)	-	
Dash P	-1.664	(0.214)	2.603	(0.388)	Dash, γ_i^0	-0.943	(0.109)	0.408	(0.09)
64 oz	0.839	(0.072)	1.308	(0.194)	Dash, σ_{i0}^2	-0.528	(0.079)	0.287	(0.079)
96 oz	0.045	(0.105)	2.056	(0.324)	Dash, Size (σ_{j1}^2)	-0.025	(0.008)	-	
128 oz	-0.491	(0.165)	2.974	(0.595)	Dash, Inc (σ_{j2}^2)	-0.003	(0.003)	-	
34-49 oz	0.402	(0.093)	2.251	(0.271)	Exogenous				
65-84 oz	0.316	(0.119)	2.897	(0.419)	Variables				
144-157 oz	-0.551	(0.172)	5.851	(0.768)	Price (α_{i0})	-2.382	(0.471)	11.915	(5.502)
S.C. (η_{i0})	1.311	(0.081)	2.002	(0.253)	Price, Size (α_1)	-0.094	(0.006) -		
S.C. Size (η_1)	-0.079	(0.008)	-		Price, Inc (α_2)	0.014	(0.006)	-	
S.C. Inc (η_2)	0.068	(0.01)	-		Coupon (α_{i0c})	-0.673	(0.17)	0.229	(0.102)
					Feature	0.888	(0.055)	0.483	(0.104)
					Display	0.944	(0.041)	0.242	(0.053)

This table shows the estimated parameters of the consumer flow utility (Section 3.1). In most parameters I allow normally-distributed heterogeneity across the population, and so I have estimated the population mean of the coefficient (b) and the variance (W). Because my model estimation procedure is Bayesian, the numbers in this table show statistics from the simulated posterior distribution of each parameter. The first two columns of numbers in the table, under the heading "Mean", show the posterior means of the mean parameters of the taste distributions, and the standard deviations of these mean parameters (in parentheses). The third and fourth columns show the mean and standard deviation of the variance parameters of the taste distributions. The last four columns of numbers show these quantities for parameters other than the taste distribution parameters. The posterior means and standard deviations in this table may be interpreted in the same way as estimated parameters and estimated standard errors that are produced by classical procedures. Some utility coefficients, such as the price coefficient and the consumer uncertainty (see Equations (6) and (5)), are transformations of the parameters in the table. For some utility coefficients, such as the Other Powder taste, the population variance was restricted to be 0. These parameters are shown with dashes.

Table 6: Parameter Estimates: Coupon Probabilities

Coefficient	Mean	Standard Err.
Non-Introductory Periods (p^0_{cj})		
Other L	0.268	0.018
Era L	0.259	0.023
Wisk L	0.314	0.014
Tide L	0.263	0.017
Solo L	0.349	0.019
Cheer L	0.16	0.016
Surf L	0.237	0.018
Dash L	0.001	0.001
Other P	0.191	0.024
Tide P	0.221	0.023
Cheer P	0.214	0.013
Surf P	0.039	0.012
Dash P	0.038	0.012
Introductory Adjustment		
Cheer (p^1_{cj})	-0.001	0.007
Surf (p^1_{cj})	-0.078	0.007
Dash (p^1_{cj})	0.13	0.004
Est., After Cheer ($p^{Cheer,1}_c$)	-0.038	0.012
Est., After Surf ($p^{Cheer,1}_c$)	-0.038	0.012
Est., After Dash ($p^{Cheer,1}_c$)	0.015	0.009

This table shows the estimates of the coupon distribution described in Section 3.2. The numbers in the first column under the heading "Non-Introductory Periods" are the probability a consumer receives a coupon for a given product after any new product's "introductory" period: the period after the first 3 months after a new product introduction. The numbers under the heading "Introductory Adjustment" are added to the probabilities under the previous heading during a given product's introductory period (the first 3 months after its introduction). For example, the probability of getting Surf during its introductory period is 0.237 - 0.078 = 0.159, and the probability of getting a Liquid Tide coupon during Surf's introductory period is 0.263 - 0.038 = 0.225.

Table 7: Average Values of Consumer Uncertainty for New Products

Product	Mean of σ^2	Population Variance
Cheer	2.03	0.47
Surf	1.85	0.50
Dash	1.81	0.34

I computed the uncertainties in the table using the individual-level draws denoted as θ_i in the body of the paper: for each consumer I save her individual-level parameter draws in each step of the MCMC algorithm, and her individual level σ^2 for each product, which is computed according to equation (5). In a given step I compute the population mean of σ^2 and its variance, and average calculate her uncertainty. These values are averaged across steps.

