

Consumer learning, switching costs, and heterogeneity: A structural examination

Matthew Osborne

Received: 28 January 2009 / Accepted: 6 October 2010 / Published online: 27 October 2010
© Springer Science+Business Media, LLC 2010

Abstract This paper develops and estimates a model of forward-looking consumer learning with switching costs using household level scanner data from a frequently purchased product category. This is novel because current models of consumer purchase behavior assume that only one of these types of dynamics is present, not both at the same time. My model estimates support the presence of both learning and switching costs in this product category. The estimates show that before consuming new products, consumers are unsure of their tastes for them, and subsequently learn their tastes by purchase and consumption of new products. Switching costs are large, comprising roughly 30 percent of the cost of a medium sized package of the product. Additionally, the model incorporates very rich individual level unobserved heterogeneity in price sensitivities, tastes, and switching costs, and the amount by which consumers learn. To show that my model produces different implications than a model with learning or switching costs only, I estimate two more specifications, one without each type of dynamics, and simulate counterfactuals that are of interest to managers and policymakers. I find that intertemporal elasticities

The views expressed are not purported to reflect those of the Bureau of Economic Analysis. I am indebted to my advisors, Susan Athey, Timothy Bresnahan and Wesley Hartmann for their support and comments. I thank Patrick Bajari, Lanier Benkard, Andrew Ching, Liran Einav, Andriy Norets, Dan Quint, and Chuck Romeo for valuable comments. My thanks also go to seminar participants at Stanford University, University of Pennsylvania, Kellogg School of Business, Duke University, UT Austin, Haas School of Business, University of Arizona, Harvard University, the US Department of Justice, the Bureau of Economic Analysis, the 2008 IIOC, and the 2009 SBIES for comments. I 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.

M. Osborne (✉)

Bureau of Economic Analysis, US Department of Commerce, Washington, DC 20230, USA
e-mail: matthew.osborne@bea.gov

are underestimated when either type of dynamics is left out, by as much as 90%. Informative advertising is also affected by the presence of switching costs, although the direction of the bias is not signed. Leaving out dynamics also has a large impact on long-term elasticities, which are used by antitrust policymakers to evaluate the impact of mergers. When learning is ignored, cross elasticities are underestimated by as much as 45%. When switching costs are ignored, both own and cross elasticities are underestimated.

Keywords Consumer learning · Switching costs · Bayesian econometrics · Dynamic programming · Discrete choice

JEL Classification M31 · C11 · D12 · D83 · L41

1 Introduction

Consumer dynamics such as consumer learning by experience and switching costs have played an important role in industrial organization and marketing for many years. Models of consumer purchase behavior in frequently purchased product categories typically include only one or the other type of dynamics. This paper develops and estimates a model of consumer purchase behavior which combines these two types of dynamics into a single model. The model is estimated using household level scanner data on a frequently purchased product category, and the estimates show that both learning and switching costs play a significant role in demand dynamics. Additionally, I demonstrate that leaving out one or the other type of dynamics can result in significant biases in model predictions that are of interest to managers and policymakers, such as the impact of introductory promotions and long-term price elasticities.

In models of consumer learning by experience, individuals are assigned a permanent taste, or match value, for each available product. A consumer learns about her taste for a product by purchasing and consuming 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. Learning produces state dependence in consumer purchase behavior: a consumer's prior purchases will impact her current decisions, because they impact what she believes about her tastes for currently available products. In marketing, learning has been offered as an explanation for state dependence for a long time (Givon and Horsky 1979). Structural approaches, which estimate the parameters of consumer utility functions, such as beliefs about tastes for new products, have been applied to several different product categories.¹

¹Some examples are laundry detergents (Erdem and Keane 1996), pharmaceuticals (Crawford and Shum 2005), automobile insurance (Israel 2005) and personal computers (Erdem et al. 2005).

An alternative way of capturing state dependence in consumer purchase behavior is to model switching costs: one assumes consumers know their true tastes for all available products, but that there is a cost of switching between products. A popular way to specify switching costs in frequently purchased products is to include a parameter in a consumer's utility that lowers it if she purchases a product that is different from her previous choice. Switching costs in markets for frequently purchased goods arise due to consumer-specific investment in a product. This investment can be psychological, as discussed in Klemperer (1995)'s pioneering paper on switching costs. The article argues that brand loyalty can create switching costs, which means that switching costs can play a role in many consumer packaged goods markets. Indeed, models incorporating switching costs of this type have seen widespread use in the empirical marketing literature for frequently purchased products.²

In contrast, in the model of this paper consumers are forward-looking, learn about new products and have costs of switching between products. The model is estimated on household level scanner data of laundry detergent purchases, where three new products are introduced during the period when the data was collected. Learning can be empirically separated from switching costs through differences in the effect of having made a first purchase of a new product on a consumer's current purchase relative to the effect of having used a product in the previous purchase event. The model estimates show that consumers display a significant amount of uncertainty about their tastes for new products before they purchase them. The average cost of switching products, when measured in 1985 dollars, is roughly 1.35, about 30% of the cost of an average sized package of detergent.³

To show how the implications of my model differ from models which do not include either learning or switching costs, I estimate two restricted versions of the model that do not include one or the other form of dynamics. I then run three counterfactual simulations to assess the differences between these models. Two of these quantify consumer response to introductory promotional behavior: I compare the impact of short term introductory price discounts on future market shares across all three models, and the impact of introductory informative advertising on market shares in the full model to those of the model without switching costs. Relative to the model with learning and

²Some product categories where evidence of switching costs have been found are nondiet soft drinks and liquid laundry detergents (Chintagunta 1999), ketchup (Roy et al. 1996), margarine, yogurt and peanut butter (Erdem 1996), breakfast cereals (Shum 2004), and orange juice (Dubé et al. 2008). In the empirical marketing literature, these types of switching costs are equivalently referred to as habit persistence or inertia. Often, when consumer switching costs are modeled, consumers are assumed to be myopic, rather than forward-looking. There are exceptions to this, however; see Chintagunta et al. (2001) and Hartmann (2006).

³It is theoretically possible for switching costs to be negative, a behavior that is often called variety-seeking. My modeling procedure does not restrict the sign of the switching cost. Because the estimation produces evidence for switching costs rather than variety-seeking, for brevity the term switching costs will be used to refer to the type of dynamics that can be modeled by both switching costs and variety-seeking behavior.

switching costs, the no learning model underpredicts the impact of short-term introductory price discounts by as much as two to three times. This happens because learning and switching costs can reinforce each other. A consumer who purchases a new product, believing her match value for it is high, will develop a cost of switching away from it even if she learns her actual taste for the product is low. The model with no switching costs also underpredicts the longer term impact of temporary price discounts on longer term shares: price discounts induce more current purchases, and when switching costs are present it takes longer for consumers to switch away once the discount has ended. I find that the model with only learning mispredicts the impact of introductory informative advertising on new product market shares and revenues, although the direction of the bias can be either positive or negative. The results of these counterfactual experiments have managerial importance: in markets where both of these dynamics are present, leaving out one type of dynamics will lead to incorrect predictions about the impact of promotions. They will also be of interest to industrial organization economists who seek to understand the role of information in markets where switching costs may be present: by ignoring switching costs, researchers will misestimate firms' incentives to use introductory advertising.

I also compute a counterfactual that is of interest to policymakers: the long term price elasticities implied by each model. Own and cross-price elasticities are used by competition authorities to assess the impact of mergers. I find that in both the learning only and switching cost only models, own and cross price elasticities are usually underestimated by leaving out dynamics by as much as 90%. Thus, if competition authorities are examining a merger in a market where both switching costs and learning are present, ignoring either type of dynamics will lead to an underprediction of the impact of the merger on prices. This exercise is in the spirit of recent research in industrial organization which has shown that econometric models which ignore dynamics can lead to incorrect policy implications (for an example, see [Hendel and Nevo 2006](#)). This paper then takes this idea one step further, and shows that even if demand dynamics are modeled, mis-specifying those dynamics can still lead to serious biases.

Biases may also result from failing to model unobserved heterogeneity. As an example of this sort of bias, suppose that consumers are heterogeneous in their sensitivities to price discounts. If, as is common in practice, there are introductory price discounts on new products, price sensitive consumers will purchase the new products initially, and switch away from them to something else when their prices rise. This behavior is also what one would expect if consumers were learning about the new product, and some of them found they disliked it. Thus, ignoring unobserved heterogeneity could lead to overestimates in the amount of learning. To account for this potential bias, I model continuously distributed unobserved heterogeneity in consumer tastes, price sensitivities, switching costs, prior beliefs, and the amount by which consumers learn.

