Consumer Learning, Habit Formation, and Heterogeneity: A Structural Examination

Matthew Osborne*

November 15, 2005

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, such as habit formation. 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, and that once consumers have learned their true match values they are very heterogeneous. The estimation results also suggest significant habit formation. Using counterfactual computations derived from the estimates of the structural demand model, I demonstrate that the presence of habit formation 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 consumers only form habits as opposed to learning and forming habits at the same time, which suggests that firms should combine price cuts with introductory advertising or free samples to increase their impact.

*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 and Dan Quint for helpful comments. I would like to thank the Stanford Institute for Economic Policy Research for financial support, and the James M. Kilts Center, GSB, University of Chicago, for provision of the data set used in this paper.

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 (2000)). In these papers it is assumed that the only type of dynamics in demand come from learning, and alternative types of dynamics, such as habit formation, are not modeled. Similarly, papers that estimate other forms of dynamics (see Chintagunta, Kyriazidou and Perktold (1999) 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 habit formation. Learning can be empirically separated from habit formation 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 habit formation in addition to learning changes the implications of the standard empirical learning model. For example, switching products becomes more costly, so consumers may be less likely to experiment with new products. Also, the intermediate run impact of an introductory price cut may be increased when compared to the learning only case, 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 nonetheless become habituated to it. Another 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.

To motivate the research I will present in this paper, I will discuss a simple example of learning in a packaged goods market. Consider a market for a frequently purchased packaged good with two products: an established product that has been available for a long time, and a new product for which we observe the introduction. Suppose that consumers have an individual-level intrinsic match value for the new product that does not change over time. A researcher in economics or marketing may be interested in knowing whether consumers need to learn that match value by purchasing and consuming the product (if consumers need to learn by experience, there is a potential role for informative advertising or free samples), or if they know their match value beforehand through other means, such as experience with the established product or by examining the new product's package. Suppose that consumers in fact do not perfectly know their true match values, but only have expectations about their true match values and must consume the new product to learn about it. What should the researcher expect to observe? First, if consumers are forward-looking they will recognize that there is value to learning about the new product, since they might like it and keep purchasing it in the future. Forward-looking consumers will therefore have an incentive to *experiment* with the product, which means that they will purchase it sooner than they would have were they myopic. Therefore, the researcher should observe consumers purchasing the new product very soon after its introduction. Second, the researcher should be able to infer whether consumers' match values for the product are higher or lower than for the established product after their first purchase of it. If the researcher observes individual behavior over time, consumers who have high match values for the new product will continue to purchase it after experimenting, and consumers who have low match values will switch back to the established product.

A problem for the researcher is that there may be 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, some consumers could be habit-formers: holding fixed their intrinsic match values, their marginal utility for the new product could be increased by a previous purchase. When a habit-former makes a first purchase of the new product, she will be likely to keep on purchasing it. To the researcher it may look like these consumers have high match values for the new product. The researcher will therefore need to take into account that these other types of dynamics exist in order to properly isolate learning.

A second problem for the researcher is that consumers may be heterogeneous in their price sensitivities. Suppose that when the new product is introduced, its price is initially low and then it is raised. 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. It the researcher does not take into account that they are price sensitive, it may look like they experimented with the product and disliked it.

This brings me to the first contribution of this paper, which is to estimate a model of consumer learning and experimentation on household panel data that nests alternative sources of dynamics in demand, such as habit formation and consumer taste for variety. In my model, consumers are forward-looking and take into account the effect of learning and alternative 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. This paper is the first to estimate such a demand model.

The model is estimated on household-level panel data for laundry detergent purchases. During the time the data was collected, three new product introductions are observed. The results of the estimation support the hypothesis that consumers learn about the three new products by experience: before consumers make their first purchases of the new product, they have very similar expectations about their intrinsic match values. After they purchase it for the first time, consumers' realized match values are very heterogeneous across the population. The estimation results also suggest that more learning occurs among smaller and lower income households, and that most households form habits with products in addition to learning.

An important question to consider is why it might be important for a researcher to differentiate between learning and alternative sources of dynamics in demand. As I mentioned above, one reason is that learning provides a role for informative advertising. Another reason is that the type of dynamics that exist in demand will impact pricing policy for new products. As an example, suppose our researcher wants to target coupons at some households in order to increase the new product's intermediate run market share. Suppose further that prior to the new product introduction, the researcher has observed the purchase behavior of households (this could be done using magnetic swipe cards which are popular in many grocery stores today; if such data is not available the researcher may know that certain demographic characteristics are positively correlated with habit formation), and can split people into habit formers and non habit-formers. Assume that the researcher knows the new product is an experience good, so all consumers will have to learn about the product. The researcher may wish to know whether targeting the habit-formers will have a greater impact on the product's intermediate run market share than the non habit-formers. If the researcher targets the non habit-formers, then the result will be that some of these consumers make a purchase as a result of the coupon will find they have a high intrinsic match value for the new product and will continue to purchase it in the future. The intermediate run impact of targeting the habit-formers could be greater or smaller than the non habit-formers. It could be smaller because when consumers form habits, they lose utility from switching brands. These consumers will realize that if they have a low match value for the new product, they will incur a future utility loss from having to switch back to the established product. On the other hand, under strong habit formation the impact of the price cut could be greater: some of the consumers who learn they have a low match value for the new product will become habituated to it, and will be less likely to switch away. With estimates of the magnitudes of these forces in hand, the firm could evaluate its optimal pricing policy.

The demand model that I estimate is structural, which means that it is possible to take the model away from the data and to examine the effect of "what-if" experiments. I perform two such experiments. In the first experiment I compare the impact of an initial price drop on the intermediate run market share of a new product under different assumptions on the type of dynamics in demand. Another contribution of my paper is to compute the effect of such a price cut in a partial equilibrium setting for each of the three new products. I find that the intermediate run effect of the price cut is greater when consumers both learn and form habits as opposed to when they only learn. Also, the impact of the price cut is greater when consumers only form habits as opposed to learning and forming habits at the same time, which suggests that firms should combine price cuts with introductory advertising or free samples to increase their impact.

In my second "what-if" experiment I examine the impact of informative introductory advertising on the new product's intermediate run market share in the presence of habit formation, and when there is no habit formation. The results of this exercise suggest when there is habit formation, informative advertising can reduce the market shares of new products that are mainstream. Informative advertising for niche products is still very beneficial, even in the presence of habit formation.

The last significant research contribution of my paper is in the area of estimation of dynamic structural models. Previous papers that estimate structural demand models where consumers are forward-looking (for example, Erdem and Keane (1996), or Crawford and Shum (2000)) use classical methods such as the maximum-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.

I overcome this problem by estimating my model using the Bayesian method of Markov Chain Monte Carlo, which is often better 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 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.

2 A Discussion of Previous Literature

In this section I will discuss previous literature about structural estimation of models of consumer learning and experimentation, and I will survey some papers that quantify habit formation. My research differs from both of these literatures in that it is the only paper to estimate a structural model of consumer experimentation and learning which nests alternative sources of state dependence and models consumers as forward-looking agents who explicitly solve their discrete choice dynamic programming problem. Another way in which my paper differs from this literature is in my estimation methods. The literature I will review uses classical methods, while I estimate my model with Bayesian methods, which can more easily deal with rich distributions of unobserved heterogeneity.

A pioneering paper in the estimation of structural models of consumer learning and experimentation is Erdem and Keane (1996), which specifies and estimates a Bayesian learning model on panel data on individual household purchases of liquid laundry detergents. In their model, consumers choose between 8 different products and are learning about 1 unobserved attribute for each product, which is interpreted as the detergent's cleaning power. This unobserved attribute is assumed to not vary across the population or across time, so that under full information it is not possible for one consumer to have a higher intrinsic preference for a particular product than another consumer. 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.

The paper assumes that consumers do not have full information and are learning about the attribute level for each product. Each time an individual purchases a product she receives a signal of the product's quality, which is her perceived product quality. The signal is drawn from a normal distribution where the mean is the true attribute level and the variance is denoted as the signal noise. Television advertising, which was collected for some households during the final year of the panel, is also allowed to signal product quality. Consumer expected utility for a particular product is a linear function of the product's perceived level, the squared attribute level, the price, and an idiosyncratic error. Learning is identified in this model by the time-series behavior of the share of consumers who repurchase each product among consumers whose previous purchase was the same product. Under learning we would expect this share to rise over time, controlling for any price variation. Initially, the share will consist of consumers who are experimenting with the products, while later on consumers will know their tastes for each product and the repurchase rates will stabilize.

The model is estimated using maximum likelihood, which requires the repeated solution of each individual's dynamic programming problem at the model's state space points. This method of estimating the learning model is extremely computationally demanding, so the paper makes restrictive assumptions about the underlying behavioral model. For example, the paper assumes that individual price coefficients are the same across the population, and that the distribution of prices does not change over time. The effects of such assumptions may not be innocuous. For example, suppose for the sake of argument that the true data generating process has individual-level heterogeneity in price sensitivity, with no consumer learning. If prices and other exogenous variables are constant over time, we would expect the model to estimate large prior variances and low signal noise variances. In the data I analyze, there are three new product introductions, and prices for new products are initially low and then rise over time. When consumers have heterogeneous price sensitivities, we will observe more brand switching right after the new product introductions because price sensitive consumers will be purchasing the new products early in the price cycle, and then switching away as the prices rise. Hence, Erdem and Keane (1996)'s structural model applied to this data would infer that there was learning, even though there is none in the underlying data generating process.

Crawford and Shum (2003) estimate a Bayesian learning model on ulcer medications. Their model is richer than Erdem and Keane (1996) because it 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 argues that illness heterogeneity will segment the pharmaceuticals market, with less sick consumers purchasing cheaper, less effective drugs. Furthermore, consumers with less serious conditions will have less of an incentive to experiment. Learning is identified from the behavior of sick consumers - in particular, the paper argues that the last spell length with a particular drug should be the longest under learning.

As with Erdem and Keane (1996), this paper estimates the model using maximum likelihood. To keep the estimation computationally tractable, the researchers assume that the distribution of unobserved heterogeneity is discrete: in each of the 2 dimensions, consumers fall into a small number of types. This type of heterogeneity may still not be rich enough to properly identify learning in the presence of price variation.

Ackerberg (2003) estimates a learning model that is very similar to Crawford and Shum's in individual-level panel data on a consumer's decision of whether or not to purchase a newly introduced brand of yogurt. This paper focuses on distinguishing two different effects of advertising on consumer utility for the new yogurt: informative (search, product existence, or experience characteristics) versus prestige effects. This paper also extends Erdem and Keane (1996) and allows 2 dimensions of individual-level heterogeneity: the intercept of each consumer's utility for the 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 (2003), who assume that the population distribution of unobserved heterogeneity is discrete, this paper 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: consumers either purchase the new product or they do not. This method would be less tractable in markets where there are multiple new product introductions.

An important point about these papers is that they do not account for any types of dynamics in demand that are not learning. For example, it could be costly for consumers to recalculate their utility if they switch products. This will create habit formation in demand. Habit formation will make brand switching more difficult, and to a researcher who is looking for learning, it may look like there is less learning than is actually going on. Conversely, consumers could have a taste for variety in the product category being examined. This will tend to increase the amount of brand switching in the market, which could to make it look like there is more learning than is actually going on.

A paper that addresses this problem is Israel (2005), which looks for learning in the time-series behavior of departure probabilities from an automobile insurance firm. An empirical fact that is observed in the paper is that the probability a consumers leaves the automobile insurance firm is high after the first non-chargeable claim with the firm, and this probability drops off over time. 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. The paper also does not directly model the forward-looking behavior of consumers; although there is a term in the utility function which is interpreted as a reduced-form value function, there is no solution of the consumer's dynamic programming problem.

There are important aspects of demand that my model takes into account which are not addressed in Israel's paper. First, because the paper only observes consumer tenure with a single firm, it is only possible for the paper to isolate learning when there is positive tenure dependence in demand; this is not possible when the tenure dependence is negative. This is probably not a problem in insurance markets, but it may be important in markets for packaged goods. Second, the paper does not distinguish between consumer lock-in and unobserved heterogeneity in preferences. The researcher may observe a 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. 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 habit formation as opposed to taste heterogeneity. Under habit formation, a temporary price cut will increase a product's future market share; under heterogeneity, this will not be the case. Third, 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 parameters of this term will be a function of policy variables, such as future prices, which will make it impossible to perform "what-if" experiments with the model.