Table 8: Average Consumer Uncertainty, Across Demographics

Cheer

Size/Income	< 20,000	20,000 - 40,000	40,000 - 60,000	60,000+	Averages
1	1.969	2.059	2.194	2.393	2.021
2	1.837	2.021	2.253	2.399	2.042
3	1.789	1.927	2.128	2.304	2.073
4+	1.686	1.865	2.025	2.189	1.988
Averages	1.864	1.936	2.096	2.265	2.029

Surf

Size/Income	< 20,000	20,000 - 40,000	40,000 - 60,000	60,000+	Averages
1	2.006	1.907	1.727	1.497	1.944
2	2.046	1.89	1.799	1.603	1.889
3	2.032	1.902	1.782	1.682	1.821
4+	1.997	1.933	1.801	1.679	1.831
Averages	2.024	1.911	1.792	1.669	1.852

Dash

Size/Income	< 20,000	20,000 - 40,000	40,000 - 60,000	60,000+	Averages
1	1.868	1.827	1.786	1.914	1.851
2	1.863	1.83	1.813	1.764	1.829
3	1.927	1.813	1.798	1.759	1.802
4+	1.893	1.826	1.792	1.77	1.805
Averages	1.878	1.823	1.796	1.766	1.814

This table shows the average uncertainty in the population for each new product, which corresponds to the variable σ^2 from Section 3.1. They are computed in the same way as the numbers from the previous table.

Table 9: Effect of Removing Learning On New Product Market Share (Full Model)

Product	Predicted Market Share, Learning	Predicted Market Share, No Learning	% Change
Entire Period			
Cheer	3.0	4.0	34
Surf	6.9	8.5	24
Dash	1.4	2.2	58
1st 12 Weeks After Intro			
Cheer	8.8	9.6	9.0
Surf	12.4	13.7	10
Dash	6.9	9.0	30

The first column of the table shows the simulated market share at the parameter estimates (average of market shares predicted at each step of the MCMC algorithm). The second column of the table shows the market share when every consumer knows her true taste draws for all three products. The market shares are predicted at the data, so prices, features, etc. are not changed. The first three rows show the market shares aggregated over the entire data length, and the last 3 show the market shares for each new product during the first 12 weeks after its introduction.

Table 10: Effect of Removing Learning On New Product Market Share (No switching costs Model)

Product	Predicted Market Share, Learning	Predicted Market Share, No Learning	% Change
Entire Period			
Cheer	3.5	3.8	9.9
Surf	6.8	7.8	15
Dash	1.8	2.4	34
1st 12 Weeks After Intro			
Cheer	9.3	10.2	9.5
Surf	11.9	13.6	14
Dash	8.1	9.9	22

Table 11: Actual and Predicted Market Shares

	Actual	Full Model	No switching costs	No Learning
Other L	12.0	11.5	11.3	9.1
Era	6.8	6.8	6.8	6.4
Wisk	10.2	10.3	10.1	8.6
Tide	11.0	10.7	11.1	10.5
Solo	3.3	3.2	3.3	3.0
Cheer	3.2	3.0	3.5	1.4
Surf	6.6	6.9	6.8	5.5
Dash	1.6	1.4	1.8	1.8
Other P	18.6	19.1	17.1	15.1
Tide P	16.1	15.4	17.0	18.9
Cheer P	7.4	7.4	7.3	9.8
Surf P	3.1	3.4	3.2	8.7
Dash P	0.75	0.90	0.86	1.13
Avg Abs Prediction Error				
Full Period (12/85 - 08/88)		0.26	0.33	1.8
Cheer Intro (05/86 - 08/86)		1.3	1.7	3.2
Surf Intro (09/86 - 11/86)		0.9	0.9	2.0
Dash Intro (03/87 - 06/87)		0.7	0.7	1.8
Last 63 Weeks (06/87 - 08/88)		0.6	0.5	0.6

The first thirteen rows of the first column of the table shows the actual market share of each of the products during the times they were available. The first thirteen rows of the second, third and fourth columns show the simulated market shares at the parameter estimates from each model. The last 5 rows show the absolute difference between the predicted and actual shares, averaged over products in different periods. The periods Cheer Intro, Surf Intro, and Dash Intro refer to the first 12 weeks after the introduction of each new product.