The presence of unobserved consumer heterogeneity raises computational issues in estimation. As discussed above, in my model consumers are forward-looking and take into account the effect of learning and other sources of dynamics on their future utility. Estimating models where consumers are forward-looking is extremely computationally burdensome, and as a result most papers that have tackled these problems have been parsimonious in how they specify consumer heterogeneity, if it is modeled at all.⁴ To tackle this problem, I employ a newly developed Bayesian estimation procedure which makes it possible to estimate dynamic discrete choice models with consumer heterogeneity in a reasonable amount of time (the procedure in this paper is similar to those discussed in Imai et al. 2009 and Norets 2009b).⁵ A related contribution of this paper is that I include unobserved heterogeneity in consumer uncertainty about future match values: some consumers may be very sure about how much they will like a new product, and some may be less so. As far as I am aware, prior work on learning has assumed that the amount of uncertainty about new products is constant across consumers. In Section 3.3, I demonstrate that if one observes multiple new product introductions, it is possible to identify unobserved heterogeneity in the amount by which consumers learn. I find evidence of unobserved heterogeneity in consumer learning.

Another estimation issue arises due to the significant amount of coupon use that is observed in the data set - consumers use a coupon in roughly half of all purchase events. Previous papers in dynamic estimation have not included coupons, because including coupon usage as a right hand side variable raises serious endogeneity concerns: the researcher only observes coupons for products that were purchased, but not for products the consumer did not choose. Thus, coupons for non-purchased products must be treated as latent unobservables and integrated out during estimation, adding significantly to computational time (for an example of this correction in the estimation of static

⁴For example, Erdem and Keane (1996) do not allow for consumer heterogeneity that is persistent over time, while Crawford and Shum (2005) allows for individual level heterogeneity in two dimensions. The paper assumes that the distribution of unobserved heterogeneity is discrete: in each of the two dimensions, consumers fall into a small number of types. Akerberg (2003)'s model of learning in the yogurt product category allows for normally distributed unobserved heterogeneity, but the model is kept computationally tractable since consumer choice is binary: there is only one new product introduction, and consumers either purchase the new product or they do not. This approach would be less tractable in markets where there are multiple new product introductions. An exception is Hartmann (2006), who employs an importance sampling technique developed by Akerberg (2009) to allow for a rich distribution of unobserved heterogeneity.

⁵It would also likely be possible to estimate this model using the method of Akerberg (2009). An interesting topic for future research would be to compare the computational speed and accuracy of these two estimation techniques.

demand models, see Erdem et al. 1999). Not including coupons at all creates measurement error in prices, which will lead to biased estimates of price sensitivities. The estimation procedure I use allows me to easily incorporate coupon usage.

Before turning to the body of the paper, I note that the implications derived in this paper could easily arise in other product categories. A wide variety of other products where learning and switching costs play a role were cited above, and it is not difficult to conceive of others. For example, it is very likely that both learning and switching costs play a role in computer operating systems, a market which has been the subject of a number of high profile antitrust cases. Additional examples of markets where switching costs play an important role are discussed in Farrell and Klemperer (2007); some examples are cellular phones, credit cards, cigarettes, air travel, online brokerage services, automobile insurance, and electricity suppliers. It is certainly conceivable that consumer learning could also play a role in these markets. For example, in cellular phone markets consumers will likely learn about aspects of their provider's service after signing a contract with them. Cellular phone contracts often penalize consumers for switching providers, which creates switching costs and makes the decision to invest in a plan a forward-looking one.

1.1 Related literature

Characterizing state dependence in demand as learning dates back to at least Givon and Horsky (1979), a paper which estimates the parameters of three models of state dependence: a linear learning model, which is presented as a reduced form model of slow learning; a Markov model, which is interpreted as a model of fast learning; and a Bernoulli model of brand choice, which is analogous to a static discrete choice model. The paper finds support for all three models, depending on the product category in question. The modeling approach, while informative about the time series process which best characterizes demand, does not distinguish between different structural models which could underlie that process. For example, fast learning may imply transition probabilities in choices which are consistent with the Markov model. However, switching costs will do this as well. Without controlling for a consumer's first purchase event of a new product and time-varying exogenous variables such as prices, it is difficult to separate these different sources of demand dynamics. The structural approach taken in my paper addresses this concern.

A paper that examines an issue similar in spirit to my paper is Moshkin and Shachar (2002), which distinguishes between two different models of state dependence in demand: a model of asymmetric information, and a pure switching cost model. The paper estimates a model where each consumer's behavior may be explained by one model or the other using television viewing data, finding that most consumers' behavior is better explained by the learning model rather than the traditional switching costs model. There are a number of important

differences between Moshkin and Shachar (2002)'s work and mine. First, the model of learning is different: firms offer new products every period and consumers learn via search rather than through experience; learning through experience is important for goods that are repeatedly purchased. Second, consumers are myopic in the model of Moshkin and Shachar (2002), which is probably a reasonable assumption for television viewing choices; however, as I have argued above, when consumers are learning by experience modeling forward-looking behavior is key. Modeling forward-looking behavior is a more difficult problem computationally. Third, my model is more flexible: in Moshkin and Shachar (2002) consumers are designated to display either learning, or switching costs, but not both behaviors at the same time as my model allows. Because I model unobserved heterogeneity in both switching costs and learning, my model nests that of Moshkin and Shachar (2002) as a special case.

There are also examples of models which nest learning and switching costs being applied to markets for services. Israel (2005) proposes and estimates a model of consumer learning and lock-in for an automobile insurance company. Very recent work by Goettler and Clay (2010) estimates a structural model of learning and switching costs on tariff choice data for an online grocer. Both of these papers are important contributions to our understanding about demand dynamics, but they differ in both the questions they answer and the complexity of the modeling approach. Most importantly, these papers provide evidence on these dynamics in the demand for services; I am the first to find evidence for both types of dynamics in a frequently purchased product category, which are the subject of a significant amount of interest in marketing and industrial organization. Also, there are a number of complications which arise when modeling learning with switching costs in packaged goods markets. In the two previous papers, the consumer learning process and the decision process is much simpler: there is a single service, and consumers are learning about their preference to use the service. In the case of Israel (2005), consumers are deciding whether or not to stay with the auto insurance company; in Goettler and Clay (2010)'s case they are deciding between three different tariff schedules for the grocer. Because these papers examine demand for a single product, it is impossible for them to compute some of the counterfactuals I do, such as cross-price elasticities. Including multiple products also adds significantly to the computational difficulty of the exercise. In my model, there are 30 different products, and three new products for consumers to learn about, making the model's state space much larger.⁶ Finally, neither paper characterizes the biases that arise from leaving out one or the other type of dynamics by estimating restricted models.

⁶Additionally, Israel (2005)'s approach models the value function in rather than solving for it explicitly as I do. The estimated parameters of that paper are potentially subject to the Lucas critique: they will be functions of policy variables, such as the distribution of future prices, which makes performing counterfactuals problematic.

2 Data set

2.1 Discussion of the scanner data

The data set I am using is A.C. Nielsen supermarket scanner data on detergent purchases in the city of Sioux Falls, South Dakota between December 29, 1985 and August 20, 1988. This data is particularly useful for identifying consumer learning and switching costs 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.

During the time the data set was collected three large companies dominate the market: 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. Market shares of all the brands used in the analysis are shown in Fig. 1. Smaller brands of each type I group into an Other category. Liquid detergents comprise roughly fifty percent of all units sold. Additionally, well known brands, such as Wisk and Tide, have high market shares.

Table 1 shows the market shares of selected brands of liquids over different periods of time. It is notable that for all three new products, their market share tends to be significantly higher in the first 12 weeks after introduction than it is for the remainder of the sample period. This fact is consistent with learning, since the option value of learning induces consumers to purchase new products early. However, it is also consistent with consumer response to introductory pricing. The average prices of different brands at different periods of time are shown in the same table, underneath the shares. There are two noteworthy facts in this table. First, prices of the new brands Cheer and Surf tend to be lower in the first 12 weeks after introduction than they are later on in the data. This fact suggests that we should be aware of possible biases due to consumer heterogeneity: for example, price sensitive consumers could purchase the new products initially when they are cheap, and switch away from them as they get more expensive, which could be mistaken for learning. Second, when Cheer is introduced to the market by Procter and Gamble, the price of Wisk, a popular product of Unilever, goes down. Similarly, when Unilever's Surf is new, Procter and Gamble's Tide drops in price. Cheer and Surf have been successful products since their introductions, but Dash was discontinued in the United States in 1992. One possible reason for this is that Dash was more of a niche product: it was intended for front-loading washers, which constituted about 5% of the market at the time.