There is a substantial empirical literature in economics and marketing about habit formation and variety-seeking. In economics, perhaps the most well-known work about forward-looking habit formation 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 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 in demand, they usually do not model consumers as forward-looking. I will briefly discuss two exceptions to this.

Chintagunta, Kyriazidou and Perktold (1999) formulate a dynamic model of brand purchase that allows a consumer's previous purchase of a product to affect her current utility. Although consumers are modeled as being forward-looking, the paper shows that under the assumption that consumer's expectations about future variables (such as prices) are independent of their current realizations and some symmetry assumptions, the model can be reduced to a linear utility model. This model is estimated on household panel data of consumer purchases of yogurts.

Hartmann (2005) 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. In the data set, consumers are randomly given coupons which allow them to play golf for a lower price at a specific date in the future. This creates an incentive for consumers to wait and play golf in the future. This paper also allows for a richer distribution of heterogeneity than in the learning papers I have previously discussed. The paper employs a new importance sampling method developed by Ackerberg (2001) to reduce the computational burden induced by the heterogeneity.

3 Theoretical Example

In this section I will present a simple theoretical model of consumer learning and experimentation that nests alternative sources of dynamics in demand by allowing individual consumers to form habits or have a taste for variety, and briefly discuss its testable implications. I will also briefly discuss previous research that finds support for these implications in the same data set I am using. The structural model I estimate nests the model that I will present here: since this model is simpler, it is easier to examine the model's working parts and explain the intuition behind some of its implications. In my model, learning happens when a consumer purchases a new product and finds out her taste for it. If consumers are forward-looking, they will recognize that if they purchase the new product and like it they will be better off in the future. This means that there will be an option value of learning, which will lead to experimentation: consumers will purchase the new product sooner than if they were myopic.

There are two reasons I wish to discuss this simple model and examine its implications. First, as I discussed in the introduction, one of the tasks I wish to perform is to examine the impact of an introductory price cut for a new product on its intermediate run market share (the product's market share in periods after the price is raised) under three different sets of assumptions about the dynamics in demand:

i) consumers only learn and do not form habits;

ii) consumers only form habits, and know their true match values;

iii) consumers learn and form habits at the same time.

The impact of the price cut could be larger in case i) or ii) compared to iii), or it could be smaller. By solving for the option value of learning in these cases, we can get a better idea of when the impact will be larger or smaller. Second, by solving for the model's testable implications we will better understand what type of variation in the data isolates learning from other forces. These implications will still hold in the more complicated structural model, and I will refer to them again in Section 5.3, where I discuss its identification. Further, the fact that support has been found for these implications in previous research in the data set I use suggests that the variation in the data is of the right kind to isolate learning.

Let us consider a market with 2 products. The first, which I denote product 1, is an established product which everyone knows their taste for. The second, which I denote product 2, is a new product which consumers may have to purchase and consume in order to find out how much they like it. The new product in this market is an experience good; other methods of learning, such as learning by search or social learning, are not considered. I assume that the set of consumers in the market stays constant over time, and that consumer purchase one unit of each product every period.[1]

Consumer tastes for each product consist of three parts, as shown in Equation 1: a permanent part which takes learning into account, a part that accounts for habit formation or variety-seeking, and an idiosyncratic component of tastes that is i.i.d. across consumers, products and time.[2]

$$\begin{aligned} \text{Product 1 } &: 0 + \eta_i \mathbf{1}\{y_{t-1} = 1\} \\ \text{Product 2, expected } &: \gamma_i^0 + \eta_i \mathbf{1}\{y_{t-1} = 2\} + \varepsilon_{it} \\ \text{Product 2, taste known } &: \gamma_i + \eta_i \mathbf{1}\{y_{t-1} = 2\} + \varepsilon_{it} \end{aligned} \tag{1}$$

The permanent part of tastes for product 1 is normalized to 0. For product 2, before consumer i has purchased it for the first time, she does not know how much she likes it, but she has a prediction of how much she expects to like it, γ_i^0, that is correct on average. The consumer's true taste or intrinsic match value for product 2, γ_i, becomes known to her when she makes her first purchase of the new product. I assume that at time 0 each consumer is assigned a value of γ_i^0 from $N(\mu^0, (\sigma^0)^2)$, and that when the consumer first purchases and consumes product 2 she receives learns γ_i, which is draw from a normal distribution with mean γ_i^0 and variance σ^2. The parameter σ^2 accounts for the consumer's uncertainty about her true taste draw for product 2. If σ^2=0, then the expected and true taste draws will be the same and there is no learning. I interpret the γ_i as a consumer's match value with product 2. If the products are detergents, then the match value could be how well the product cleans the consumer's clothes. This could be individual-specific since wardrobes may vary across individuals, and different detergents may do better jobs on different types of fabrics.

The term η_i allows dynamics in demand even if $\sigma^2 = 0$. A consumer's utility is increased by η_i if she purchases the same product in period t as she did in period $t-1$. I interpret a positive value of η_i as habit formation (Pollak (1970), Spinnewyn (1981)). Habit formation could arise due to some sort of switching cost or lock-in; for example, there may be costs of recalculating utility if

[1] In my thesis research (see Osborne (2005)), this last assumption is relaxed; the two implications I described in the introduction still hold, and a third implication is derived: consumers will purchase smaller sizes of the new product on their first purchase. Since I do not model size choice in my econometric model, I will not discuss it in the theoretical model either.

[2] The function $\mathbf{1}\{\cdot\}$ returns 1 when its argument is true, and 0 when it is false.

a consumer decides to switch products, which could bias them to purchase the same product over and over. I interpret a negative value of η_i 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.

I assume that consumers are forward-looking and discount the future at a rate $\delta \geq 0$. This means that there when a consumer decides to make a first purchase of the new product, she will look at the future benefits of consuming it: she might like it better than product 1 and continue to purchase it. This means there will be an option value of experimentation, which will be positive when there are no alternative dynamics in demand. If there is habit formation in demand it will be possible for it to be negative, since if the consumer ends up not liking the new product she will lose utility from having to switch brands. The option value of experimentation is also always increasing in σ^2, which will lead consumers to purchase the new product sooner than they would have if $\delta = 0$. I denote this behavior as experimentation.

As I mentioned in the introduction, the option value of experimentation will affect consumer responses to an introductory price cut, which could in turn affect intermediate run market shares. As an example, if consumers are only learners ($\eta_i = 0\ \forall i$ and $\sigma^2 > 0$), a price cut will draw in new consumers, some of whom will find they have a high intrinsic match value (a high γ_i) for the product and repurchase it. If consumers are learners and habit-formers ($\eta_i > 0\ \forall i$ and $\sigma^2 > 0$), it is possible for the price cut to be less effective, since consumers dislike switching brands and will realize if their true match value for the product is low, they will be worse off in the future from having to switch again. It is also possible for the price cut to be more effective under habit formation and learning than learning only if the habit formation is particularly strong. There are two reasons this could happen. First, if the habit formation is strong, then consumers who respond to the price cut and learn that they have a low intrinsic match value may become habituated to the product and will continue to purchase it. Second, if consumers expect to like the new product, the habit formation could actually increase the option value of learning - consumers will want to become habituated to a product they could end up liking very much.

In summary, when there is positive state dependence one of two things can happen to the option value of experimentation:

1. If consumers expect to have a low match value for the product (i.e. γ_i^0 is low), then increasing η_i can decrease the option value of experimentation.
2. If consumers expect to have a high match value for the product (i.e. γ_i^0 is high), then increasing η_i can increase the option value of experimentation.

To see these two cases, I have solved the model above numerically and graphed the option value of learning in Figure 1 for $\eta_i > 0$ and $\eta_i = 0$ for a number of values of γ_i^0. When consumers expect to have low match values for the new product the new product, the option value for $\eta_i > 0$ lies below that of $\eta = 0$.

These numerical findings could be interesting to researchers who are interested in targeted coupons for newly introduced experience goods. For example, suppose that through previous market research, such as observing individual household purchases through the use of magnetic swipe cards, the researcher is able to infer each consumer's η_i. If the researcher knows that an experience good will be introduced to the market, then she will want to target the coupons at consumers who will be more likely to keep purchasing the product in the long run. If consumers on average expect to have low match values for the product, then she should target low η_i consumers; otherwise she should target high η_i consumers.

It is also useful to examine the relative impact of an introductory price cut on a new product's intermediate run market share when there is habit formation only versus habit formation and learning. When I discuss habit formation only, I am referring to the case where consumers know their true taste draws for the new product, and the distribution of true tastes is $N(\mu^0, (\sigma^0)^2 + \sigma^2)$. A firm could potentially neutralize the impact of learning in a market with informative advertising, or by distributing free samples of the new product.

A price cut could be more effective under habit formation only ($\eta_i > 0\ \forall i$ and $\sigma^2 = 0$) as opposed to habit formation and learning ($\eta_i > 0\ \forall i$ and $\sigma^2 > 0$) for the following reason: when there is habit formation only, the price cut draws in consumers who will become habituated to the product and continue to purchase it. When there is habit formation and learning, some of these consumers will find they have a low intrinsic match value for the product and will switch away from it. In this case the firm may want to combine its price cut with advertising in order to remove the learning[3]. As with the case of learning only versus learning and habit formation, it is also possible for the price cut to be more effective under habit formation and learning as opposed to habit formation only. Again, this could occur if the habit formation is particularly strong. When there is only habit formation, consumers who know they have a low intrinsic match value for the new product will be less likely to respond to the price cut. If there is habit formation and learning, these consumers will not know their true match value until they have purchased the new product. They will be more responsive to the price cut and once they find their true match value, the habituation will induce them to keep purchasing the new product.

Another task that may be of interest to researchers is to test for the importance of learning; the null hypothesis for this test is that $\sigma^2 = 0$, while the alternative is that $\sigma^2 > 0$. There are two ways to do this; one is to use simple models to estimate demand and to construct the test statistics associated with the two testable implications I mentioned in the introduction, and will

[3]This argument does not take into account that advertising alone could increase the market share of a new product - if most consumers have low expected tastes, then many of them may not experiment with the product even though their actual match value for the product was high. Advertising could inform these consumers of their high match values and increase the product's intermediate run market share.

describe again in a moment; the other is to estimate the structural model and to directly test if $\sigma^2 = 0$, which is the approach taken in this paper. Although the second approach is more difficult to implement and requires more restrictive modeling assumptions, it has the advantage that we can take the model away from the data and perform "what-if" experiments.

The two testable implications to this model are examined in Osborne (2005), 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 because the option value of experimentation induces consumers to purchase the new product sooner rather than later[4]. When there is no learning, the test statistic will be zero since the order of purchase does not matter. The test may also be used when consumers form habits ($\eta_i > 0$ for all i), but it may be less powerful. The reason for this is that the test statistic tends to be negative when there is no learning and positive η_i; since the test statistic is a continuous function of σ^2, it will still be negative for some values of σ^2 close enough to zero. This turns out to be an issue in Osborne (2005), who finds that the test statistic is in fact negative for one of the new products. Estimating the structural model allows us to shed light on this issue: estimating the structural model allows the researcher to recover the population distribution of habit formation and variety-seeking, the η_i's, and the learning parameter, σ^2, directly.

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. This testable implication is more robust than the first one, because it is true for all values of the discount factor and any type of state dependence in demand. However, the fact that it is true for all values of the discount factor means that it does not tell the researcher about the option value of experimentation. Support for this implication is found for all new products in Osborne (2005).

[4]Since evidence in favor of this implication is found in the data set I use in Osborne (2005), it is reasonable to conclude that for some new products the option value of learning is positive, and that consumers are forward-looking.

4 Data Set

4.1 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 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.

The second table in Table 2 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. Table 3 shows the average prices of different brands at different periods of time. 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.

4.2 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 (Chemical Week, Jan 21, 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.

4.3 Selection of Household Sample

Although there are 1646 households in the total sample, I remove many of them from the sample before estimation. The main reason I do this is to avoid having to deal with inventory behavior. Since laundry detergents are a storable good, some price sensitive households may wait until until

[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).

they observe a low price in the product category before making a purchase. Modeling inventory behavior is computationally difficult (see Erdem, Imai and Keane (2002) for an example), and adding this element to my model of learning and habit formation would make the model computationally intractable. Therefore, I believe it is better to simply remove households who coordinate their purchase behavior with sales so that I do not have to model this behavior. The households that are left in the sample will tend to be households who do not pay attention to store prices unless they have run out of laundry detergent and need to make a purchase in the product category. An added advantage to dropping sale-sensitive households is that the purchase timing of the households who are left can probably be taken to be exogenous. I will be discussing the importance of this point later when I discuss the identification of my structural model. The last advantage to dropping sale-sensitive households is that leaving them in adds a potential source of bias that is similar to the problem of ignoring price sensitive consumers. 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.