Table 12: Counterfactual: Impact of Price Cuts in Periods 1,3 and 5

Model	Full Model									No Learning Model			No S.C. Model		
Dynamics	All Dynamics			No Learning			No S.C.								
Period	1	3	5	1	3	5	1	3	5	1	3	5	1	3	5
Cheer															
1	12.3			9.65			11.5			11.9			9.95		
2	0.87			1.17			0.16			2.18			0.08		
3	0.41	10.6		0.60	7.85		0.15	10.8		1.54	9.09		0.06	9.43	
4	0.22	0.99		0.31	1.14		0.14	0.13		1.29	1.22		0.05	0.05	
5	0.17	0.46	9.72	0.20	0.60	7.37	0.14	0.12	10.3	1.20	0.55	8.61	0.05	0.04	9.18
6	0.14	0.26	1.00	0.14	0.32	1.11	0.12	0.11	0.11	1.15	0.29	1.17	0.05	0.04	0.05
7	0.12	0.18	0.48	0.10	0.21	0.58	0.12	0.11	0.11	1.12	0.19	0.54	0.04	0.04	0.04
8	0.11	0.14	0.26	0.08	0.14	0.30	0.11	0.10	0.10	1.10	0.13	0.29	0.04	0.04	0.04
9	0.10	0.12	0.20	0.06	0.11	0.20	0.10	0.09	0.09	1.08	0.10	0.19	0.04	0.03	0.03
10	0.09	0.11	0.15	0.05	0.09	0.14	0.10	0.09	0.08	1.07	0.07	0.13	0.03	0.03	0.03
Surf															
1	11.9			10.2			11.3			13.7			10.2		
2	0.80			1.34			0.00			3.22			0.08		
3	0.33	10.3		0.69	8.33		0.03	10.8		2.60	10.1		0.13	9.84	
4	0.13	1.10		0.35	1.27		0.05	0.04		2.34	1.39		0.15	0.15	
5	0.09	0.50	9.53	0.24	0.69	7.80	0.06	0.05	10.5	2.25	0.62	9.66	0.16	0.14	9.48
6	0.07	0.22	1.18	0.17	0.36	1.25	0.07	0.06	0.05	2.20	0.31	1.38	0.15	0.14	0.15
7	0.06	0.14	0.58	0.13	0.24	0.67	0.07	0.06	0.06	2.18	0.19	0.65	0.14	0.13	0.14
8	0.06	0.09	0.28	0.10	0.17	0.34	0.07	0.06	0.06	2.16	0.12	0.33	0.13	0.12	0.13
9	0.06	0.07	0.19	0.08	0.12	0.23	0.06	0.06	0.05	2.14	0.09	0.21	0.12	0.12	0.12
10	0.06	0.06	0.14	0.07	0.10	0.16	0.06	0.06	0.05	2.14	0.07	0.14	0.12	0.11	0.11
Dash															
1	14.2			11.8			13.9			15.0			13.3		
2	1.34			1.45			0.39			2.25			0.61		
3	0.78	12.5		0.71	9.64		0.35	13.2		1.52	12.8		0.49	12.4	
4	0.52	1.43		0.38	1.43		0.31	0.31		1.25	1.53		0.40	0.40	
5	0.42	0.77	11.8	0.24	0.69	9.10	0.28	0.27	12.7	1.15	0.63	12.4	0.34	0.34	11.9
6	0.35	0.50	1.41	0.17	0.36	1.39	0.25	0.24	0.24	1.09	0.31	1.62	0.29	0.29	0.29
7	0.31	0.41	0.77	0.12	0.24	0.71	0.23	0.22	0.22	1.07	0.19	0.67	0.25	0.25	0.25
8	0.28	0.33	0.48	0.10	0.16	0.39	0.21	0.20	0.20	1.06	0.13	0.35	0.22	0.22	0.22
9	0.25	0.28	0.37	0.08	0.12	0.25	0.19	0.18	0.18	1.04	0.10	0.23	0.20	0.20	0.20
10	0.22	0.24	0.30	0.06	0.10	0.16	0.17	0.17	0.16	1.02	0.07	0.16	0.18	0.18	0.18

This table shows simulated percentage change in market share due to a ten percent, 1 period price cut in Cheer, Surf or Dash. The columns labeled “Period” denotes the period when the price cut takes place. The first three columns show the impact of the price cut under the estimates of the full model. The next three show the impact of the price cut under the full model estimates when consumers know their tastes, and the next three when the switching costs parameters are set to zero. The final six columns show the impact of the price cut using the estimates of the restricted models.

Table 13: Counterfactual: Effect of Informative Advertising

Product	Full Model				No S.C. Model	
Dynamics	All Dynamics		No S.C.			
	Share	Revenue	Share	Revenue	Share	Revenue
Cheer						
1	15.06	17.51	17.74	28.92	10.96	14.87
2	17.58	22.99	16.65	27.64	8.82	12.00
3	18.97	26.47	15.70	26.39	7.52	10.23
4	19.22	27.72	14.78	25.13	6.60	8.97
5	19.14	28.29	13.93	23.90	5.92	8.04
⋮						
10	16.90	26.84	10.38	18.45	3.96	5.38
Total		261.36		234.53		85.34
Surf						
1	-6.04	-9.77	5.01	10.39	10.88	18.30
2	-2.61	-4.25	5.38	11.01	13.17	21.40
3	0.18	0.96	5.69	11.60	13.95	22.56
4	1.80	4.18	5.86	11.95	14.02	22.76
5	2.89	6.45	5.88	12.02	13.79	22.54
⋮						
10	5.04	11.14	5.06	10.55	11.39	19.40
Total		46.86		113.33		211.34
Dash						
1	24.49	12.75	32.09	26.18	33.09	21.42
2	28.26	16.40	29.77	24.93	30.37	20.43
3	29.81	18.62	27.96	23.95	28.01	19.46
4	29.97	19.49	26.33	23.04	25.97	18.54
5	29.66	19.98	24.94	22.26	24.15	17.70
⋮						
10	26.58	19.88	19.34	18.63	18.08	14.46
Total		188.12		220.56		175.66