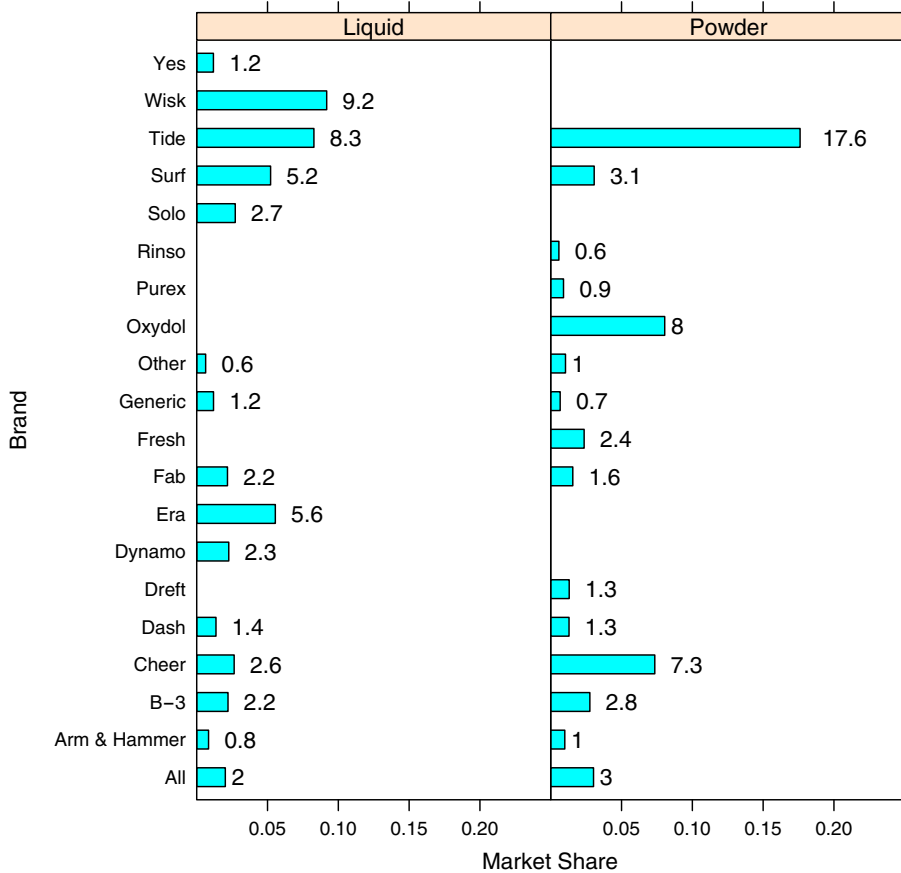


Fig. 1 Market shares of major detergent brands

Table 2 summarizes household level information that is relevant to learning and brand choice. The first panel summarizes the number of purchase events that are observed for each household. As will be discussed in further detail below, separating out learning from switching costs will require observing a time series of purchases for each household. The fact that the average number of observed purchases is 22 is heartening. The next panel in the table shows the number of different brands a household is ever observed to purchase. 75% of households purchase at least 3 different products, and the median household purchases 4. The third panel of the table shows the fraction of households who ever make a purchase of one of the new products. It is notable that, even though the new products are heavily discounted on their introduction, and there is potentially value to learning about them, most households never purchase the products. One explanation for this is that most households have a favorite product, and the value of learning about a new product is not high enough to induce them to experiment. It is also possible

Table 1 Market shares, average prices: selected liquid brands at different periods

Period	Actual time YYYY/MM	Wisk	Tide	Cheer	Surf	Dash
Entire sample	1985/12–1988/08	0.19	0.17	0.06	0.11	0.03
		2.90	3.97	3.57	2.67	3.12
Before any product intro	1985/12–1986/05	0.19	0.16	0.00	0.00	0.00
		3.03	4.41	–	–	–
First quarter after cheer	1986/05–1986/08	0.27	0.11	0.20	0.00	0.00
		2.79	3.98	3.13	–	–
First quarter after surf	1986/09–1986/11	0.15	0.17	0.05	0.19	0.00
		3.05	3.10	3.76	2.01	–
First quarter after dash	1987/03–1987/06	0.18	0.10	0.07	0.15	0.12
		2.88	3.96	2.90	2.70	3.15
Remaining time	1987/06–1987/06	0.18	0.21	0.05	0.12	0.04
		2.88	4.01	4.83	2.95	3.11

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

that switching costs make it less likely consumers will experiment with a new product. Consumers who purchase the new product usually make their first purchase during the first 12 weeks after the product's introduction, as shown in the fourth panel. The fact that consumers disproportionately make

Table 2 Summary statistics: household purchase behavior

Total number of observed purchase events					
Min	First Q.	Median	Mean	Third Q.	Max
1	10	18	22.47	30	137
Number of different products ever purchased					
Percentile	5	25	50	75	95
	1	3	4	6	10
Fraction of consumers who ever purchase					
	Cheer	Surf	Dash		
	25%	41%	16%		
Fraction of consumers whose first purchase is					
First 12 weeks after intro	53%	39%	57%		
Fraction of switches after					
First purchase	0.78	0.80	0.83		
Non first purchase	0.62	0.68	0.70		

The sample used is the entire sample of purchase events, which contains 38,123 observations and 1,693 households. The fraction of switches panel shows the fraction of purchases of a product where the subsequent purchase is observed to be a different product

their first purchases soon after the product's introduction is consistent with there being value to learning about the new product. Household switching behavior also provides evidence for learning, and this is shown in the last panel of the table. A purchase event is denoted as a switch if a household purchases a different product in their subsequent purchase. The panel shows the fraction of purchases of each new product where a switch occurs, split up by whether the purchase is a household's first purchase of the product, or non-first purchase. For all three products, households are much more likely to switch to something else after their first purchase, as opposed to later purchases. This is consistent with household experimentation: after a household's first purchase, the household learns how much they like the product. If they dislike it, they switch to something else. Households who make second purchases tend to like the product and will be more likely to repurchase it. A caveat is that these statistics are only suggestive of the presence of learning. The last statistic could also reflect consumer heterogeneity in price sensitivities. Since we observe introductory pricing for the new products, most consumers' first purchases will occur when the product is on discount. Price sensitive consumers will purchase the products when they are cheap and will subsequently switch away from them. More sophisticated tests for the presence of learning which allow the researcher to account for unobserved heterogeneity are discussed in Osborne (2006). Evidence of learning in this data set is found in that work. The approach taken in this paper is also able to identify learning in the presence of consumer heterogeneity.

2.2 An overview of the laundry detergent market prior to 1988

The fact that the three new products were liquid detergents was not a coincidence, and to see why it is useful to briefly discuss the evolution of this industry. The first powdered laundry detergent for general usage to be introduced to the United States was Tide, which was introduced in 1946. Liquid laundry detergents were introduced later: the popular brand Wisk was introduced by Unilever in 1956. The market share of liquid laundry detergents was much lower than powders until the early 1980's. The very successful introduction of liquid Tide in 1984 changed this trend, and detergent companies began to introduce more liquid detergents. Product entry in this industry is costly: an industry executive quoted the cost of a new product introduction at 200 million dollars (Cannon et al. 1987). Industry literature suggests a number of reasons for the popularization of liquids during this time: first, low oil and natural gas prices, which made higher concentrations of surfactants⁷

⁷The 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).

more economical; second, a trend towards lower washing temperatures; third, increases in synthetic fabrics; fourth, on the demand side, an increased desire for convenience. In the third and fourth points, liquids had an advantage over powders since they dissolved better in cold water, and did not tend to cake or leave powder on clothes after a wash was done.

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

3 Econometric model

3.1 Specification of consumer flow utility

Below I describe the elements of my model of consumer learning and switching costs. In my structural econometric model an observation is an individual consumer's purchase event of a package of liquid or powdered laundry detergent. In the following discussion, I index each consumer with the subscript i , and number the purchase events for consumer i with the subscript t . The dependent variable in this model is the consumer's choice of a given size of one of the 30 different laundry detergents listed in Fig. 1.⁸ I index each product with the variable j , and each size with of the product with s .⁹ In a particular purchase event t for consumer i , not all of the choices may be available. I denote the set of products available to consumer i in purchase t as J_{it} . I assume that a consumer's period utility is linear, as in traditional discrete choice models. The

⁸Although the Other product category is an amalgamation of smaller brands, grouping these brands together is unlikely to contribute much bias to the model estimates, because the Other category of products only comprises about 1.6 percent of the total market share.