In total, around three quarters of the households are dropped, leaving a subsample of 472 households. As I just described, I choose households who appear to be unlikely to make a purchase of any laundry detergent in response to the product category's price being low in the store in a given week. In order to do this, it is necessary to observe whether a household visits a store during a given calendar week. Fortunately, there is a file in the data set that keeps track of a household's daily store visits. Because I observe a household's laundry detergent purchases in a given week as well, I can determine whether a household bought any detergent at all in a given shopping trip.

To determine whether a specific household is sensitive to price drops in its decision of whether to purchase at all, I estimate each household's decision to purchase a laundry detergent separately using binary logit models. There are 1646 households in the entire data set, so I estimate 1646 logit models, where an observation in each logit is a household shopping trip. The dependent variable is whether or not the household chooses to purchase any laundry detergent in that shopping trip or not. I control for average price in the store in the current week[6], average price in the next week, a measure of household inventory, and the number of products on feature and display. Any households whose price coefficients are estimated to be less than zero are dropped from the sample. Also, households who make less than 5 purchases in total are dropped. Multiple brand purchase is also not considered in the paper, so any purchase events that include multiple purchases on the same shopping trip are dropped from the sample (this only accounts for 4% of purchases in the entire sample). Last, any households whose first purchase of the new product occurs at the same time as purchases of other brands of detergent are dropped from the sample. In total, 1174 households are

[6]Some product prices are not directly observed, and must be inferred. This issue is discussed in detail in the Appendix.

dropped, leaving 472 households in the subsample I use for my estimation.

5 Econometric Model

5.1 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 one of the 13 different laundry detergents listed in Table 2. I index each product with the variable j. In a particular purchase event t for consumer i, not all 13 products 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 $j \in J_{it}$ on purchase event t is assumed to be

$$\begin{aligned} & u_{ijt}(s_{it-1}, \alpha_i, p_{ijt}, c_{ijt}, \beta_i, x_{ijt}, \eta_i, y_{ijt-1}, \varepsilon_{ijt}) \\ & = \Gamma_{ij}(s_{ijt-1}, y_{ijt-1}) + \alpha_i(p_{ijt} - \alpha_{ic} c_{ijt}) + \beta_i x_{ijt} + \eta_i y_{ijt-1} + \varepsilon_{ijt}, \end{aligned} \tag{2}$$

where $\Gamma_{ij}(s_{ijt-1}, y_{ijt-1})$ is consumer i's taste for product j. Consumer taste 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:

$$s_{ijt} = s_{ijt-1} + 1\{s_{ijt-1} = 0 \text{ and } y_{ijt-1} = 1\}. \tag{3}$$

For the 10 established products, I assume that consumer tastes do not change over time, so $\Gamma_{ij}(s_{it-1}, y_{it-1}) = \gamma_{ij}$. For identification purposes, I normalize every consumer's taste 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:

$$\begin{aligned} \Gamma_{ij}(s_{ijt-1}, y_{ijt-1}) &= \gamma_{ij}^0 \text{ if } s_{ijt-1} = 0, \text{ and } y_{ijt-1} = 0 \\ \Gamma_{ij}(s_{ijt-1}, y_{ijt-1}) &= \gamma_{ij} \text{ if } s_{ijt-1} = 1, \text{ or } y_{ijt-1} = 1. \end{aligned} \tag{4}$$

The consumer's taste for the new product is γ_{ij}^0 if the consumer has never purchased the product before, and it is γ_{ij} once she has. For the three new products, γ_{ij}^0 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" taste for the product.

I assume that

$$\gamma_{ij} \sim N(\gamma_{ij}^0, \sigma_{ij}^2), \tag{5}$$

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

$$\sigma_{ij}^2 = \sigma_{max} \frac{\exp(\sigma_{0ij} + \sigma_{1j} INC_i + \sigma_{2j} SIZE_i)}{1 + \exp(\sigma_{0ij} + \sigma_{1j} INC_i + \sigma_{2j} SIZE_i)}. \tag{6}$$

Note that there is unobserved heterogeneity in σ_{ij}^2 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_{ij}^2 is always positive and bounded above by σ_{max}, which I assume is equal to 5.[7]

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

$$\alpha_i = \alpha_{max} \frac{\exp(\alpha_{0i} + \alpha_1 INC_i + \alpha_2 SIZE_i)}{1 + \exp(\alpha_{0i} + \alpha_1 INC_i + \alpha_2 SIZE_i)}, \tag{7}$$

where α_{max} is set to -10. α_i is assumed to always be negative and like σ_{ij}^2 it is bounded[8]. p_{ijt} is the price in dollars per ounce 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

$$\alpha_{ic} = \frac{\exp(\alpha_{0ic})}{1 + \exp(\alpha_{0ic})}, \tag{8}$$

where α_{0ic} lies on the real line.

In Equation (2), 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

[7]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.

[8]Again, we might be worried that α_i could be greater than or equal to zero in the population. I will discuss the estimated distribution of α_i's across the population in the model estimates section. The estimated distribution appears to be right-skewed, so this is not likely to be a problem.

in the first week after the introduction of Surf, and j is Surf. The next 11 elements are the same thing for weeks 4 to 14 after the Surf introduction (weeks 2 and 3 were dropped due to identification issues). The next 12 elements of the vector are the same time-product dummy variables for the Dash introduction (weeks 7 and 9 were dropped for lack of identification). These time dummies 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 $t-1$. 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 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 *habit formation* or *variety-seeking*. If $\eta_i > 0$, consumer i's utility is greater if she consumes the same product twice in a row. This behavior is habit formation. If $\eta_i < 0$, the consumer will prefer to consume something different than her previous product choice: I label this as variety-seeking. As with the price coefficient and consumer uncertainty, I allow both observed and unobserved heterogeneity in η_i:

$$\eta_i = \eta_{i0} + \eta_1 INC_i + \eta_2 SIZE_i \tag{9}$$

Last, the ε_{ijt} is an idiosyncratic taste component that is i.i.d. across i, j and t, and has a logistic distribution. I assume this error is observed to the consumer but not the econometrician and 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, α_{0ic}'s and σ_{0ij}'s for the three new products, the intercept of the habit formation 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. This assumption means that the household's uncertainties about tastes for the new products, σ_{ij}^2's, and the price sensitivities α_i's will be transformations of normals as shown in Equations (6) and (7). 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 habit formation, which are $\sigma 1j$ 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.

5.2 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

$$V(\Sigma_{it};\theta_i,\theta) = \max_{\Pi_i} E\left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} u_{ij\tau}(s_{i\tau-1}, p_{ij\tau}, c_{ij\tau}, x_{ij\tau}, y_{ij\tau-1}, \varepsilon_{ij\tau}, \theta_i)|\Sigma_{it}, \Pi_i;\theta_i,\theta\right], \quad (10)$$

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

$$V(\Sigma_{it};\theta_i,\theta) = E_{\varepsilon_{ijt}}[\max_{j\in J_{it}}\{u_{ijt}(s_{it-1}, p_{ijt}, c_{ijt}, x_{ijt}, y_{ijt-1}, \varepsilon_{ijt}, \theta_i) + \delta EV(\Sigma_{it+1};\theta_i,\theta)\}]. \quad (11)$$

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 (11) 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 (5),

ii) future prices,

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]

[9] In my thesis research (Osborne (2005)), 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.

[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

I account for consumer expectations about future prices p_{ijt} 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 (2002) which I will briefly summarize. A detailed description of the estimation of this process can be found in the Appendix. 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.

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 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. 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. I assume that the introductions of new products are a surprise to consumers, so this aspect of

they purchase will function. Future advertising is likely to be more important with "prestige" products, such as shoes or clothing.

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 6.1, 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 receiving a coupon 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 receiving 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^0_{ij} and σ^2_{ij}. Let γ be a random variable with the distribution of true tastes for product j, where $f(\gamma|\gamma^0_{ij}, \sigma^2_{ij})$ is $N(\gamma^o_{ij}, \sigma^2_{ij})$, 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 $v_{ikt+1}(\gamma)$ as consumer i's utility for product k in purchase event $t+1$ as a function of γ, minus the logit error ε_{ijt+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.

$$\begin{aligned}&\text{Product } k = j : v_{ikt+1}(\gamma) = \gamma - \alpha_i p_{ikt+1} + \eta_i y_{ikt} + \delta EV(\Sigma_{it+2}; \theta_i(\gamma), \theta)\\&\text{Product } k \neq j : v_{ikt+1}(\gamma) = \Gamma_{ik}(s_{ikt}, y_{ikt}) - \alpha_i p_{ikt+1} + \eta_i y_{ikt} + \delta EV(\Sigma_{it+2}; \theta_i(\gamma), \theta).\end{aligned} \tag{12}$$

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

$$EV(\Sigma_{it+1}; \theta_i, \theta) = E_{p_{it+1}|p_{it}} E_{J_{it+1}|J_{it}} \left[\int_{\gamma_{ij}} \sum_{l=1}^{L} \ln \left(\sum_{k \in J_{it+1}} \exp(v_{ikt+1}(\gamma_{ij})) \right) f(\gamma_{ij}|\gamma_{ij}^0, \sigma_{ij}^2)\, d\gamma_{ij} \right]. \tag{13}$$

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 (12) will be a function of the true taste γ_{ij} rather than γ and the value function in (13) 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 $t+1$.

5.3 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.

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_{ijt}, has infinite support, eventually everyone in the market will purchase the new product once. After every consumer has experimented with the new product, the only dynamics left in demand will be the habit formation 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 is a habit-former. 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_{ij}^2 and γ_{ij}^0 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, we can see how σ_j^2 is identified by recalling the test statistics associated with the implications of the model discussed in section 3. The first test statistic was the share of consumers who purchase the new product and then do not minus the share of consumers who do not purchase the new product and then do. This share difference is an increasing function of σ_j^2, 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_j^2. If this share difference is greater in the data than the model would predict at $\sigma_j^2 = 0$, then σ_j^2 will pick up that difference.

We can make a similar argument with the second testable implication, which says that among consumers whose previous purchase was the new product, the share of consumers who repurchase the product will rise over time. We know that 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_{ij}^0'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_j^2. 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_j^2 will increase the share of consumers who repurchase the 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 γ_j^0 can be identified from the share of consumers who purchase the new product twice in a row on their first two purchase events after the new product's introduction. It is straightforward to see that this share is an increasing function of γ_j^0.

Now let us relax the assumption that γ_j^0 is constant across the population. In this case we have to identify the mean of γ_{ij}^0 and its variance. The mean is identified from the share of consumers who purchase the new product twice in a row. To identify the variance of γ_{ij}^0, it appears that the third moment that is necessary is the derivative of the probability of a first purchase with respect to the price in purchase event t. In order to calculate an empirical counterpart to this theoretical moment, it is necessary to observe variation in prices across consumers on their first purchase events after the new product's introduction. In previous research (Osborne (2005)), I solve a simple version of the learning model numerically and demonstrate that these three moments appear to be sufficient for local identification of σ_j^2, the mean of γ_{ij}^0, and its variance.

Last, I will relax the assumption that σ_j^2 is constant across the population. I will offer an heuristic explanation of how the distribution of σ_{ij}^2 could be identified, given we have identified the distribution of γ_{ij}^0, η_i and the distribution of non-dynamic parameters. Identification of the variance in σ_{ij}^2 will be obtained from the share difference at points in the new product's price path where its price is high.

To start, I will mention that in previous research (Osborne (2005)) I solve a simple version of my structural model with heterogeneity in σ_{ij}^2, and simulate the share difference for two different price paths for the new product: where its price is high over time, and low over time. I observe the following numerical result: when the population variance in σ_{ij}^2 increases, the share difference at the low point in the price path does not change very much, but the share difference at the high point in the price path increases significantly. In particular, at both low and high price path points when the variance in σ_{ij}^2 increases, the share of consumers who purchase the new product and then do not drops a small amount. At high price paths, the share of consumers who do not purchase the new product and then do drops off much more as the variance in σ_{ij}^2 increases than at low price paths. This provides the key to how we can identify the mean and variance of σ_{ij}^2: first, the share difference at low points in the new product's price path will pin down the mean, since this moment doesn't move around as the variance does. Second, if the share difference at high points in the price path is larger than it should be if the variance in σ_{ij}^2 were zero, then this moment will pin down the variance.