This table shows the impact of informative advertising in the first period on product market shares, in percentage terms, and on product revenues, which are in dollars. Consumers are given a signal on the product's quality which has half the variance of their uncertainty about their match value with the product. The first two columns show the simulated shares and revenues using the results of the full model, while the third and fourth use the results of the full model with the switching cost parameters restricted to 0. The final two columns show the results for the model with no switching costs.

Table 14: Counterfactual: Long Term Empirical Price Elasticities

	Full Model				No Learning Model				No S.C. Model			
	Cheer	Surf	Dash	Tide L	Cheer	Surf	Dash	Tide L	Cheer	Surf	Dash	Tide L
Cheer												
1	-1.85	0.14	0.04	0.12	-1.19	0.06	0.03	0.14	-1.10	0.10	0.05	0.09
3	-1.73	0.15	0.05	0.12	-1.21	0.08	0.04	0.14	-1.02	0.09	0.05	0.09
5	-1.64	0.15	0.05	0.11	-1.22	0.09	0.04	0.14	-0.99	0.09	0.05	0.09
Surf												
1	-	-1.80	0.05	0.18	-	-1.37	0.04	0.20	-	-1.14	0.05	0.11
3	-	-1.67	0.07	0.16	-	-1.44	0.05	0.19	-	-1.08	0.05	0.11
5	-	-1.60	0.08	0.16	-	-1.47	0.06	0.19	-	-1.06	0.05	0.11
Dash												
1	-	-	-1.84	0.27	-	-	-1.50	0.20	-	-	-1.42	0.16
3	-	-	-1.87	0.27	-	-	-1.62	0.22	-	-	-1.42	0.17
5	-	-	-1.87	0.27	-	-	-1.66	0.23	-	-	-1.39	0.17
Wisk L												
1	0.11	0.19	0.06	0.20	0.05	0.10	0.04	0.23	0.08	0.12	0.06	0.13
3	0.12	0.21	0.07	0.19	0.07	0.14	0.06	0.21	0.08	0.11	0.06	0.13
5	0.12	0.20	0.08	0.18	0.08	0.15	0.06	0.21	0.08	0.12	0.06	0.13
Tide L												
1	0.10	0.17	0.05	-1.55	0.04	0.07	0.03	-1.56	0.06	0.09	0.04	-0.82
3	0.12	0.19	0.07	-1.48	0.06	0.10	0.04	-1.48	0.06	0.09	0.04	-0.82
5	0.12	0.19	0.07	-1.45	0.06	0.11	0.05	-1.47	0.06	0.09	0.05	-0.82
Tide P												
1	0.10	0.19	0.05	0.12	0.05	0.07	0.03	0.13	0.05	0.08	0.03	0.07
3	0.09	0.17	0.05	0.11	0.05	0.08	0.03	0.12	0.05	0.07	0.03	0.07
5	0.09	0.16	0.05	0.11	0.05	0.08	0.03	0.12	0.05	0.07	0.04	0.07

This table shows the empirical price elasticities implied by my model estimates, $\frac{\partial Q_{it}}{\partial P_j}\frac{P_j}{Q_{it}}$. The row labels show the product i that is being affected, and the period of interest. The column labels show the product j whose price is being changed. The price change is assumed to occur from period 1 onwards, and consumers are assumed to understand this, which is why this is a long term elasticity. Furthermore, when this elasticity is computed, the prices of all products are set to be constant over time. Thus the number in the first row and second column, 0.14, shows the impact of a permanent price cut in Cheer from period 1 onwards on Surf's period 1 market share. The first four columns show the elasticities implied by the full model, the next four show the elasticities for the no learning model, and the next four for the no switching costs model.