⁹For liquids, each s denotes one of the five most popular sizes, which comprise over 99 percent of liquid purchases. Package sizes in powders were much more disaggregated, so I grouped package sizes into 5 categories. For powders, s denotes one of the size categories.

period, or flow utility for consumer i for product-size $(j, s) \in J_{it}$ on purchase event t is assumed to be

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

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

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

For the 27 established products, I assume that consumer match values do not change over time, so $\Gamma_{ij}(s_{it-1}, y_{it-1}) = \gamma_{ij}$. For identification purposes, I normalize every consumer's match for powder All (product 1) to 0. For the three new products, I assume that the evolution of the consumer's permanent taste is as follows:

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

The consumer's match value 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 match with the product.

I assume that

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

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:

$$\sigma_{ij}^2 = \sigma_{\max} \frac{\exp(\sigma_{0ij} + \sigma_{1j} \ln(INC_i) + \sigma_{2j} \ln(SIZ E_i))}{1 + \exp(\sigma_{0ij} + \sigma_{1j} \ln(INC_i) + \sigma_{2j} \ln(SIZ E_i))}. \tag{5}$$

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 3. Household size, the variable $SIZ E_i$, also varies from 1 to 4 and is defined similarly. Note that

Table 3 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

σ_{ij}^2 is always positive and bounded above by σ_{\max} , which I assume is equal to 10.¹⁰

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

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

where α_{\max} is set to -500 . α_i is assumed to always be negative and, like σ_{ij}^2 , it is bounded. p_{ijst} is the price in dollars per ounce of size s of product j in the store during purchase event t , and the variable c_{ijt} is the value of a manufacturer coupon for product j that consumer i has on hand in purchase event t , also measured in dollars per ounce. The parameter α_{ic} is consumer i 's sensitivity to coupons. I assume that α_{ic} lies between 0 and 1, and that

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

where α_{0ic} lies on the real line.

In Eq. 1, β_i is a vector that measures consumer i 's sensitivity to other variables, x_{ijt} . The first and second elements of the x_{ijt} vector are dummy variables which are equal to 1 if product j is on feature or display, respectively. The third element is a dummy variable that is 1 if purchase event t occurs in the first week after the introduction of Cheer, and j is Cheer. The fourth is the same thing for the second week of Cheer, the fifth for the third and so on up to the fourteenth week after the Cheer introduction. The next element is a dummy variable that is 1 if purchase event t occurs in the third week after the introduction of Surf, and j is Surf. The next 11 elements are the same thing for weeks 4 to 14 after the Surf introduction. The next 14 elements of the vector are the same time-product dummy variables for the Dash introduction. These time

¹⁰During initial runs of the model I simply exponentiated $\sigma_{0ij} + \sigma_{1j} \ln(INC_i) + \sigma_{2j} \ln(SIZ E_i)$ in order to ensure a positive variance. I found that occasionally when the estimation algorithm was traversing the parameter space it would choose a value of σ_{0ij} , σ_{1j} , or σ_{2j} that, when exponentiated, lead to a variance large enough to cause numerical problems. The upper bound of ten was chosen to avoid these numerical problems, but to be large enough not to be binding. Indeed, as I will show below, the final estimates are nowhere near this bound.

dummy variables are included to capture the effect of unobserved introductory advertising for the new products.

The consumer's utility in purchase event t is increased by η_i if she purchases the same product that she did in purchase $t - 1$. 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 learning process I estimate in this paper is simpler in one dimension than the learning process which has been used in some other recent papers which estimate structural learning models, such as Erdem and Keane (1996) and Akerberg (2003). In my learning process, consumers learn their true taste for the new product immediately after consuming it, while the previously cited papers model learning as a Bayesian updating process, which allows the learning to take place over several periods. Although my specification is more restrictive in this sense, I believe this restriction is reasonable for two reasons. First, consumers use laundry detergent several times in between purchases. The median number of weeks between purchases is 8, and consumers likely use laundry detergent on a very regular basis, at least once per week. Thus, in between purchases consumers will have had 8 or more consumption experiences with each product. This means that if a consumer purchases a new product for the first time, by her second purchase it is reasonable to assume that most or all of her uncertainty about the product will have been resolved. Thus, even if the underlying learning process is not a one shot model, the one shot model likely provides a very good approximation to the amount of learning that occurs between purchases. Secondly, both Erdem and Keane (1996) and Akerberg (2003) actually find that consumers learn their tastes for the new products very quickly: their estimates suggest that most consumer uncertainty is resolved after one purchase. Hence, my assumption is consistent with their findings. Even with one shot learning, my model is still very complex. Making the learning process a slower Bayesian updating process would make the estimation even more cumbersome, and for the reasons given above would not likely change the results very much. Additionally, my model of learning is more sophisticated than those of the previously cited papers in a different dimension: I allow for continuously distributed unobserved heterogeneity in both the mean of consumer priors, and the variances. In contrast, both of the previously cited papers assume that consumer priors are the same for all consumers. Thus, my model allows for some consumers to be more sure of their *ex-ante* true tastes than others, and for some consumers to expect to like the new products more than others. I think this is an interesting innovation; as I will discuss below, the effectiveness of promotions for new products will depend on consumers' initial beliefs and hence it will be important to model them flexibly.

The term η_i accounts for the dynamic behaviors of *switching costs* or *variety-seeking*. If $\eta_i > 0$, consumer i 's utility is greater if she consumes the same product twice in a row. Thus, a positive η_i induces a switching cost (Pollack 1970; Spinnewyn 1981). An alternative way to model switching costs would be to subtract a positive η_i from all products except the one that was previously chosen; since utility functions are ordinal and there is no outside good in this model, these two formulations are equivalent. As discussed in the introduction, switching costs have been found to be an important part of demand for consumer packaged goods. They could arise due to brand loyalty developing over time; an alternative explanation for switching costs in packaged goods markets is that they may proxy for costs of recalculating utility if a consumer decides to switch products. A consumer who is shopping for a large number of products and is pressed for time may more easily recall her utility for the product she last purchased, which would bias her to repurchase it. If $\eta_i < 0$, the consumer will prefer to consume something different than her previous product choice: I label this as variety-seeking (McAlister and Pessemier 1982). Variety-seeking is not likely an important behavior in laundry detergent markets, but I allow it in the model for the sake of generality. As with the price coefficient and consumer uncertainty, I allow both observed and unobserved heterogeneity in η_i :

$$\eta_i = \eta_{i0} + \eta_1 INC_i + \eta_2 SIZ E_i. \quad (8)$$

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

I model unobserved heterogeneity in a significant number of the individual-level parameters. In trial runs of the model, I found it was difficult to identify unobserved heterogeneity in many of the smaller brands, so I only model unobserved heterogeneity in 13 of the 30 brands. I also allow for unobserved heterogeneity in γ_{ij}^0 's, the α_{0i} 's, the feature and display variables, the intercept of the switching costs parameter η_{i0} , and the σ_{0ij} 's. 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.¹¹ 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 Eqs. 5 and 6. Their distribution

¹¹A possible worry may be that the assumption of normally distributed heterogeneity is too restrictive. Osborne (2006) presents the results of the estimation of an extended model where some of the heterogeneity is assumed to be a two point mixture of normals. Identification of some parameters becomes more difficult in this case, but generally the results do not change, which suggests that the assumption of normality is a good fit.

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 with small shares, the coefficients on household demographics for the learning parameters, the price sensitivities and the switching costs, which are σ_{1j} and σ_{2j} , α_{1j} and α_{2j} and η_1 and η_2 respectively, and a group of parameters which capture consumer expectations of future coupons c_{ijt} . These latter parameters will be discussed further in the next section. I denote the vector of population-fixed parameters as θ .