The intuitive reason behind why the share difference is more sensitive to changes in the variance of σ_{ij}^2 at high price paths as opposed to low price paths is as follows. As I mentioned above, most of the change occurs in the share of consumers who purchase do not purchase the new product and then do. At low price paths, the consumers who don't purchase and then do will be price sensitive consumers who get low draws on the product's epsilon and then high draws on it, and consumers who expected to like the new product but got a low epsilon draw on it the first period. When the price of the product is high the price sensitive consumers will move to the share of consumers who don't purchase the new product twice in a row. When the variance in σ_{ij}^2 is raised, the consumers who expect to like the new product will have draws on σ_{ij}^2 from the more extreme ends of the taste distribution. Those whose draws get closer to zero will probably not be affected much - their option value of learning will be lowered, and they will still be sensitive to low error term draws in the first purchase. Those who get higher draws on σ_{ij}^2 will have a higher option value of learning, and will purchase the new product sooner - they will move to the share who purchase twice in a row.

In my data, I will observe this share difference at low price paths and high price paths, since I observe low introductory prices for the new products. A potential problem with this argument is that price sensitive consumers may enter the market when the new product's prices are low, changing the composition of the two share differences and biasing the results. This problem is mitigated by the fact that I have chosen to estimate my model on a group of consumers who appear not to enter the market in response to sales in the product category (see Section 4.3).

6 Estimation Procedure

6.1 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_{ijt}, 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 5.2; 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 $\bar{c}_{ijt}$ 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 $\bar{c}_{ijt}v_{ijt}$, and the vector of population-fixed parameters, θ, contains the parameters p_{cj}^0, p_{cj}^1 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, Sun and Keane (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.[13]

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

[13] 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 habit formation.

6.2 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[14]. In the next few paragraphs I will outline in more detail how the MCMC estimator works, and the next section will describe the functional form of the posterior distribution that is implied by my modeling assumptions.

Denote the vector of model parameters as Θ, and the prior distribution as $k(\Theta)$. In my model, the vector of model parameters contains b, W, the θ_i's for all consumers, the vector of unobserved coupons c_{ijt}, and the population fixed parameters, θ. The b vector is 56-dimensional, the diagonal W vector also contains 56 parameters, and θ is 30 dimensional. These 148 parameters are the model's main parameters. Each θ_i will be a 59-vector, for $i = 1, ..., 472$. This prior is combined with the likelihood of the data, $L(\Theta|Data)$, to form the posterior distribution of Θ given the data,

$$\Lambda(\Theta|Data) \propto L(\Theta|Data)k(\Theta) \tag{14}$$

For my model, the posterior in (14) will not have a convenient form from which to take draws of Θ. Drawing from this posterior can be facilitated using the method of Gibbs sampling. To perform Gibbs sampling, I divide the parameter vector Θ into a series of blocks, where the posterior distribution of a particular block of parameters conditional on the data and the other blocks takes a convenient form. In my estimation, the model parameters are divided into 5 different blocks:

[14] 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.

the first for the θ_i's, the second for the parameters of b, the third for the W, the fourth block of parameters for the vector of all unobserved c_{ijt}'s, and the fifth block for the p_z and the p_{cj}'s. The most computationally burdensome block to draw from is the posterior of an individual θ_i given the data and the other parameters; this posterior does not take a convenient form and requires use of the Metropolis-Hastings algorithm (an overview of this method is presented in Chib and Greenberg (1995)), which is essentially a more complex version of Gibbs sampling. The other conditional posteriors have distributions that are well-known, such as normal or Bernoulli, and are computationally tractable.

The Gibbs sampling algorithm (Casella and George (1992) provides an overview of this method) proceeds by drawing iteratively from the conditional densities of the blocks of Θ. To be precise, denote the five different Θ blocks as Θ_1 to Θ_5. To start the Gibbs sampling procedure, I choose an initial value for blocks 2 to 5, which I denote Θ_2^0,..., Θ_5^0. I then take a draw on Θ_1 from its posterior, which is $\Lambda(\Theta_1|Data, \Theta_2^0, ..., \Theta_5^0)$. Denoting this draw as Θ_1^0, I then draw Θ_2^1 from $\Lambda(\Theta_2|Data, \Theta_1^1, \Theta_3^0, ..., \Theta_5^0)$. The vectors Θ_1^g from $\Lambda(\Theta_1|Data, \Theta_2^{g-1}, ..., \Theta_5^{g-1})$ and Θ_2^g to Θ_5^g from their respective posteriors are a sequence in g that converges to draws from the density of Θ.

The initial draws in the sequence will be affected by the parameters' starting values, and are discarded. Draws that are taken after the sequence has converged to the joint posterior are retained. The MCMC estimator consists of these retained draws from the simulated posterior. They can be used to construct statistics of the parameters, such as their means or standard deviations. In my work, I choose to discard the first 7,500 draws, and I retain every tenth draw out of the next 7,500. To a researcher who is familiar with classical methods, the simulated distribution produced by the Bayesian estimator may seem difficult to interpret. This problem can be overcome by appealing to the Bernstein-von Mises theorem, which states that the mean of the simulated posterior and its variance are asymptotically equivalent to the estimated parameters and the variance matrix produced by classical estimation (see Train (2003), pg. 291-294, for an overview).

6.3 Markov Chain Monte Carlo Blocks: A Short Description

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 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 y_{it} as the vector of observed y_{ijt}'s, c_{it} as the vector of c_{ijt}'s, x_{it} as the vector of x_{ijt}'s and v_{ijt} as the consumer's flow utility minus the logit error. The probability of the consumer's choice in purchase event t will be

$$Pr(y_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) = \sum_{j \in J_{it}} y_{ijt} \frac{\exp(v_{ijt} + \delta EV(\Sigma_{it+1}; \theta_i, \theta))}{\sum_{k \in J_{it}} \exp(v_{ikt} + \delta EV(\Sigma_{it+1}; \theta_i, \theta))}. \tag{15}$$

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

$$\begin{aligned} \Lambda(\theta_i \forall i, b, W, c_{it} \forall i \text{ and } t, \theta) \propto & \prod_{i=1}^{I} \left[\prod_{t=1}^{T_i} \{Pr(y_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) Pr(c_{it}|\theta)\} g(\theta_i|b, W) \right] \\ & \cdot k(b) IG(W) \end{aligned} \tag{16}$$

As I described above I draw from this posterior in 5 different blocks, where each block is convenient to draw from. I will briefly described the distributions of the conditional posteriors of Θ_1 through Θ_5 in the next paragraph. The formulas for the posteriors are given in detail in the Appendix.

The first block draws θ_i for each household conditional on the y_{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 $\prod_{t=1}^{T_i} \{Pr(y_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it})\} g(\theta_i|b, W)$. This distribution is not conjugate, which means that the Metropolis-Hastings algorithm (see the Appendix for the steps I use to implement this) must be used in this step.[15] The next 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 y_{it}. The posterior distribution of θ conditional on θ_i, the $\bar{c}_{ijt}$'s, v_{it} and the y_{it}'s is

$$\prod_{i=1}^{I} \prod_{t=1}^{T_i} \{Pr(y_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) Pr(c_{it}|\theta)\}. \tag{17}$$

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

6.4 Value Function Solution

In this section I will broadly describe how I solve for the value function in Equation (15) using the method of Imai, Jain and Ching (2005). The innovation of this new method is that discrete

[15]Note that when we perform this step, we will need to evaluate the consumer's expected value function in Equation (15), $EV(\Sigma_{it+1}; \theta_i, \theta)$. The procedure I use to do this is described in Section 6.4.

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_{ijt}'s, and the consumer's future tastes for products she has not yet purchased, the γ_{ij}'s. Integrating out the ε_{ijt}'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 (13), 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 $E_{(p',J')|(p,J)} V_n(s', p', J', y', n'; \theta_i^l, \theta)$.[16] Then the consumer's utility for product j at simulation l,

[16] 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

v_{ij}^l, will be

$$\text{Product } k = j : v_{ik}^l = \gamma_{ik}^l - \alpha_i(p_k - c_{ik}^l) + \eta_i y_k + \delta E_{(p',J')|(p,J)} V_n(s',p',J',y',n';\theta_i^l,\theta)$$

$$\text{Product } k \neq j : v_{ik}^l = \gamma_{ik}(s_k) - \alpha_i(p_k - c_{ik}^l) + \eta_i y_k + \delta E_{(p',J')|(p,J)} V_n(s',p',J',y',n';\theta_i^l,\theta),$$

which corresponds to Equation (12).

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

$$\hat{EV}_g(s,p,J,y,n;\theta_i,\theta) = \frac{1}{L}\sum_{l=1}^{L} \ln\left(\sum_{k=1}^{J} \exp(v_{ik}^l)\right). \tag{18}$$

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

$$E_{(p',J')|(p,J)} V_g(s,p,J,y,n,\theta_{i,g},\theta_g) = \frac{\sum_{r=1}^{N(g)} \left[\hat{EV}_r(s,p,J,y,n;\theta_{i,r},\theta_g)\right] k((\overline{\theta}_{i,g} - \overline{\theta}_i^r)/h_k)}{\sum_{i=1}^{N(g)} k((\overline{\theta}_{i,g} - \overline{\theta}_{i,r})/h_k)}, \tag{19}$$

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 (6.4).

7 Estimation Results

The main estimation results are shown in Table 4. 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 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 first and second columns show the mean and standard deviation of this simulated posterior for each element of b; similarly, the third and fourth columns show the mean and variance of the simulated posterior for W. Estimates of parameters that are fixed across the population are also shown in the first column; the third and fourth columns are dashed for

details associated with the value function solution, are described in the Appendix.

these parameters. Although the numbers in the table are posterior means and variances, they can be interpreted in the same way as estimated coefficients and standard errors 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 habit formation are fixed across the population. The first element of the first row shows the population average of consumer tastes for liquid Era, which is -1.545. 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 3.380. 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.

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 -1.092, and this estimate is statistically different from zero. The population variance of predicted tastes is statistically significant, but small relative to the mean at 0.240. 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 1.176, while the parameters on household size and income are positive 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 (6)). As an example, for a household of income 3 and size 3 that has population average value of σ_{0ji}^2, the variance in her true taste for Cheer is $5\frac{\exp(1.176+0.040\cdot 3+0.160\cdot 3)}{1+\exp(1.176+0.040\cdot 3+0.160\cdot 3)}$, which is about 4.28. If the consumer's expected taste for Cheer is -1.092, the population average, then her true taste will be drawn from a $N(-1.092, 4.52)$. 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 close to zero, but its variance is large at 2.913. The coefficients on household size and income, η_1 and η_2, are precisely estimated and positive. 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 2.74, and its variance is 3.00. This means that most households are habit-formers, but a portion of them are variety-seeking. Further, the amount of habit formation is increasing in household size and increasing in household income. The fact that habit formation is increasing in income is consistent with the idea that habit formation may be caused by a cost of recalculating utility: for high income consumers, time is likely more valuable and the cost of recalculating utility may be higher.

The fact that most consumers are habit formers 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. Since most consumers are habit-formers, the price drop will draw consumers away from Tide who will become habituated 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 (7)). The price sensitivity of a consumer with household income of 3 and size of 3 is -7.89. This number may seem large, but since prices are measured in dollars per ounce they range between about 0.02 and 0.05. A puzzling result is that the average population parameter on household income is positive, which suggests that higher income households are more price sensitive. The estimates of the coupon sensitivity parameter, α_{0ic}, show that its mean is -0.591 and its variance is 0.27. 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 (8)). The population mean of α_{ic} is 0.37, and its variance is 0.02, 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 5 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 0.354. 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 $p_{cj}^{0} + p_{cj}^{1} = 0.246 - 0.039 = 0.207$. The probability a consumer gets a coupon for Tide Liquid during the introductory period for Surf Liquid is $p_{cj}^{0} + p_{cj}^{Surf,1} = 0.354 - 0.029 = 0.325$.

7.1 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 4, 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 that, in general, larger and higher income households are more uncertain about their true tastes for the new products. 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.

The first column of Table 6 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.[17] 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

[17] 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 (6)). 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.