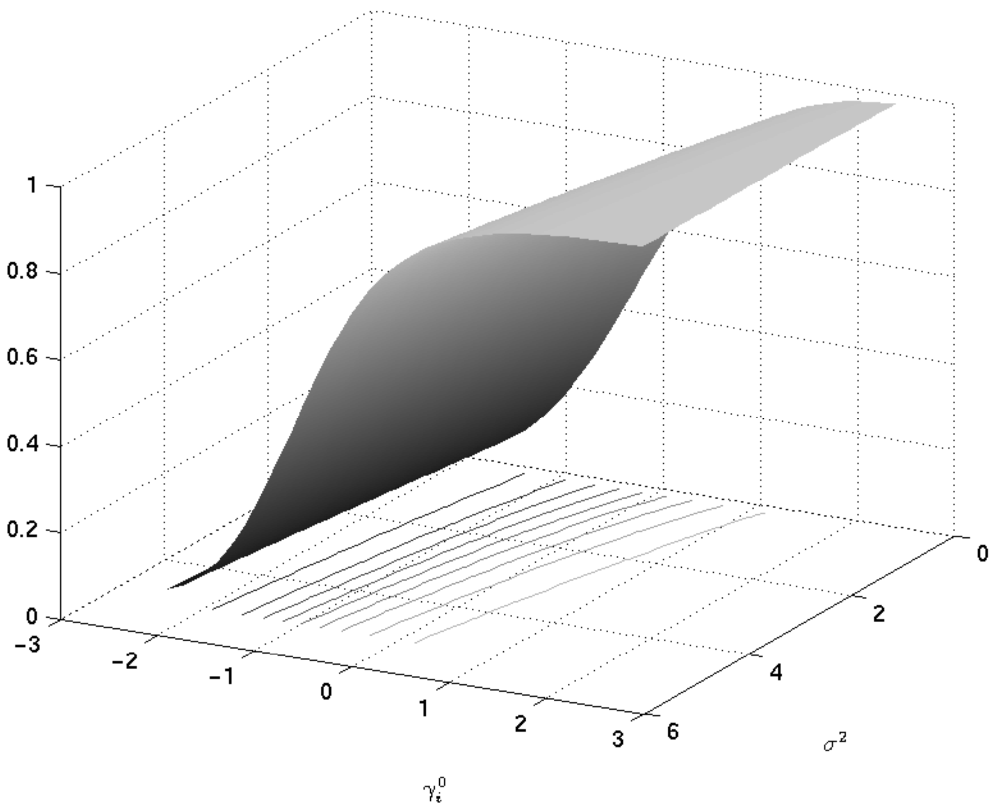

Figure 1: Probability of Experimenting

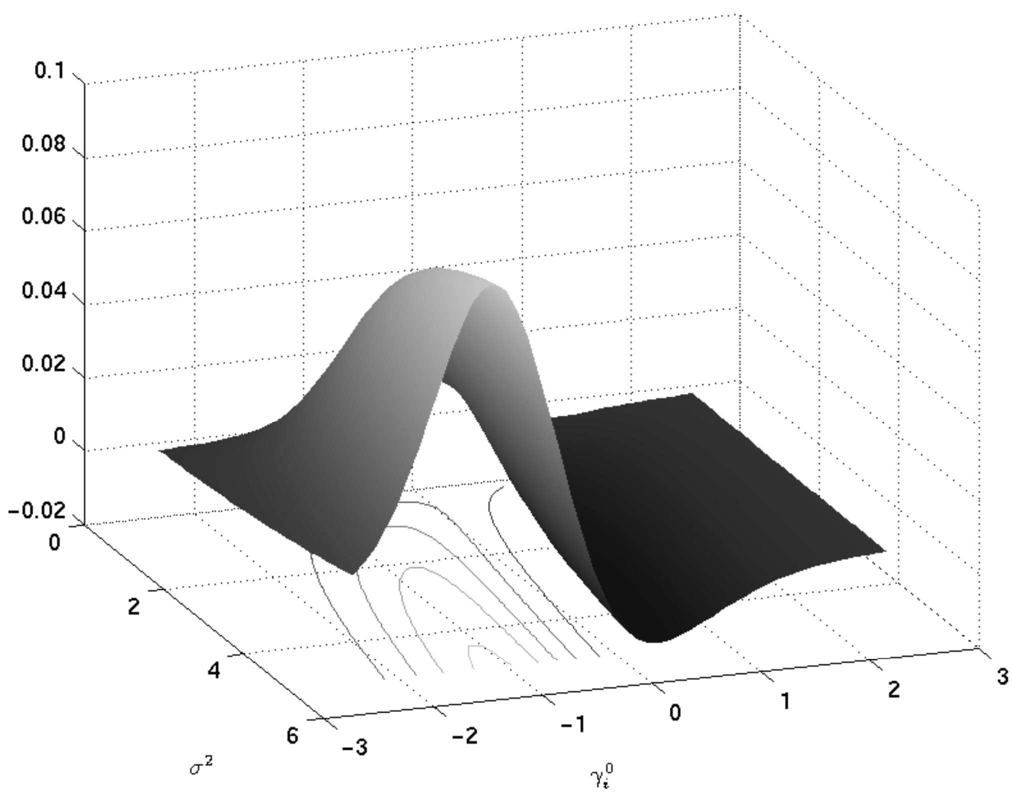

Figure 2: Response of Probability of Experiment to a Future Price Increase