A feature of the model that the reader may have noted is that there is no outside good. This means that if the prices of all products were to increase, the total amount of laundry detergent sold would not decrease. In reality, if the price of all detergents increased significantly, consumers would either switch to a substitute, or wash their clothes less often. Since substitutes for detergent are not readily available, one would expect a quantity response. However, it seems probable that quantity response would be inelastic: most consumers likely do laundry on a regular basis, once or twice a week. Thus, if a price increase in a product is observed, consumers will likely switch to another brand, rather than washing their clothes less. In this paper, I only consider partial equilibrium counterfactuals, that is, the impact of changing the price of one product on consumer switching, holding fixed the prices of other products. Since this is going to lead to brand switching, rather than overall quantity reduction, not modeling an outside good should not impact the implications I examine. This issue would be a greater concern if I was calculating market equilibria, though. An additional concern is that leaving out the outside good, which in this case is consuming less detergent, could impact my parameter estimates. Hendel and Nevo (2006) shows that if consumer utility is linear in tastes for brands, and discounting is low, then the consumer's quantity decision and the brand choice decision can be modeled separately. These are both true in the model I have discussed, which alleviates this concern. A final issue I wish to note is that although I have not modeled usage, I have modeled quantity choice. An alternative approach would be to model brand choice conditional on size choice, or to construct an average price for each brand. I choose to model size choice because the prices of sizes vary over time, and a consumer who prefers a certain brand may choose a different size than they normally would if that size happens to be on sale. Thus, modeling size choice in this way avoids potential measurement error in prices.

3.2 Consumer dynamic optimization problem

I assume consumers are forward-looking¹² and in each purchase event they maximize the expected discounted sum of utility from the current purchase

¹²In Osborne (2006), evidence is provided that consumers are forward-looking in this data set.

into the future. The consumer's expected discounted utility in purchase event t is

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

where Π_i is a set of decision rules that map the state in purchase t , Σ_{it} , into actions, which are the $y_{ij\tau}$'s in purchase event t . The parameter δ is a discount factor, which is assumed to equal 0.95.¹³ The function $V(\Sigma_{it}; \theta_i, \theta)$ is a value function, and is a solution to the Bellman equation

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

The state vector in purchase event t , Σ_{it} , has the following elements: the $S_{ij\tau-1}$'s for the new products, the $y_{ij\tau-1}$'s for all 30 products, the prices of all products, $p_{ij\tau}$, 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 Eq. 10 will be taken over the distributions of future variables, which are

- i) the true tastes for new products the consumer has never purchased, as in Eq. 4,
- 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 $x_{ij\tau}$'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.¹⁴

¹³The discount factor is usually difficult to identify in forward-looking structural models, so it is common practice to assign it a value. Since the timing between purchase events varies across consumers, it is possible that the discount factors may also vary across consumers. As I will discuss in a few paragraphs, I assume that all consumers have the same expectations about when their next purchase will occur, which removes this problem. Also, the estimation method I will use requires that the discount factor is not an estimated parameter.

¹⁴Assuming that consumers do not expect future advertising is probably not that unrealistic in the laundry detergent market. For this product category, it is likely that consumers will care more about future prices and how well the product they purchase will function. Future advertising is likely to be more important with "prestige" products, such as shoes or clothing.

I account for consumer expectations about future prices p_{ijst} and product availability J_{it} in the following way. I estimate a Markov transition process for prices and availability from the data on a store-by-store basis, using a method similar to Erdem et al. (2003) which I will briefly summarize. A detailed description of the estimation can be found in the paper's Technical Appendix, Section 3, which is available on request from the author. I assume that consumers' actual expectations about these variables are equal to this estimated process. This process captures consumer expectations about the values prices will take when they run out of detergent and need to purchase it again. In the data, the number of weeks between purchases is clustered between 6 and 8 weeks. I assume that all consumers expect to make their next purchase in exactly 8 weeks, which is the median interpurchase time. Additionally, prices tend to be clustered at specific values, so the transition process for prices is modeled as discrete/continuous. I model the probability of a price change for a product conditional on its price 8 weeks ago, previous prices for other products, and whether a new product was recently introduced 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 occurred. Note that there are almost 150 possible brand-size combinations, which makes the state space of prices very large. To reduce the size of the state space, the Markov process for prices is only estimated on the most popular sizes of liquids and powders. The prices of other sizes are assumed to be a function of the prices of the popular sizes.

An important part of the price process is that we observe introductory pricing for the new products. I assume consumers understand that the prices of new products will rise after their introduction, so I include a dummy variable in both the price transition logit and regression which is 1 for the first 12 weeks after the introduction of Cheer, a separate dummy variable which is 1 for the first 12 weeks after the introduction of Surf, and one for the first 12 weeks after Dash's introduction. Allowing for introductory pricing in this way will complicate the state space. To see why, consider a consumer who purchases a laundry detergent on the week of Cheer's introduction. Suppose further that this person purchases detergent every 8 weeks, and she knows exactly when she will make her future purchases. This person's next purchase will occur in 8 weeks, when the price of Cheer is still low. Her next purchase after that will occur in 16 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 . Because I assume that all households expect to make their next purchase in 8 weeks, 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 afterward.

For the state variable J_{it} , I estimate the probability of each detergent being available in a given calendar week for a given store separately using a binary logit. As was the case with prices, the process for availability is only estimated for the most popular sizes of each product, and so the only part of J_{it} that

is a true state variable are the availabilities of these products. This means I estimate 30 logits, one for each product, where one of the regressors is whether the product was available 8 weeks ago. The availabilities of less popular sizes are assumed to be a function of the availability of the popular sizes. I assume that the introductions of new products are a surprise to consumers, so this aspect of the state space is not taken into account by my availability estimation. A result of this assumption is that consumers will recalculate their value functions after each new product introduction: there will be a value function for after the Cheer introduction, a new one after the Surf introduction, and another one after the Dash introduction. Hence, there will be three times where it will be possible for n_t to be equal to 1, right after the introduction of each new product.

I treat consumer expectations about future coupons, which are the c_{ijt} 's, differently than future prices. I specify a process for the distribution of coupons and estimate the parameters of this process along with the other model parameters. I assume that the future c_{ijt} 's are composed of two random variables: a binary random variable \bar{c}_{ijt} which is 1 if consumer i receives a coupon for product j in purchase t , and a random variable v_{ijt} which is the value of the coupon received. Denote probability of a consumer having a coupon on hand and available for use for product j as p_{cj} . 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 period¹⁵ by tabulating the number of redeemed coupons of that value for that brand in that period, and dividing by the total number of redeemed coupons for that product in that period.

The last part of the state space is the process on the state variables summarizing purchase history, S_{ijt-1} and y_{ijt-1} . Because these state variables are influenced by consumer choices, it is instructive to examine how we compute the value functions as these parts of the state space change. Suppose first that $S_{ijt-1} = 0$ for some product j . If the consumer decides to purchase product j for the first time, then S_{ijt} will be zero and y_{ijt} will be 1. When we construct the next period value function we will integrate out the consumer's true taste for product j , conditional on γ_{ij}^0 and σ_{ij}^2 . Let γ be a random variable with the distribution of true tastes for product j , where $f(\gamma|\gamma_{ij}^0, \sigma_{ij}^2)$ is $N(\gamma_{ij}^0, \sigma_{ij}^2)$, and denote $\theta_i(\gamma)$ as the vector of individual level parameters for consumer i with her true taste draw for product j replaced by γ . Denote $v_{ikst+1}(\gamma)$ as consumer

¹⁵There 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. In an earlier draft of the paper I allowed the p_{cj} 's to be functions of n_t . The earlier version of the paper grouped more products into the Other category, and when I expanded the product space I found it was difficult to identify separate p_{cj} 's for different time periods.

i 's utility for size s of product k in purchase event $t + 1$ as a function of γ , minus the logit error ε_{ijst+1} :

$$\begin{aligned} \text{Product } k = j: v_{ikst+1}(\gamma) &= \gamma + \xi_{is} - \alpha_i(p_{ikst+1} - \alpha_{ic}c_{ikst+1}) + \eta_i y_{ikt} \\ &+ \delta EV(\Sigma_{it+2}; \theta_i(\gamma), \theta) \end{aligned}$$

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

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

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

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

3.3 Model identification

The identification argument can be summarized in two steps. First, one considers the period after most or all of the learning has occurred. During this period, it is possible to identify the distribution of consumer heterogeneity, and the distribution of consumer switching costs or variety-seeking. Once these distributions are known, the parameters which quantify learning will be identified from ways in which consumer behavior in the first few periods deviates from a model with no learning. 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 (recall that the average number of purchases that is observed for each consumer is 22). Although the estimation procedure I am using is likelihood-based, for brevity I will discuss it in the context of method of moments estimation. Thus, I will consider which moments in the data will be necessary to solve for the model's parameters. This is sufficient to show that the likelihood-based estimates are identified since the likelihood-based estimator, if it is correctly specified, is consistent and will converge to the same value as the method of moments estimator.