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 about the same as Cheer. This may be because Dash was a niche product which was primarily for use in front-loading washers, so consumer uncertainty about the product may have been greater. Second, we can see that the standard deviation 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 7 shows the average consumer uncertainty broken down by household income and size. Overall, an interesting pattern emerges for all three new products: there appears to be more learning among larger and higher income households. For Cheer and Dash, the consumer uncertainty in true tastes is 8 to 9% lower for households with income of less than \$20,000 as opposed to those with income greater than \$60,000. This pattern is less pronounced for household size: for both products the average uncertainty for 1 person households is about 6% lower than for households with 4 or more members. For Surf, the uncertainty among households with income of less than \$20,000 is only about 2% lower than those with income greater than \$60,000, when we do not condition on household size. Conditional on household size, however, the uncertainty in tastes for Surf decreases rather than increases. Consumer uncertainty in tastes for Surf is increasing in household size, whether or not we condition on 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 weekly market share for each product from the simulated choices, averaged over the g draws. The first column of Table 8 shows the average of this simulated market share over all the weeks that the 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 8, and are substantially larger than the shares in the first column: the market share of Cheer rises by 103%, Surf by 105%, and Dash by 244%. 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 5, 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 -1.092 and variance of 0.240 (Table 4). 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 -1.092, and a variance of 4.47. The number 4.47 is the variance in γ_{ij}^0, 0.24, plus the average of σ_{ij}^2 across the population, which is 4.23.

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 (average tastes, no habit formation), assuming consumers expect prices to stay the same over time. This new cutoff is shown by the line $\delta = 0.95$; it can be seen that the option value of learning is not that large, which means that although the total number of consumers who experiment increases when consumers are forward-looking, the increase is not that large. The shaded area to the right of $\delta = 0.95$ line on the expected tastes distribution is much 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 a significant increase in the product's short run market share, even in when consumers are forward-looking. 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.

7.2 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.

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. 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. I then solve for every consumer's value function, assuming that they know the path of future prices, and simulate each persons's choice at each purchase event. This means that I draw new ε_{ijt}'s. 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 habit formation 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 is no habit formation, 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 tabulate the simulated short run market share, which I define to be the first 12 weeks after the new product introduction, and the intermediate run market share, which I define to be the next 24 weeks after the short run, for each new product at constant prices in the first column of Table 9. Simulated revenues are also tabulated, and are shown in brackets beside the market share. To understand the revenue calculation, recall that the price variable is measured in dollars per ounce. I keep each consumer's size choice in each purchase event fixed. Thus, if a person's actual purchase was a 32 ounce bottle of Tide, and her simulated purchase from the counterfactual exercise is Cheer, then I assume that she purchases the 32 ounce bottle of Cheer. Simulated revenue for a product in a given week is the price per ounce for the product multiplied by the total number of ounces sold that week.

The first row of this column shows the short run market share for Cheer at the estimated parameters, which is 22.2. When there is no habit formation, this share drops to 12.2, as shown in the third row. An explanation for this drop is that the high value of σ_{ij}^2 means the option value of learning will be fairly large, so there will be significant experimentation. When there is habit formation some consumers who find that their intrinsic match value for Cheer is low will have formed a habit with the product, and will continue to purchase it; under no habit formation these consumers will switch away from Cheer to something else. If we compare row 1, column 1 to row 5, column 1, we can see that the market share of Cheer is lower when there is no learning as opposed to learning and habit formation. An explanation for this is that when there is habit formation only, there is no option value of experimentation to induce consumers to purchase the new product sooner rather than later. Thus, since consumers will have formed a habit with some established product, they will be less likely to switch into the new product early. We can see that if we compare the short run market share in row 5 column 1 to its intermediate run value in row 6, the market share rises to a value that is close to the intermediate run market share when there is both learning and habit formation. For Surf, the results are very similar, but for Dash, the market share of the product is higher when either learning or habit formation are removed. A possible explanation for this comes from the fact that consumers expected match values for Dash are very low when compared to the other new products. If habit formation is added to learning for Dash, consumers may be even less likely to experiment with it since if they do not like it they will lose future utility from switching brands. Similarly, if there is no learning and only habit formation, consumers who know that they like Dash will purchase it right away, whereas under learning and habit formation these consumers would have expected to dislike the product and would have been unlikely to experiment with it due to the low option value of experimentation.

I compute the effect of a price cut for a new product as follows. I drop the price of the new product by one half for its first three months, holding fixed the number and attributes of competing products. This is a partial equilibrium analysis: I do not take competitor responses into account. I then simulate consumer choices and tabulate simulated market shares for all the new products, which are shown in the second column of Table 9. The third column of this table shows the percentage change in market share resulting from the price cut.

In the first row, we can see that in the first 12 weeks after Cheer's introduction, the price cut results in a 14% increase in market share. The price cut reduces revenue substantially, which is not surprising considering that it is a large cut. The second row shows the intermediate run market share. We can see that the intermediate run market share for Cheer rises by about 1.7% when there is an initial price cut. The price cut causes some consumers to experiment with the new product, and the consumers who like it will continue to purchase it. Some consumers drawn in by the price cut will also become habituated to Cheer. Now let's look at the fourth row, which shows the effect of the price cut on Cheer's intermediate run market share when there is no habit formation. The percentage change in market share is smaller, only 0.5%. Clearly the price drop is more effective when consumers learn and form habits, as opposed to learning only. The reason for this is that, under learning and habit formation, some of the consumers who respond to the price cut will find that they dislike the new product, but the habit formation will induce them to keep purchasing it in the future. Under learning only, consumers who dislike Cheer will switch to something else.

Last, consider the effect of the price cut for Cheer on its intermediate run market share when there is only habit formation, which is shown in the sixth row of the table. In this case the intermediate run market share for Cheer increases more than it does in the learning and habit formation case, by 4.1%. The intuition behind this result is that when there is habit formation only, most of the consumers who are drawn in by the price cut will become habituated to it and will continue to purchase the product. When there is learning and habit formation, some of these consumers will find they dislike the product and will switch away from it. This result suggests that firms should combine their price cuts with advertising or free samples to increase their impact.

For both Surf and Dash, the effect of the price cut is similar to that of Cheer: in the intermediate run, the impact of the price cut is reduced when there is no state dependence, and it is increased when there is no learning.

The second counterfactual, shown in Table 10, 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):

$$\begin{aligned} \gamma_{uij}^0 &= \frac{\frac{\gamma_{ij}^0}{\sigma_{ij}^2} + \frac{a_{ij}}{\sigma_{aj}^2}}{\frac{1}{\sigma_{ij}^2} + \frac{1}{\sigma_{aj}^2}} \\ \sigma_{uij}^2 &= \frac{1}{\frac{1}{\sigma_{ij}^2} + \frac{1}{\sigma_{aj}^2}} \end{aligned} \tag{20}$$

For each product, I assume that the signal variance σ_{aj}^2 is one half of the population variance in Table 6, so that for the Cheer counterfactual σ_{aj}^2 is 2.115, for Surf it is 1.91, and for Dash it is 2.10. This counterfactual is simulated when there is habit formation and no habit formation.

The simulated market shares in Table 10 show an interesting result: for Surf and Cheer, informative advertising reduces the new product's market share in the presence of habit formation, and increases it when there is no habit formation. The reason for this is similar to the reason that the market shares for Cheer and Surf dropped under no learning in Table 9: when consumers have a better signal of how much they will like the new product, their option value of learning is reduced. Because most consumers will have a habit with some established product, they will be even less likely to switch into the new product. When the habit formation is removed, the short run market share of the new products decreases due to the reduced option value of learning, but the intermediate run effect of the advertising is positive. This happens for the same reason that removing learning increased market shares in the simulation experiment discussed in Section 7.1. [18] For Dash, I have calculated the market share for three different time periods rather than just two. The rows labeled Short Run and Intermediate Run show the simulated market share calculated over the same time periods they were for Surf and Cheer; the row labeled Intermediate Run (2) shows the market share for Dash for the entire sample period after the short run period, a period of 62 weeks in length. We can see that for the short run and intermediate run, advertising decrease Dash's market share. However, for the longer intermediate run period, Intermediate Run (2), advertising increases Dash's market share by 3.5%.[19] To understand why this happens, it is best to look at the effect of adver-

[18] It may seem counterintuitive that removing learning increased market shares in Section 7.1, while in the counterfactual experiment advertising reduces the market share. A reason for this is that the simulation experiment performed in Section 7.1 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 habit formation, making the results look more like the no habit formation case.

[19] I have also calculated the intermediate run market share for Cheer and Surf for periods as long as Intermediate Run

tising on the market share for Dash when there is no habit formation, which is in the last three rows of the table. Advertising increases both the short run and intermediate run market shares of Dash, in contrast to Cheer and Surf where advertising decreases the short run market share. 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. The advertising 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 the market share for Dash than for Cheer or Surf. In the presence of habit formation, the advertising decreases Dash's market share initially for the same reason it decreased for Surf and Cheer. Because the advertising has a stronger effect on the market share of Dash, eventually this will outweigh the effect of the habit formation, leading to an increase in the product's market share. In summary, these results suggest that in the presence of strong habit formation, informative advertising will be more effective for niche products.

8 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 habit formation 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 habit formation 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 form habits in addition to learning. 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 form habits, as opposed to when there is no habit formation and they only learn. The impact of the price cut is also greater only form habits than when consumers learn and form habits, which suggests that price cuts may be more effective when they are combined

(2); for these products advertising still significantly decreases the intermediate run market share.

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 habit formation. For a niche product, informative advertising is beneficial.

There are a number of extensions for this research that would be useful. First, the assumption that learning is a one-shot process is possibly restrictive. If learning takes several purchases, a consumer may purchase a new product a few times in a row in order to learn about it. This would tend to positively bias the parameter on habit formation and negatively bias the learning parameter. This could be overcome by allowing the learning to take a longer period of time. Previous literature that estimates structural models of learning with forward-looking consumers has allowed this by modeling the learning process as Bayesian (although this literature does not take alternative sources of state dependence into account). A learning model such as the Bayesian learning model of Crawford and Shum (2000) would fit well in this context, since in that paper consumer match values are heterogeneous. An issue with adding this is that it would complicate the state space for the learning process - instead of just keeping track of which products a consumer had or had not tried, it would be necessary to keep track of her posterior means for each product, and how many times she had purchased each product. Although this is not likely to result in a large increase in computational time, it will increase the model's memory requirements significantly.

It would also be interesting to examine more carefully the supply side under learning and habit formation. For example, the counterfactuals I calculated do not include competitive responses. Also, the price cut I have chosen is somewhat arbitrary and it would be useful to examine the effect of a price cut that is optimal from the firm's perspective on market share. This is a more difficult problem; however, some recent research has emerged which examines firm pricing under learning or state dependence (see Villas-Boas (2002) for an example of firm pricing under learning, and Che, Sudhir and Seetharaman (2005) for the state dependence case).

Last, it would be useful to examine learning in other product categories. For example, learning has been examined in the yogurt product category (Ackerberg (2003)). As I discussed earlier, my estimation results suggest that there is evidence for more learning among smaller and lower income households. It would be interesting to see if this result existed in other product categories as well.

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.

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

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

   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 $Pr(y_{ijt}|\theta_i,c_{it},p_{it},x_{it})$ 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

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

   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 $\overline{c}_{ijt}$'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 5.2. The $\overline{c}_{ijt}$'s are binary, and their distribution is:

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

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

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

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

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

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 each brand 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 availability for all products in all the stores in the data. The data set also 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 this data 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 similar to Nielsen's to impute prices and availability of products in a store during a given calendar week[20]. First, I run through all household purchases and store the price of the product purchased in that purchase event[21]. If no consumer purchases a particular product from a store for an interval greater than 4 weeks, I assume that product is unavailable for that period. Some stores were identified by Nielsen to be stores in the same chain and were observed to have very similar price processes. For these stores, I assume the prices are the same in a given week. If different prices are observed in a given week for the same product in these chain stores, then I assume the true price is the modal price (or the lower if there are multiple modes). Some stores had very few observed purchases, and these stores were not included in the estimation. When a product is assumed to be available, the products shelf price is imputed forwards during the weeks when no purchases are observed. Periodically products are marked below their shelf price, which is

[20]It would also be possible to estimate a price distribution along with the model parameters, treating prices for non-purchased brands as latent unobservables like I did for coupons.