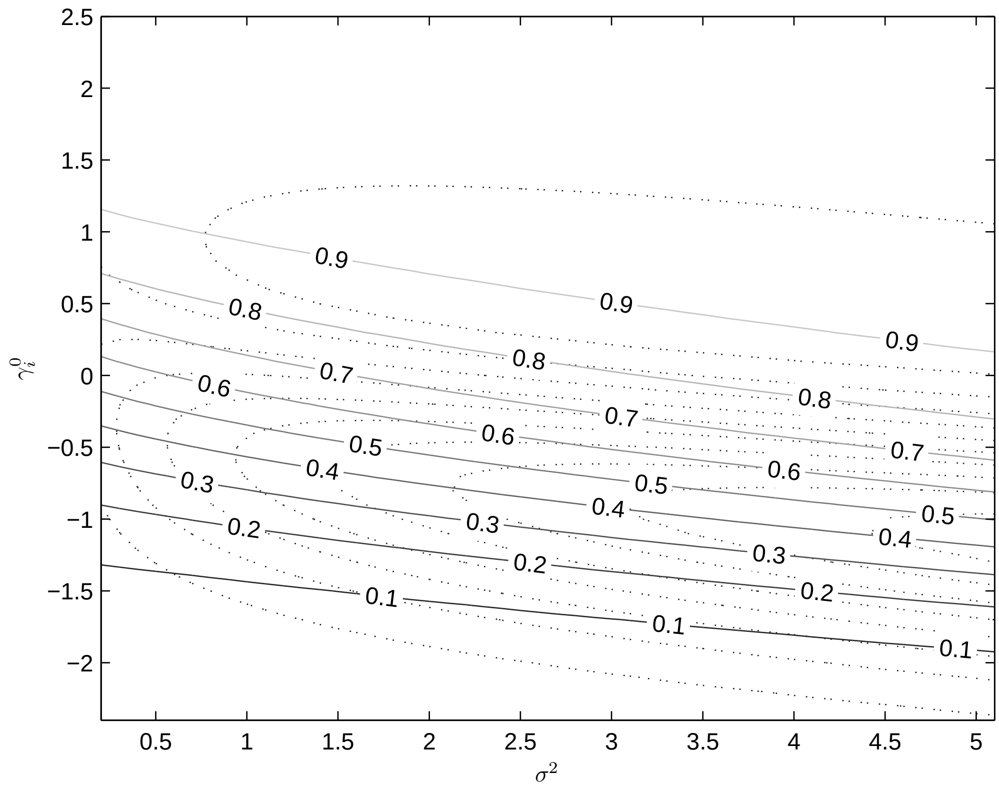

Figure 3: Overlay of Level Curves From Previous Figures

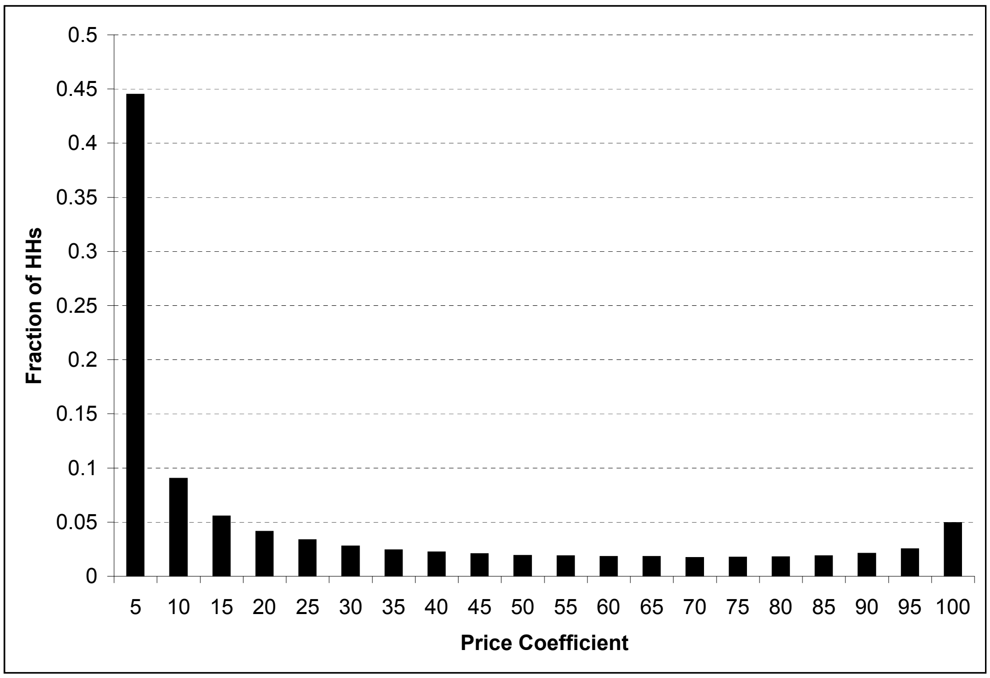

Figure 4: Histogram of Estimated α_i