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

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

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 consumer uncertainty, σ_{ij}^2 , and consumers' expected tastes for the new products, γ_{ij}^0 . In my model I allow these parameters to vary across the population, but to get a feel for identification it is easier to start with the case where there is no heterogeneity. Hence, for the next few paragraphs I will drop the i subscript. First, consider the identification of σ_j^2 . This parameter will be identified from the change over time in the likelihood a consumer repurchases the new product. In particular, it can be shown 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 $\sigma^2 > 0$.¹⁶ This is because initially the consumers whose previous purchase was the new product consist mostly of consumers who are experimenting; later it consists mostly of consumers who like the new product. The share of consumers who repurchase the new

¹⁶See Osborne (2006) for detailed numerical experiments which verify this finding.

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 . Note that although the η_i parameter also impacts the likelihood of repurchasing a new product, η_i is constant over time so it will not cause the repurchase rate to change over time. Thus, learning and switching costs are separately identified.¹⁷

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

Last I consider identification of unobserved heterogeneity in σ_j^2 . Unobserved heterogeneity in σ_j^2 can be identified if I assume that it is the same across j 's, and I observe multiple new product introductions. In particular, assume that σ_j^2 is specified as

$$\sigma_j^2 = 5 \frac{\exp(\sigma_{0j} + \sigma'_{1j} D_i + \varepsilon_{i,\sigma})}{1 + \exp(\sigma_{0j} + \sigma'_{1j} D_i + \varepsilon_{i,\sigma})}, \quad (13)$$

¹⁷Osborne (2006) presents an alternative identification strategy which exploits the difference between 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 can be shown to be an increasing function of σ_j^2 . The intuition behind this finding is that 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.

where $\varepsilon_{i,\sigma}$ is a mean zero error term. We want to estimate the variance of $\varepsilon_{i,\sigma}$, which I denote σ_σ^2 . An argument for the identification of this variance can be made by extending the argument for the identification of σ_j^2 in the case where there is no heterogeneity. Consider two groups of consumers: those who purchased Cheer prior to the introduction of Surf, and those who did not. We can estimate a value of σ_{Surf}^2 for each of these groups of consumers from the change over time in the repurchase share within each group. If σ_σ^2 is zero, then the two estimates of σ_{Surf}^2 will be the same, because there is no heterogeneity in σ_{Surf}^2 . However, if σ_σ^2 is positive, then all else equal those consumers who purchased Cheer will have higher values of σ_{Cheer}^2 than those who did not. Because σ_{Surf}^2 is positively correlated with σ_{Cheer}^2 through $\varepsilon_{i,\sigma}$, the estimate of σ_{Surf}^2 will be higher for the consumers who bought Cheer than for the consumers who did not. The difference between these two estimates will rise the higher is σ_σ^2 .

The argument presented above for separating learning from switching costs is not dependent on the learning being one shot learning or the switching costs occurring in only one period. Other learning papers have allowed the learning process to be a multi period Bayesian learning process. However, even in these learning models, consumers eventually learn their taste for the new products. One simply needs a long enough time series to allow the impact of the learning to decrease. One could also allow the switching cost to last for more than one period: it could be allowed to build over time. The identification argument outlined above would remain unchanged, except that after the learning had died one could use exogenous variables from periods prior to the previous purchase to identify the switching cost. Thus, observing more time periods after the learning has died out will aid in identifying switching costs that build over time.

4 Estimation procedure

The model is estimated on a subsample of 550 households. The original data set contains 1,693 households in total. Roughly 600 of the households were dropped due to data issues, and a 50 percent random subsample of the remaining households was kept for estimation. The random sampling was done to ease computational burden. The model takes roughly one week to run due to its complexity. I note that my sample is similar in size to that used in other dynamics papers that estimate models similar to this one: for example, Erdem and Keane (1996) includes 167 households, and Hendel and Nevo (2006) includes 221. Details on the construction of the data set are discussed in the Technical Appendix, Section 1.

4.1 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 the full model, I

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

I assume that for each purchase event every coupon c_{ijt} for a non-purchased product (one for which $y_{ijt} = 0$) received by the consumer is drawn from the same distribution as consumer expectations about future coupons that is described in Section 3.2; hence, consumer expectations about future coupons are rational. To summarize the notation developed in that section, recall that the c_{ijt} for a non-purchased product is composed of two random variables, the binary random variable \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} . Note that because these coupon parameters must be between zero and one, they are transformations of underlying parameters. Each p_{cj} is equal to $\exp(\tilde{p}_{cj})/(1 + \exp(\tilde{p}_{cj}))$, where \tilde{p}_{cj} lies in the real line.

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 et al. (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.

4.2 Estimation algorithm

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.¹⁸

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.¹⁹ My underlying demand model is the random coefficients logit model, with two differences: the coupon parameters and the value function solution. Thus, the setup for my Gibbs sampler is very similar to that used to estimate the random coefficients logit model. This estimator is well understood and is described in Train (2003), pp. 307–308.

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 and the \tilde{p}_{cj} 's, a normal prior on b which I denote $k(b)$, and inverse gamma priors on the elements of the diagonal matrix W , which I denote as $IG(W)$. The posterior distribution of the model parameters will depend on the parameters' prior distribution and the probability of the data given the parameters. The priors on b and W are assumed to be non-informative, so that $k(b)$ has zero mean and infinite variance. The prior on W is also chosen to be non-informative, so that the scale is set to 1 and the degrees of freedom approaches 1. The posterior distribution of the model parameters will depend on the parameters' prior distribution and the probability of the data given the parameters.

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

¹⁸Bayesian techniques have become widely used in the estimation of discrete choice demand models. For an overview, please refer to Rossi et al. (2005) and Train (2003).

¹⁹Norets (2009a) and Gelman and Rubin (1992) discuss how to formally assess convergence in MCMC algorithms. This involves running the Gibbs sampler from different random starting points and testing when the posterior distributions calculated from the converged sequences are equal across runs. My model takes roughly one week of computational time to finish, so it was not feasible for me to run it many times. My approach, which is less formal, is to run the sampler for a long time, and to observe when the time series of draws appears to be traversing the posterior.

flow utility minus the logit error. The probability of the consumer’s choice in purchase event t will be

$$Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) = \sum_{(j,s) \in J_{it}} d_{ijst} \frac{\exp(v_{ijst} + \delta EV(\Sigma_{it+1}; \theta_i, \theta))}{\sum_{(k,l) \in J_{it}} \exp(v_{iklt} + \delta EV(\Sigma_{it+1}; \theta_i, \theta))}. \tag{14}$$

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

$$\Lambda(\theta_i \forall i, b, W, c_{it} \forall i \text{ and } t, \theta) \propto \prod_{i=1}^I \left[\prod_{t=1}^{T_i} \{Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) Pr(c_{it}|\theta)\} g(\theta_i|b, W) \right] \cdot k(b) IG(W) \tag{15}$$

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

The first block draws θ_i for each household conditional on the d_{it} ’s, the c_{it} ’s, b and W . Because of the assumption that the error term is logit, the conditional posterior likelihood of a particular vector of θ_i is proportional to $\prod_{t=1}^{T_i} \{Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it})\} g(\theta_i|b, W)$. This distribution is not conjugate, which means that the Metropolis-Hastings algorithm (see Technical Appendix, Section 2 for the steps I use to implement this) must be used in this step.²⁰

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

$$\prod_{i=1}^I \prod_{t=1}^{T_i} \{Pr(d_{it}|\theta_i, \theta, \Sigma_{it}, c_{it}, x_{it}) Pr(c_{it}|\theta)\}. \tag{16}$$

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

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

Because one draws parameters out of conditional posteriors, initial parameter draws may be influenced by the starting points. To avoid this, it is common

²⁰Note that when we perform this step, we will need to evaluate the consumer’s expected value function in Eq. 14, $EV(\Sigma_{it+1}; \theta_i, \theta)$. The procedure I use to do this is described in Section 4.3.

practice to drop the initial draws. I run the estimator for 20,000 iterations, and drop the first 10,000 draws. A more detailed discussion of convergence is presented in the Technical Appendix, Section 5.

4.3 Value function solution

The method of Imai et al. (2009) and Norets (2009b) works in conjunction with the Gibbs sampler to obtain a solution of the value function. In this section I will broadly describe how I solve for the value function in Eq. 14 and how the method works. The innovation of this new method is that discrete choice dynamic programming problem is solved only once, along with the estimation of the model parameters.