[21]In this step I treat a product as a brand-size. When the final prices are constructed, I average over available sizes for a brand in a store during a given week

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, similar to Erdem, Imai and Keane (2002). An observation in this estimation is the price/availability of a product in a given store during a given week. 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

$$\frac{\exp(\kappa_{0j}+\kappa_{1j}d_1+\kappa_{2j}d_2+\kappa_{3j}d_3+\kappa_{4j}(p_{jt-1}-1/J\sum_{k=1}^{J}p_{kt-1})+\kappa_{5j}(p_{jt-1}-1/J\sum_{k=1}^{J}p_{kt-1})^2)}{1+\exp(\kappa_{0j}+\kappa_{1j}d_1+\kappa_{2j}d_2+\kappa_{3j}d_3+\kappa_{4j}(p_{jt-1}-1/J\sum_{k=1}^{J}p_{kt-1})+\kappa_{5j}(p_{jt-1}-1/J\sum_{k=1}^{J}p_{kt-1})^2)}$$

The d's are dummy variables for the first 3 months after the new product introduction to allow the price process to be different during this time. The price of the product includes the prices of other products to allow competitor response. If the price changes in period t then I assume the density of the price change is

$$\ln(p_{jt-1})=\lambda_{0j}+\lambda_{1j}d_1+\lambda_{2j}d_2+\lambda_{3j}d_3+\lambda_{4j}\ln(p_{jt-1})+\lambda_{5j}[1/J\sum_{k=1}^{J}\ln(p_{kt-1})]+\varepsilon_{itj},$$

where 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

$$\frac{\exp(\zeta_{0j}+\zeta_{1j}a_{jt-1}+\zeta_{2j}(1-a_{jt-1})(p_{jt-1}-\sum_{k=1}^{J}p_{kt-1})+\zeta_{3j}a_{jt-1}(\sum_{k=1}^{J}p_{kt-1}))}{1+\exp(\zeta_{0j}+\zeta_{1j}a_{jt-1}+\zeta_{2j}(1-a_{jt-1})(p_{jt-1}-\sum_{k=1}^{J}p_{kt-1})+\zeta_{3j}a_{jt-1}(\sum_{k=1}^{J}p_{kt-1}))}.$$

I run these estimations in Stata and keep the results in data files my fortran programs can access. Parameter estimates are shown in Tables 11 to 13.

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[22]. 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.

[22] 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.

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 6.4. 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_{ijt} and the set of available products, J_{it}, in a given purchase event. Because there are 13 products, this portion of the state space is high-dimensional. Recall that the expected utility which is calculated in (18) must be retained for future use. During 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 the Appendix). 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

$$E_{(p',J')|(p,J)}V_i(s, p, a, y, n; \theta_i, \theta) \approx \frac{\sum_{m=1}^{M} \hat{EV}_i(s, p^m, J^m, y, n; \theta_i, \theta) f(p^m, J^m|p, J)}{\sum_{m=1}^{M} f(p^m, J^m|p, J)}. \tag{21}$$

This equation is plugged into Equation (19) in the second step of the value function calculation, so the version of Equation (19) that is used in practice is

$$E_{(p',J')|(p,J)}V_g(s, p, J, y, n, \theta_{i,g}, \theta_g) = \frac{\sum_{r=1}^{N(g)} \left[\frac{\sum_{m=1}^{M} \hat{EV}_r(s, p^m, J^m, y, n; \theta_{i,r}, \theta_r) f(p^m, J^m|p, J)}{\sum_{m=1}^{M} f(p^m, J^m|p, J)} \right] k((\overline{\theta}_{i,g} - \overline{\theta_{i,r}})/h_k)}{\sum_{i=1}^{N(g)} k((\overline{\theta_{i,g}} - \overline{\theta_{i,r}})/h_k)}. \tag{22}$$

For the kernel function $k(\cdot)$, I use the Epanechnikov kernel for computational efficiency, and choose $h_k = 2$.

When I estimate the model, I make a simplification to steps 1 and 2. I choose to save $N(g) = 500$ previous value functions. Saving 500 previous value functions at all the state space points for all 472 households will still require a large amount of computer memory. 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 (19) will in practice not depend on i.

References

[1] Ackerberg, D. (2001), "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," Working Paper.

[2] Ackerberg, D. (2003), "Advertising, Learning, and Consumer Choice in Experience Goods Markets: A Structural Empirical Examination", *International Economic Review,* 44 (3), 1007-1040.

[3] Becker, G., Murphy, K. (1988), "A Theory of Rational Addiction," *The Journal of Political Economy*, 96 (4), 675-700.

[4] Becker, G., Grossman, M., Murphy, K. (1994), "An Empirical Analysis of Cigarette Addiction," *The American Economic Review*, 84 (3), 396-418.

[5] Bergemann, D., Valimaki, J. (1997), "Market Diffusion with Two-Sided Learning," *The RAND Journal of Economics,* 28 (4), 773-795.

[6] Casella, G., George, E. (1992) "Explaining the Gibbs Sampler," *The American Statistician,* 46 (3), 167-174.

[7] Chamberlain, G. (1985), "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data,* ed. J.J. Heckman and B. Singer, no. 10 in Econometric Society Monograph series, Cambridge, New York and Sidney: Cambridge University Press, 3-38.

[8] Che, H., Sudhir, K., Seetharaman, P. (2005) "Pricing Behavior in Markets with State Dependence in Demand," Working Paper.

[9] Chib, S., Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician,* 49(4), 327-335.

[10] Ching, A. (2002), "Consumer Learning and Heterogeneity: Dynamics of Demand for Prescription Drugs After Patent Expiration," Working Paper.

[11] Chintagunta, P., Kyriazidou, E., Perktold, J. (1999), "Panel Data Analysis of Household Brand Choice," Working Paper.

[12] Crawford, G., Shum, M. (2000), "Uncertainty and Learning in Pharmaceutical Demand," Working Paper.

[13] Cyert, R., DeGroot, M. (1987), *Bayesian Analysis and Uncertainty in Economic Theory.* Rowman & Littlefield.

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

[15] Erdem, T., Keane, M. (1996), "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science,* 15 (1), 1-20.

[16] Erdem, T., Keane, M., Sun, B. (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.

[17] Erdem, T., Imai, S., Keane, M. (2002), "A Model of Consumer Brand and Quantity Choice Dynamics Under Uncertainty.", Working Paper.

[18] Gabszewicz, J., Pepall, L., and Thisse, J. (1992), "Sequential Entry with Brand Loyalty Caused by Consumer Learning-by-Using," *The Journal of Industrial Economics,* 12 (4), 397-416.

[19] Gelman, A., Rubin, D. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science,* 7, 457-472.

[20] Gonul, F., Srinivasan, K., (1996), "Estimating the Impact of Consumer Expectations of Coupons on Purchase Behavior: A Dynamic Structural Model," *Marketing Science,* 15 (3), 262-279.

[21] Hartmann, W. (2005), "Intertemporal Effects of Consumption and Their Implications for Demand Elasticity Estimates," Working Paper.

[22] Imai, S., Jain, N., Ching, A. (2005), "Bayesian Estimation of Dynamic Discrete Choice Models", Working Paper.

[23] Israel, M. (Feb. 2005), "Services as Experience Goods: An Empirical Examination of Consumer Learning in Automobile Insurance," Working Paper.

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

[25] McAlister, L., Pessemier, E., (1982), "Variety-Seeking Behavior: An Interdisciplinary Review," *The Journal of Consumer Research,* 9 (3), 311-322.

[26] Nelson, P. (1970), "Information and Consumer Behavior," *The Journal of Political Economy,* 78 (2), 311-329.

[27] Osborne (2005), "A Test of Consumer Experimentation and Learning in Packaged Goods Markets," Unpublished Manuscript.

[28] Pollack, R. (1970), "Habit Formation and Dynamic Demand Functions," *The Journal of Political Economy,* 78 (4), 745-763.

[29] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner," *Econometrica,* 55, 993-1033.

[30] Spinnewyn, F. (1981), "Rational Habit Formation," *European Economic Review,* 15, 91-109.

[31] Stiglitz, J. (1989), Imperfect Information in the Product Market, *Handbook of Industrial Organization: Volume 1*, Richard Schmalensee and Robert Willig, eds. Amsterdam: North-Holland.

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

[33] Villas-Boas, M. (2004), "Dynamic Competition with Experience Goods," Forthcoming in *Journal of Economics and Management Strategy.*

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: Market Shares

Powders and Liquids

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

Liquids Only at Different Periods

Period	Actual Time YYYY/MM	Other	Era	Wisk	Tide	Solo	Cheer	Surf	Dash
Entire Sample	1985/12 - 1988/08	0.26	0.12	0.19	0.17	0.06	0.06	0.11	0.03
Before Any Product Intro	1985/12 - 1986/05	0.41	0.14	0.19	0.16	0.10	0.00	0.00	0.00
First Quarter After Cheer	1986/05 - 1986/08	0.24	0.11	0.27	0.11	0.07	0.20	0.00	0.00
First Quarter After Surf	1986/09 - 1986/11	0.24	0.13	0.15	0.17	0.06	0.05	0.19	0.00
First Quarter After Dash	1987/03 - 1987/06	0.24	0.10	0.18	0.10	0.05	0.07	0.15	0.12
Remaining Time	1987/06 - 1988/08	0.24	0.11	0.18	0.21	0.04	0.05	0.12	0.04

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. 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. 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.

Table 3: Average Prices, Adjusted For Coupon Use

Period	Actual Time YYYY/MM	Other	Era	Wisk	Tide	Solo	Cheer	Surf	Dash
Entire Sample	1985/12 - 1988/08	2.80	4.21	2.90	3.97	4.12	3.57	2.67	3.12
Before Any Product Intro	1985/12 - 1986/05	2.56	4.12	3.03	4.41	3.26	.	.	.
First Quarter After Cheer	1986/05 - 1986/08	2.69	3.55	2.79	3.98	4.10	3.13	.	.
First Quarter After Surf	1986/09 - 1986/11	2.91	3.87	3.05	3.10	3.85	3.76	2.01	.
First Quarter After Dash	1987/03 - 1987/06	2.80	4.15	2.88	3.96	4.42	2.90	2.70	3.15
Remaining Time	1987/06 - 1988/08	2.91	4.42	2.88	4.01	4.83	4.07	2.95	3.11

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 4: Parameter Estimates of b and W (Utility Function)

Coefficient	Mean	Standard Err.	Variance	Std. Err.
Taste Parameters				
Era L	-1.545	0.169	3.380	0.507
Wisk L	-1.081	0.130	2.465	0.352
Tide L	-0.754	0.098	1.796	0.249
Solo L	-3.487	0.385	7.531	1.622
Other P	-0.228	0.001	-	-
Tide P	-0.034	0.002	-	-
Cheer P	-1.545	0.137	2.002	0.420
Surf P	-1.235	0.082	0.561	0.218
Dash P	-1.708	0.099	0.113	0.031
Habit Formation (η_{i0})	0.002	0.097	2.913	0.310
H.F. Size (η_1)	0.432	0.002	-	-
H.F. Income (η_2)	0.546	0.003	-	-
Learning parameters				
Cheer, γ_i^0	-1.092	0.060	0.240	0.066
Cheer, σ_{i0}^2	1.176	0.104	0.149	0.068
Cheer - size (σ_{j1}^2)	0.040	0.001	-	-
Cheer - inc (σ_{j2}^2)	0.160	0.002	-	-
Surf, γ_i^0	-0.875	0.081	0.372	0.102
Surf, σ_{i0}^2	0.934	0.061	0.141	0.074
Surf - size (σ_{j1}^2)	0.150	0.003	-	-
Surf - inc (σ_{j2}^2)	-0.033	0.002	-	-
Dash L, γ_i^0	-1.645	0.127	0.370	0.173
Dash L, σ_{i0}^2	1.234	0.071	0.353	0.131
Dash - size (σ_{j1}^2)	0.061	0.005	-	-
Dash - inc (σ_{j2}^2)	0.128	0.003	-	-
Exogenous Variables				
Price Dol/Oz (α_{i0})	1.235	0.099	0.183	0.034
Price - size (α_1)	-0.312	0.002	-	-
Price - inc (α_2)	0.340	0.003	-	-
Coupon Sensitivity (α_{0ic})	-0.591	0.205	0.270	0.074
Feature	0.800	0.077	0.313	0.068
Display	0.878	0.060	0.493	0.091

This table shows the estimated parameters of the consumer flow utility (Section 5.1). 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). The mean and variance are shown in the first and third columns, respectively, and the standard error of the estimates in the second and fourth columns. Some parameters are assumed to be fixed across the population. For these parameters, the third and fourth columns are dashed out. Some utility coefficients, such as the price coefficient and the consumer uncertainty (see Equations (7) and (6)), are transformations of the parameters in the table. Because my model estimation procedure is Bayesian, all the parameter estimates shown are the means of the simulated posterior distribution. The estimates in this table may be interpreted in the same way as those produced by classical procedures.