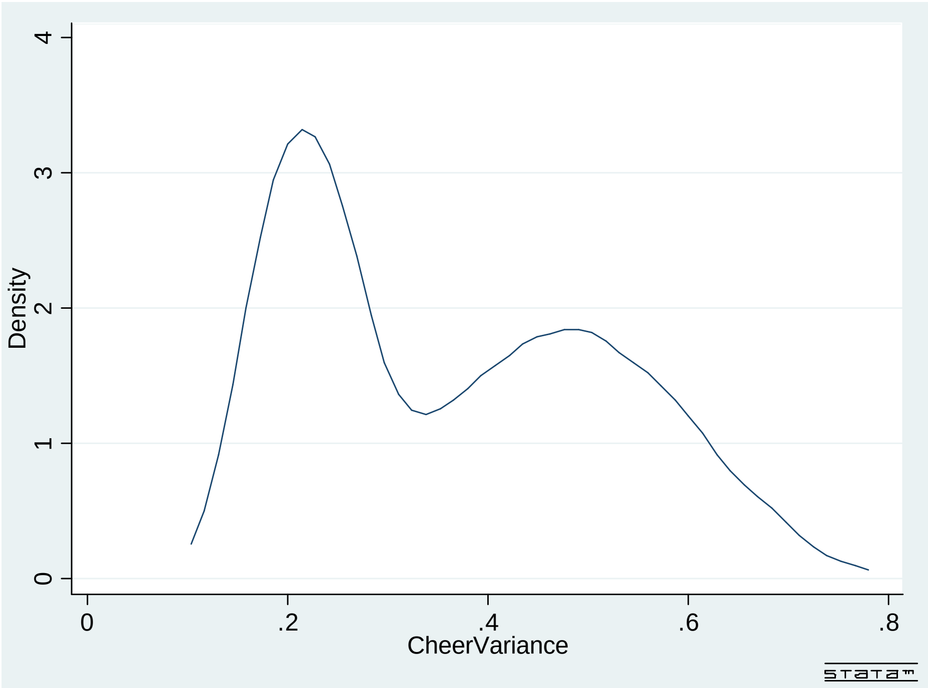

Figure 5: Posterior Density of σ^2_{i0} for Cheer

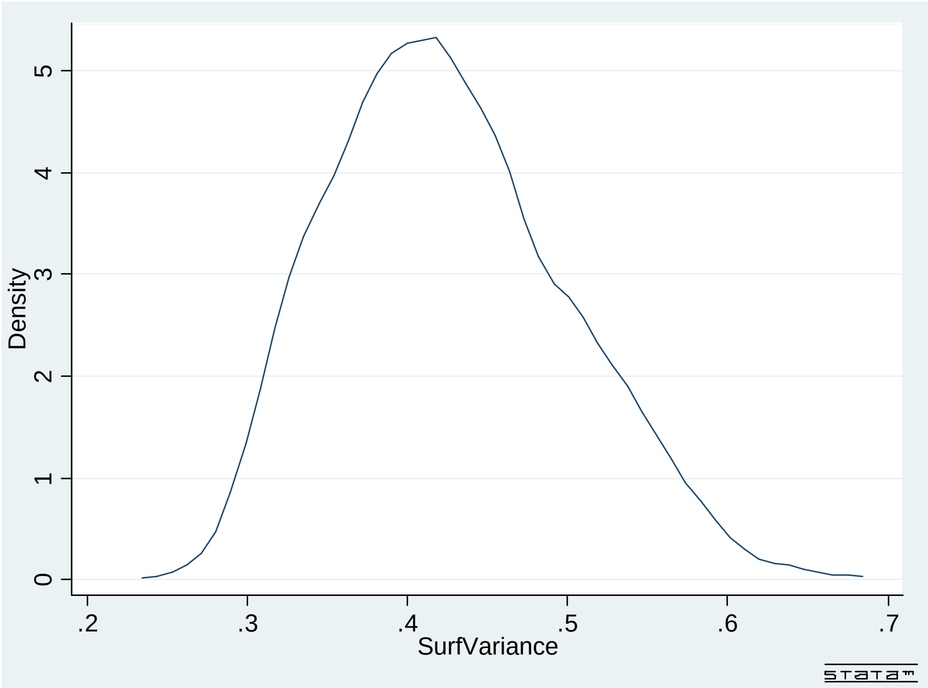

Figure 6: Posterior Density of σ^2_{i0} for Surf