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

Consider the first step, which is to draw some values of the model's unobservables and calculate expected utility. This calculation is done at points in the state space, $\Sigma = (s, p, J, y, n)$, and the expected utilities and current parameter value are retained. There are two different sets of unobservables which are unobserved to the consumer at the time she makes her purchase decision, and must be integrated out when the value function is formed: the ε_{ijst} 's, and the consumer's future tastes for products she has not yet purchased, the γ_{ij} 's. Integrating out the ε_{ijst} 's does not require numerical approximation: because of the assumption that they are logit errors, the consumer's expected utility has a closed form solution, conditional on θ_i , θ , and future coupons. This is not true when we integrate out the future γ_{ij} 's and c_{ijt} 's, so these must be approximated numerically. As an example, let us consider constructing an analogue to the consumer's expected value function in Eq. 12, which is the value at state space point $s_j = 0, y_j = 1$ for some new product j . First I draw $L = 10$ draws from the true taste distribution for product j , which is $N(\gamma_{ij}^0, \sigma_{ij}^2)$, and from the coupon

distribution implied by θ . To calculate the expected utility, we need to calculate first each consumer’s exact utility (ignoring the logit error) at each product at simulation l . Denote the l th taste draw as γ_{ij}^l and the l th 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 $E_{(p',J')|(p,J)} V_n(s', p', J', y', n'; \theta_i^l, \theta)$.²¹ Then the consumer’s utility for product j at simulation l , v_{ij}^l , will be

$$\begin{aligned}
 \text{Product } k = j : v_{iks}^l &= \gamma_{ik}^l + \xi_{is} - \alpha_i(p_{ks} - \alpha_{ic}c_{ik}^l) + \eta_i y_k \\
 &\quad + \delta E_{(p',J')|(p,J)} V_n(S', p', J', y', n'; \theta_i^l, \theta) \\
 \text{Product } k \neq j : v_{iks}^l &= \gamma_{ik}(S_k) + \xi_{is} - \alpha_i(p_{ks} - \alpha_{ic}c_{ik}^l) + \eta_i y_k \\
 &\quad + \delta E_{(p',J')|(p,J)} V_n(S', p', J', y', n'; \theta_i^l, \theta),
 \end{aligned} \tag{17}$$

which corresponds to Eq. 11.

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

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

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 $\bar{\theta}_{i,g}$. 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

$$\begin{aligned}
 &E_{(p',J')|(p,J)} V_g(s, p, J, y, n, \theta_{i,g}, \theta_g) \\
 &= \frac{\sum_{r=1}^{N(g)} \left[\hat{E}V_r(s, p, J, y, n; \theta_{i,r}, \theta_g) \right] k((\bar{\theta}_{i,g} - \bar{\theta}_{i,r}^r)/h_k)}{\sum_{i=1}^{N(g)} k((\bar{\theta}_{i,g} - \bar{\theta}_{i,r})/h_k)}, \tag{19}
 \end{aligned}$$

²¹Since the state space is quite large, and computer memory is limited, I only evaluate the value function at a subset of the state space points, and interpolate it everywhere else. The details of this procedure, as well as other computational details associated with the value function solution, are described in the Technical Appendix, Section 4 (http://sites.google.com/site/matthewosborne/home/papers/technical_appendix.pdf?attredirects=0).

where $k(\cdot)$ is a kernel density function and h_k is a bandwidth parameter, and $\hat{E}V_r(s, p, J, y, n; \theta_{i,r}, \theta)$ is the r th retained expected utility.²² The approximated value function is used to calculate the utilities in Eq. 17.²³

5 Estimation results

Estimates of selected model parameters are shown in Table 4.²⁴ Recall that 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 columns of estimates under the heading "Mean" shows the means and standard deviations (shown in parentheses) of the simulated posterior density for each element of b ; similarly, the columns of estimates under the heading "Variance" shows the mean and standard deviation of the simulated posterior for W .²⁵ Estimates of parameters that are fixed across the population, θ , are shown under the "Mean" heading; the corresponding entries under the "Variance" heading are dashed for these parameters. Although the numbers in the table are means and standard deviations of parameter posterior densities, they can be interpreted in the same way as the estimated coefficients and standard errors that are produced by classical methods.

The first 8 rows show the estimated tastes for some of the established products.²⁶ The All Powder product is normalized to 0, so the taste parameters capture a consumer's utility relative to this product. Mean taste parameters for popular brands such as Powder or Liquid Tide are positive. Furthermore,

²²Norets (2009b) proposes constructing the expected value function using a nearest neighbor approximation rather than a kernel weighted average, showing that this method has better convergence properties, and may offer computational advantages as well. In this paper, I use a Gaussian kernel function with a reasonably small bandwidth parameter, which should behave similarly to the nearest neighbor approximation: the Gaussian kernel assigns zero weight to parameter draws that are too far away from the one where the approximation is occurring.

²³Norets (2007) proposes an alternative procedure for estimating dynamic discrete choice models using MCMC, which involves solving the value function on a fixed grid of points prior to estimation, and then fitting a feedforward neural network function to the grid. The fitted values of the function are then used during the Gibbs sampling procedure to approximate the expected value function. This new procedure shows much promise and should be a fruitful area for future research.

²⁴Estimates of parameters that are not shown in the table, such as the weekly dummy variables or coupon parameters, are presented in the Technical Appendix, Section 6.

²⁵The estimated standard deviation can be affected by autocorrelation between the draws of Gibbs sampler. To reduce this, in the estimates that I present and the counterfactuals below, I use every tenth draw. Some evidence that this is sufficient is presented in the Technical Appendix, Section 5.

²⁶Recall that the model includes 30 different brands. Taste parameter estimates for the other 22 products are not shown for brevity, but are available in the Technical Appendix, Section 6.

Table 4 Selected parameter estimates of b , θ , and W (utility function)

Coefficient	Mean, b , θ	Std. dev.	Variance, W	Std. dev.
Cheer powder	-0.636	(0.214)	4.515	(0.781)
Oxydol powder	-0.207	(0.172)	4.357	(0.64)
Tide powder	1.141	(0.152)	2.983	(0.42)
B-3 liquid	1.237	(0.332)	2.256	(0.893)
Dynamo liquid	0.216	(0.175)	2.329	(0.351)
Era liquid	1.198	(0.173)	2.751	(0.546)
Tide liquid	1.534	(0.134)	2.931	(0.42)
Wisk liquid	0.358	(0.13)	3.754	(0.48)
Price, intercept	-1.284	(0.092)	0.793	(0.086)
Price, $\ln(\text{size})$	-0.053	(0.067)	-	
Price, $\ln(\text{income})$	-0.235	(0.09)	-	
Coupon	5.024	(0.43)	-	
Display	1.025	(0.081)	0.836	(0.154)
Feature	0.87	(0.07)	0.596	(0.111)
Cheer (γ_{ij}^0)	1.698	(0.165)	1.408	(0.453)
Surf (γ_{ij}^0)	1.783	(0.096)	0.371	(0.177)
Dash (γ_{ij}^0)	0.651	(0.105)	0.685	(0.2)
σ_{0i}^2 (intercept)	-2.147	(0.198)	0.563	(0.223)
σ_{Cheer}^2 , $\ln(\text{size})$	-0.279	(0.19)	-	
σ_{Cheer}^2 , $\ln(\text{income})$	-0.801	(0.264)	-	
$\sigma_{0,\text{Surf}}^2$ (intercept)	1.48	(0.222)	-	
σ_{Surf}^2 , $\ln(\text{size})$	-0.543	(0.153)	-	
σ_{Surf}^2 , $\ln(\text{income})$	-0.84	(0.431)	-	
$\sigma_{0,\text{Dash}}^2$ (intercept)	1.216	(0.208)	-	
σ_{Dash}^2 , $\ln(\text{size})$	-0.644	(0.11)	-	
σ_{Dash}^2 , $\ln(\text{income})$	1.306	(0.394)	-	
Switching cost, intercept	1.596	(0.207)	2.276	(0.266)
Switching cost, $\ln(\text{size})$	-0.18	(0.098)	-	
Switching cost, $\ln(\text{income})$	0.02	(0.177)	-	

This table shows some of the estimated parameters of the consumer flow utility (Section 3.1) for the main model specification. In most parameters I allow normally-distributed heterogeneity across the population, and so I have estimated the population mean of the coefficient (b) and the variance (W). Because my model estimation procedure is Bayesian, the numbers in this table show statistics from the simulated posterior distribution of each parameter. The columns under the heading “Mean” show the posterior means of the mean parameters, and the columns under the heading “Variance” show the variances. The standard deviations of the posteriors are shown in parentheses. Some utility coefficients, such as the price coefficient and the consumer uncertainty (see Eqs. 6 and 5), are transformations of the parameters in the table. For some utility coefficients the population variance was restricted to be 0. These parameters are shown with dashes. The utility specification also includes brand intercepts for 22 other products, dummy variables for different sizes, weekly dummy variables, and coupon probabilities. Estimated distributions of these coefficients are in the paper’s Technical Appendix, Section 6, which is available on request from the author

most of the mean taste parameters are statistically significant, except for some smaller brands such as Oxydol or Dynamo. The variances are quite large as well, ranging from about 2.2 to about 4.5, indicating a significant amount of unobserved heterogeneity in tastes. Consumer heterogeneity in tastes is very important in this market, which is consistent with these products being

experience goods. It is also consistent with important heterogeneity in factors such as the types of fabrics in a household's wardrobe, the types of soils and stains that need to be cleaned, the water temperature used, the household's washing machine quality, and the types of scents the household prefers.