Table 5: Parameter Estimates: Coupon Probabilities

Coefficient	Mean	Standard Err.
Non-Introductory Periods (p_{cj}^0)		
Other L	0.328	0.004
Era L	0.198	0.014
Wisk L	0.030	0.004
Tide L	0.354	0.010
Solo L	0.030	0.004
Cheer L	0.031	0.004
Surf L	0.246	0.006
Dash L	0.167	0.009
Other P	0.277	0.006
Tide P	0.213	0.007
Cheer P	0.318	0.011
Surf P	0.030	0.004
Dash P	0.030	0.004
Introductory Adjustment		
Cheer (p_{cj}^1)	-0.030	0.004
Surf (p_{cj}^1)	-0.039	0.004
Dash (p_{cj}^1)	0.006	0.001
Est., After Cheer ($p_c^{Cheer,1}$)	-0.029	0.004
Est., After Surf ($p_c^{Cheer,1}$)	-0.029	0.004
Est., After Dash ($p_c^{Cheer,1}$)	-0.029	0.004

This table shows the estimates of the coupon distribution described in Section 5.2. 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 0.246 - 0.039 = 0.207, and the probability of getting a Liquid Tide coupon during Surf's introductory period is 0.354 - 0.029 = 0.325.

Table 6: Average Values of Consumer Uncertainty for New Products

Product	Mean of σ^2	Population Std. Dev.
Cheer	4.23	0.07
Surf	3.82	0.12
Dash	4.20	0.17

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 (6). 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 7: Average Consumer Uncertainty, Across Demographics

Cheer

Size/Income	Less than 20,000	20,000 - 40,000	40,000 - 60,000	60,000+	Averages
1	3.94	4.08	4.21	4.29	4.08
2	4.00	4.11	4.22	4.31	4.22
3	4.05	4.16	4.24	4.34	4.28
4+	4.05	4.15	4.27	4.36	4.31
Averages	3.97	4.11	4.24	4.34	4.23

Surf

Size/Income	Less than 20,000	20,000 - 40,000	40,000 - 60,000	60,000+	Averages
1	3.68	3.65	3.61	3.56	3.65
2	3.80	3.78	3.76	3.72	3.75
3	3.88	3.89	3.87	3.85	3.86
4+	4.05	4.02	4.02	3.98	3.99
Averages	3.74	3.76	3.84	3.85	3.82

Dash

Size/Income	Less than 20,000	20,000 - 40,000	40,000 - 60,000	60,000+	Averages
1	3.94	4.04	4.14	4.21	4.04
2	3.99	4.10	4.18	4.26	4.18
3	3.99	4.13	4.22	4.30	4.25
4+	4.11	4.16	4.25	4.33	4.28
Averages	3.96	4.09	4.21	4.29	4.20

This table shows the average uncertainty in the population for each new product, which corresponds to the variable σ^2 from section 3. They are computed in the same way as the numbers from the previous table.

Table 8: Effect of Removing Learning On New Product Market Share

Product	Predicted Market Share, Learning	Predicted Market Share, No Learning	% Change
Cheer	3.1	6.4	103
Surf	4.3	8.8	105
Dash	1.9	6.4	244

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.

Table 9: Counterfactual: Effect of Introductory Price Cut

Brand	Dynamics in Demand	Time period	No Price Cut	Intro Price Cut	% Increase
Cheer	Habit Formation and Learning	Short Run	22.2 (766.58)	25.3 (436.33)	14% (-43%)
		Int. Run	18.5 (1134.01)	18.8 (1122.78)	1.7% (-1.0%)
	No Habit Formation, Learning	Sh. Run	12.2 (256.83)	14.5 (433.26)	19% (-41%)
		Int. Run	11.6 (727.39)	11.7 (710.96)	0.5% (-2.3%)
	Habit Formation, No Learning	Sh. Run	7.91 (157.59)	9.19 (270.72)	16% (-42%)
		Int. Run	10.6 (659.45)	11.0 (646.27)	4.1% (2.0%)
Surf	Habit Formation and Learning	Sh. Run	18.7 (308.82)	21.5 (536.63)	15% (-42%)
		Int. Run	18.5 (849.61)	18.7 (857.92)	1.4% (1.0%)
	No Habit Formation, Learning	Sh. Run	13.1 (388.76)	15.4 (227.57)	17% (-41%)
		Int. Run	11.9 (712.06)	11.8 (707.73)	-0.5% (-0.6%)
	Habit Formation, No Learning	Sh. Run	8.57 (244.03)	10.0 (142.37)	17% (-42%)
		Int. Run	12.3 (699.15)	12.6 (721.66)	3.1% (3.2%)
Dash	Habit Formation and Learning	Sh. Run	6.23 (129.20)	7.11 (80.04)	14% (-38%)
		Int. Run	6.20 (272.02)	6.25 (274.32)	0.7% (0.8%)
	No Habit Formation, Learning	Sh. Run	6.41 (137.69)	7.32 (85.05)	14% (-38%)
		Int. Run	6.17 (280.33)	6.18 (280.15)	$\approx$ 0.0% ($\approx$ 0.0%)
	Habit Formation, No Learning	Sh. Run	4.84 (100.62)	5.46 (62.39)	13% (-38%)
		Int. Run	6.55 (288.78)	6.68 (295.08)	2.1% (2.2%)

Table shows simulated market shares, revenues in brackets. Short run is the first 3 months after the new product introduction. The intermediate run is period is defined to be the first 6 months after the short run period ends.

Table 10: Counterfactual: Effect of Informative Advertising

Brand	Dynamics in Demand	Time period	No Advertising	Advertising	% Increase
Cheer	Habit Formation	Short Run	22.2 (766.58)	15.1 (529.51)	-32% (-31%)
		Int. Run	18.5 (1134.01)	15.4 (959.59)	-16% (-15%)
	No Habit Formation	Sh. Run	12.2 (433.26)	12.0 (431.39)	-2.1% (-0.4%)
		Int. Run	11.6 (727.39)	11.9 (751.48)	2.2% (3.3%)
Surf	Habit Formation and Learning	Sh. Run	18.7 (536.63)	14.7 (426.96)	-21% (-20%)
		Int. Run	18.5 (1072.55)	16.0 (943.14)	-13% (-12%)
	No Habit Formation	Sh. Run	13.1 (388.76)	12.9 (387.11)	-1.4% (-0.4%)
		Int. Run	11.9 (712.06)	12.0 (728.30)	0.9% (2.3%)
Dash	Habit Formation and Learning	Sh. Run	6.23 (129.20)	5.13 (109.76)	-18% (-15%)
		Int. Run	6.20 (272.02)	6.03 (272.68)	-2.8% (0.2%)
		Int. Run (2)	6.19 (693.56)	6.41 (739.20)	3.5% (6.6%)
	No Habit Formation	Sh. Run	6.41 (137.69)	6.76 (149.63)	5.5% (8.6%)
		Int. Run	6.19 (280.33)	6.81 (318.13)	11% (13%)
		Int. Run (2)	6.29 (724.28)	7.02 (825.78)	12% (13%)

For Dash, the effect of informative advertising is calculated for two "intermediate run" periods. The first intermediate run period is the 6 months after the introductory period. The second is the time after the introductory period until the end of the sample period, a length of 62 weeks. Results from the longer intermediate run period for Cheer and Surf are very similar to those shown for the 6 month period and are omitted from the table.

Table 11: Store Price Process: Probability of Same Price Logit

Product	κ_{0j}	κ_{1j}	κ_{2j}	κ_{3j}	κ_{4j}	κ_{5j}
Other (L)	0.12	0.61*	-0.16	0.33	1.02*	-0.23*
Era	-0.38*	0.61*	0.21	0.53*	0.82*	-0.20*
Wisk	0.17*	-0.02	0.15	0.29	0.31*	-0.29
Tide (L)	0.05	0.41*	-0.27	0.11	0.37	-0.33
Solo	0.51*	0.41	-0.02	-0.08	0.29	-0.27
Cheer (L)	0.88*	-0.61*	-0.13	-0.44*	0.11	-0.07
Surf (L)	0.31*	·	-0.31	-0.47*	0.04	-0.11
Dash (L)	0.45*	·	·	0.08	-0.30	-0.37
Other (P)	-0.29*	0.19	-0.07	0.19	0.04	-0.07
Tide (P)	0.17*	0.01	-0.49*	0.13	0.81*	0.78*
Cheer (P)	0.94	0.90*	-0.25	-0.30	0.21	-0.32
Surf (P)	0.53*	·	0.09	1.10*	-0.19	-0.18
Dash (P)	-0.22	·	·	-1.02*	-1.60	-0.40

Note: the dummy variables for Surf and Dash powder prior to their introduction periods could not be estimated, since there were no observed purchases of these products during these periods in the stores I use to estimate the price process. This might lead the reader to believe that these products were introduced at the same time as their liquid versions. This inference is incorrect: a few purchases of these powders were observed early on in the sample period, however this only happens at stores where very few purchases were made. The details on the construction of the price process variables are described in the Appendix.

Table 12: Store Price Process: Price Change Regression

Product Available in $t-1$

Product	λ_{0j}	λ_{1j}	λ_{2j}	λ_{3j}	λ_{4j}	λ_{5j}
Other (L)	0.64*	0.016	0.08*	-0.03	0.14*	0.40*
Era	0.86*	-0.06*	-0.04*	0.04*	0.34*	0.20*
Wisk	0.86*	-0.04*	-0.03*	-0.03	0.39*	0.07
Tide (L)	0.88*	0.02	0.01	0.02	0.44*	0.06
Solo	0.60*	0.01	0.01	0.02	0.50*	0.017
Cheer (L)	1.90*	0.01*	0.001	-0.04	0.32*	-0.46*
Surf (L)	0.65*	·	-0.01	-0.002	0.44*	0.18
Dash (L)	1.32*	·	·	-0.06*	0.34*	-0.23
Other (P)	0.63*	0.02	-0.02	-0.04	0.48*	0.11
Tide (P)	0.81*	-0.02*	0.002	0.01	0.52*	-0.05
Cheer (P)	1.08*	-0.04	-0.02	-0.01	0.33	-0.04
Surf (P)	1.37*	·	-0.10	0.01	0.22*	-0.17
Dash (P)	-1.06	·	·	-0.04	0.10	1.40

Product Not Available in $t-1$

Product	λ_{0j}	λ_{1j}	λ_{2j}	λ_{3j}	λ_{5j}
Other (L)	0.83	·	-0.06	0.05	0.31
Era	1.16*	0.10	·	-0.12	0.30
Wisk	2.42*	·	0.18	-0.03	-0.61*
Tide (L)	1.81*	0.03	0.03	0.17	-0.10
Solo	1.29*	0.001	-0.10	-0.08	0.30
Cheer (L)	1.54*	-0.05	0.07	-0.12*	0.11
Surf (L)	1.66	·	-0.01	-0.17	0.03
Dash (L)	0.94*	·	·	-0.11*	0.33
Other (P)	0.64*	·	-0.10	·	0.57
Tide (P)	0.68	0.19	-0.23	·	0.47
Cheer (P)	2.09*	0.16	-0.06	0.12	-0.41
Surf (P)	1.39*	·	-0.30	0.12	-0.08
Dash (P)	1.85*	·	·	-0.20*	-0.36

Table 13: Store Availability Process: Probability of Store Stockout Logit

Product	ζ_{0j}	ζ_{1j}	ζ_{2j}	ζ_{3j}
Other (L)	-4.98*	4.19	-0.60	1.62
Era	-1.66*	3.11	-2.09*	0.79
Wisk	-3.53*	11.05*	-1.19*	-3.19*
Tide (L)	-2.34*	7.41*	-1.38*	-1.72
Solo	-3.33*	6.66*	0.69*	-0.29
Cheer (L)	-2.43*	10.41*	-0.26	-3.70*
Surf (L)	-3.76*	12.7*	0.04	-4.50*
Dash (L)	-2.05*	9.50*	0.74*	-3.49*
Other (P)	-5.47*	11.56	0.21	-1.94
Tide (P)	-5.53*	11.67*	-1.87*	-2.23
Cheer (P)	-4.34*	13.90*	-0.50	-3.95
Surf (P)	-3.21*	3.78	-0.64*	1.73
Dash (P)	-1.99*	4.10	0.06	0.90