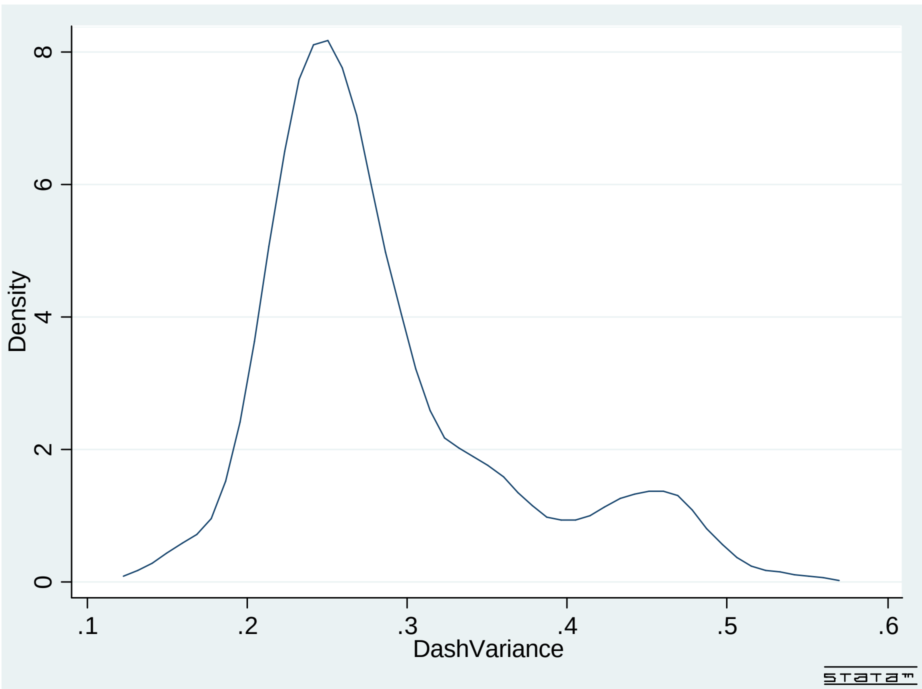

Figure 7: Posterior Density of σ_{i0}^2 for Dash

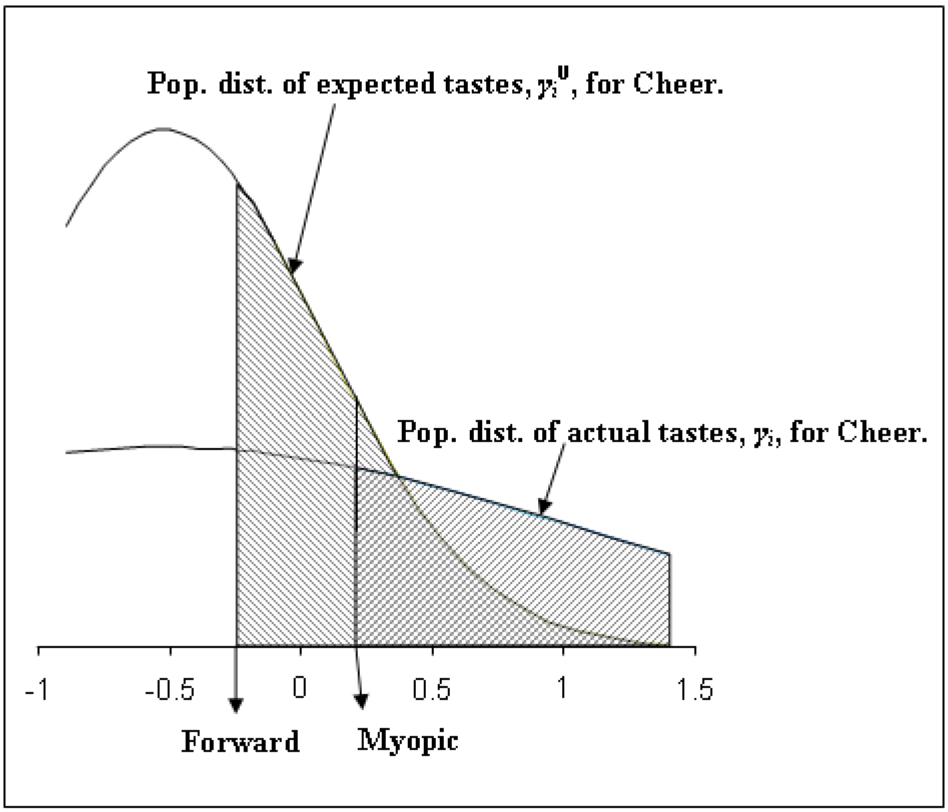

Figure 8: Estimated Taste Distributions For Cheer