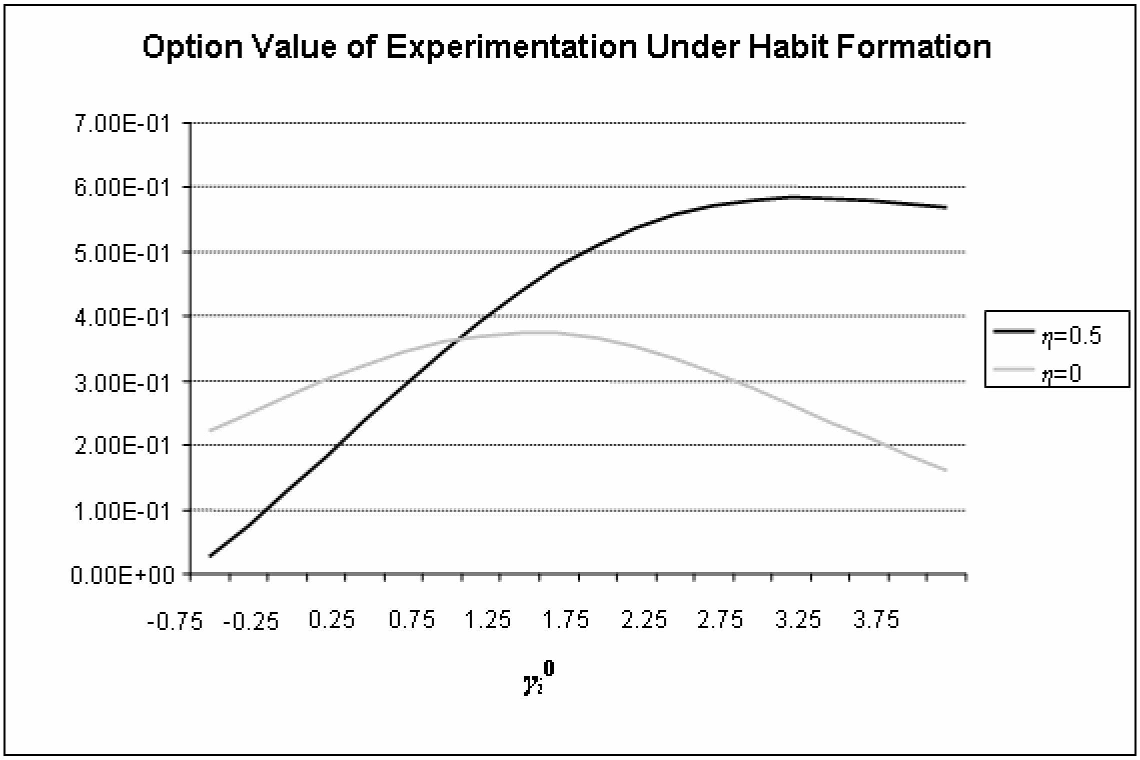

Figure 1: Option Value of Learning

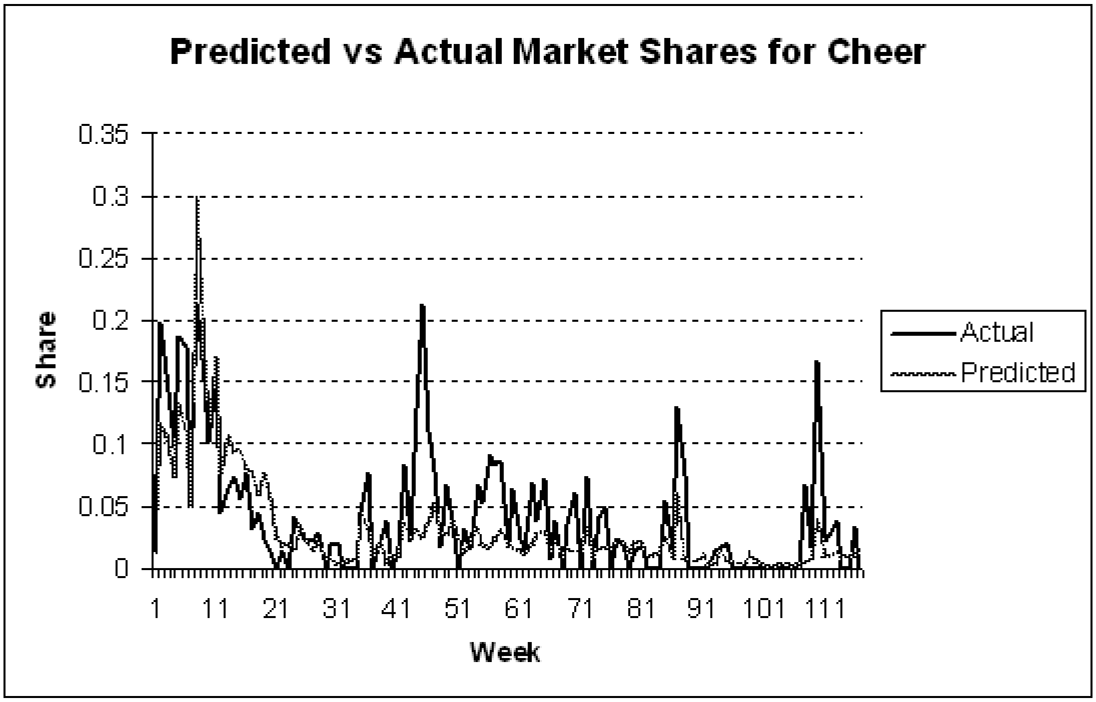

Figure 2: Cheer Predicted vs Actual Market Shares

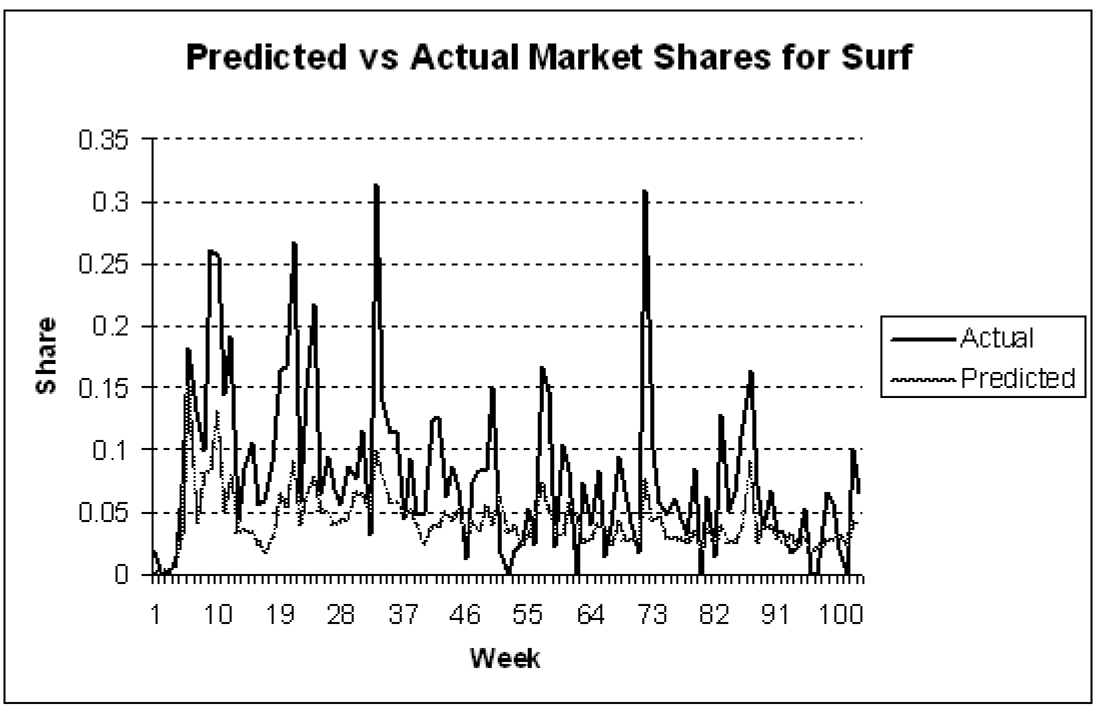

Figure 3: Surf Predicted vs Actual Market Shares

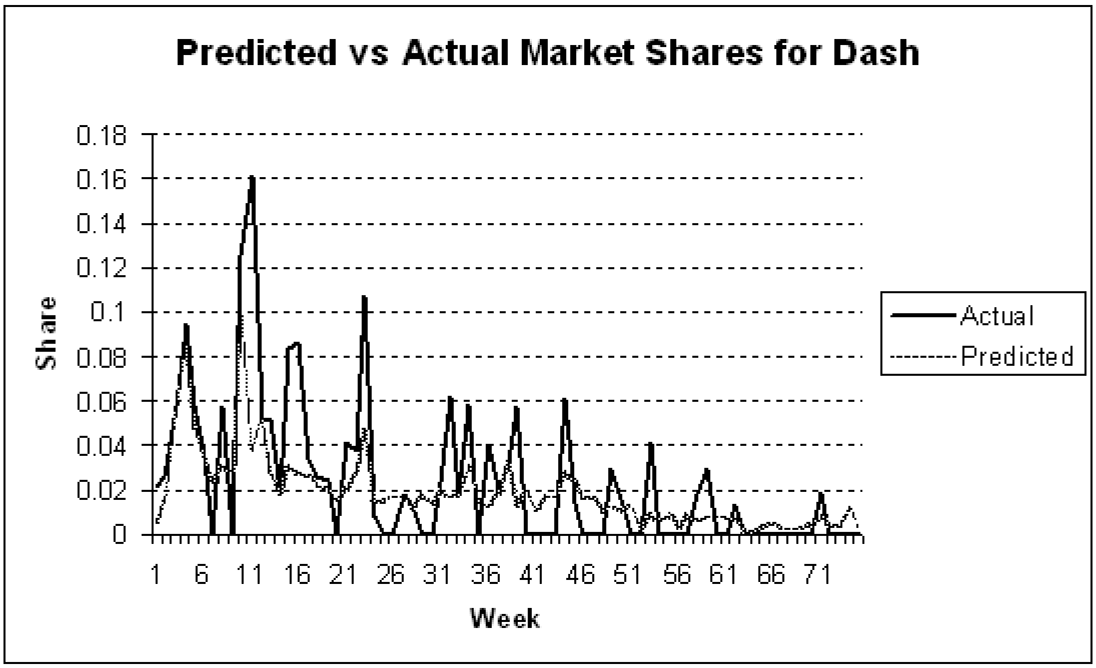

Figure 4: Dash Predicted vs Actual Market Shares

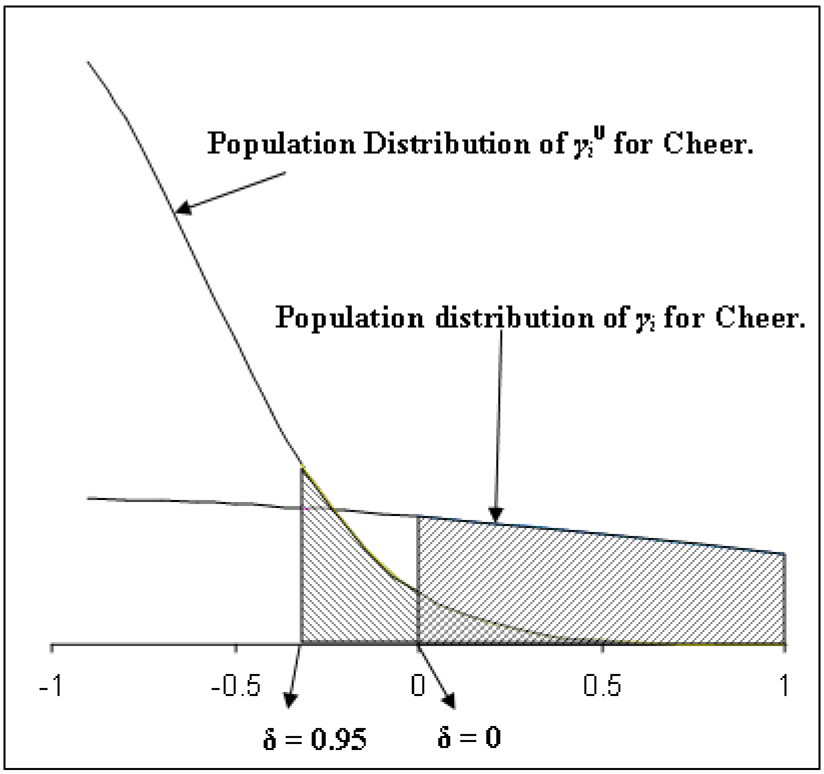

Figure 5: Estimated Taste Distributions For Cheer