The next six rows show consumer responses to prices, coupons, feature and display variables. The parameter for consumer price sensitivities is constructed in the same way as for the learning parameters (Eq. 6). The population average value of the price coefficient is about -104 , with a standard deviation of 26 , indicating that there is substantial heterogeneity in price sensitivity.²⁷ The parameter on household income is negative and significant, which is intuitive. The parameter for household size, however, is not precisely estimated. The estimate of the coupon sensitivity parameter is about 5 . Recall that the coupon sensitivity coefficient that enters the consumer flow utility, α_{ic} , is a transformation of α_{0ic} , $\frac{\exp(\alpha_{0ic})}{1+\exp(\alpha_{0ic})}$ (Eq. 7). This means that the coupon sensitivity is very close to one. The feature and display variables are both positive and precisely estimated.

5.1 Estimated consumer uncertainty and switching costs

I now turn to the estimates of the parameters that are of primary interest in this paper, those pertaining to learning and switching costs. The learning parameters are shown in the third block of parameters in Table 4. Recall that for each new product, consumers have a prediction of their taste for it, γ_{ij}^0 , and uncertainty about their taste, σ_{ij}^2 . The first three rows of this block of coefficients shows the estimated distribution of predicted tastes for each new product. On average, consumers expect to like Surf the best, then Cheer, and then Dash the least. The population variances of these parameters are large, especially for Cheer and Dash, indicating that there is a significant amount of variation in consumers' prior beliefs about these products.

The next lines show estimates of the parameters pertaining to consumer uncertainty. Recall that I model both observed and unobserved heterogeneity in learning, where the unobserved heterogeneity is persistent within a consumer, across products (see Eq. 13). For simplicity of presentation, when I estimate the model I use the following specification for consumer uncertainty

$$\sigma_j^2 = 5 \frac{\exp(\sigma_{0i} + \sigma_{0,\text{Surf}}\mathbf{1}\{j = \text{Surf}\} + \sigma_{0,\text{Dash}}\mathbf{1}\{j = \text{Dash}\} + \sigma'_{1j}D_i)}{1 + \exp(\sigma_{0i} + \sigma_{0,\text{Surf}}\mathbf{1}\{j = \text{Surf}\} + \sigma_{0,\text{Dash}}\mathbf{1}\{j = \text{Dash}\} + \sigma'_{1j}D_i)}, \quad (20)$$

²⁷Recall that prices are measure in dollars per ounce. This means that a product's price is on the order of 0.1 to 0.01.

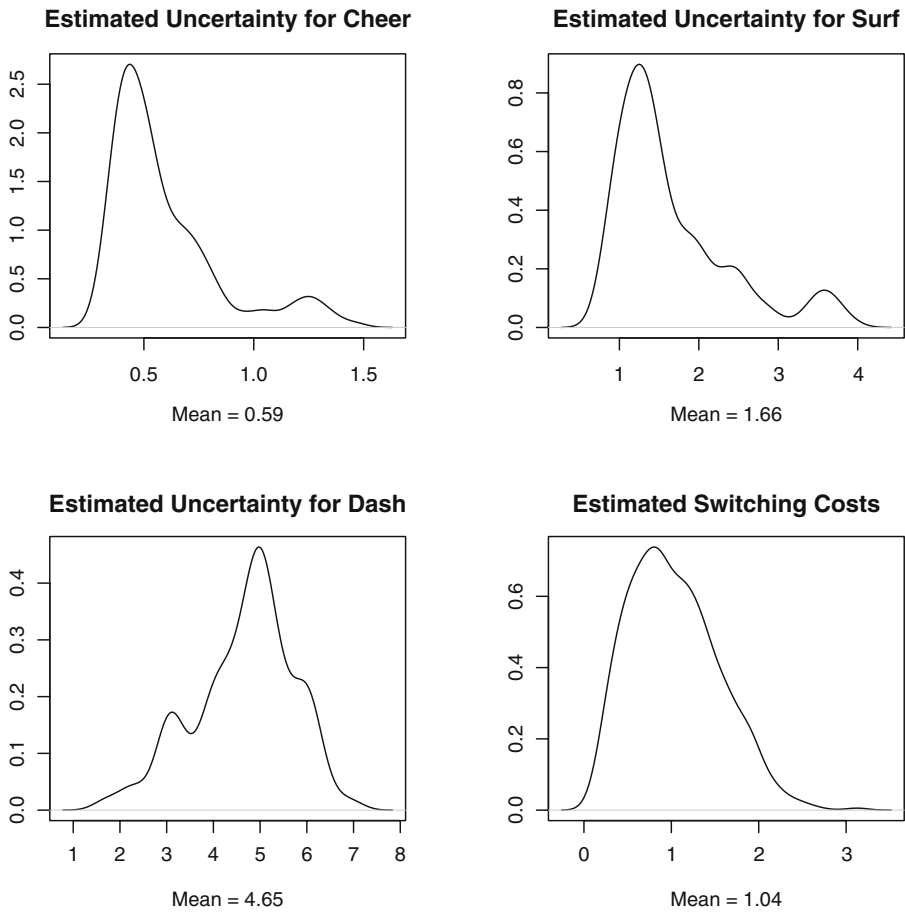


Fig. 2 Estimated population distributions of learning and switching cost parameters

which is equivalent to that of Eq. 13. σ_{0i} represents the intercept for Cheer; the intercept for Surf is $\sigma_{0i} + \sigma_{0, \text{Surf}}$, and similarly for Dash.²⁸ The variance of the intercept term, σ_{0i} , is large, indicating there is unobserved heterogeneity in learning that is not being captured by income or household size. The impact of household size on learning is always negative, indicating larger households are more certain about their true tastes for new products than smaller ones, although this effect is not precisely estimated for Cheer. The impact of household income is also negative for Cheer and Surf, but not for Dash.

To see what the population distribution of consumer uncertainty looks like, in the two top panels and the bottom-left panel of Fig. 2, I show kernel density

²⁸This specification is equivalent to setting $\sigma_{0i} = \sigma_{0, \text{Cheer}} + \varepsilon_{i, \sigma}$ and redefining $\sigma_{0, \text{Surf}}$ and $\sigma_{0, \text{Dash}}$ to be the differences between the Cheer and Surf and Cheer and Dash dummy variables in Eq. 13.

plots of the estimated σ_{ij} parameters.²⁹ Consumers are more uncertain of their true tastes for Surf than for Cheer, and are most uncertain for Dash. In all the products, there is a significant amount of heterogeneity in learning across the population.

The variance in true tastes for these products will be a function of both the variance in the prior, and the mean and variance of the learning parameter. If the amount of learning is small, most of the variance in true tastes will be comprised of the variance in priors. To see this, notice that the true taste $\gamma_{ij} = \gamma_{ij}^0 + \sigma_{ij}z_{ij}$, where z_{ij} is a standard normal random variable that is independent of both γ_{ij}^0 and σ_{ij} .³⁰ Using the decomposition of variance formula, the variance of the true taste can be written as

$$\begin{aligned} V(\gamma_{ij}) &= V(\gamma_{ij}^0) + E_{z_{ij}}(V(\sigma_{ij})z_{ij}^2) + V_{z_{ij}}(E(\sigma_{ij})z_{ij}) \\ &= V(\gamma_{ij}^0) + V(\sigma_{ij}) + E(\sigma_{ij})^2. \end{aligned}$$

The percentage of the variance in true tastes that is due to learning is therefore $(V(\sigma_{ij}) + E(\sigma_{ij})^2)/V(\gamma_{ij})$. For Cheer, this percentage is about 30%; for Surf, it is about 82%, and for Dash, it is about 87%. For all three products, the amount of learning is significant when measured in this way.

The estimated parameters related to switching costs are shown in the last three rows of the table. The intercept is large and positive, suggesting that most consumers have high switching costs. Larger households incur smaller switching costs, while the impact of household income on switching costs appears negligible. The bottom right panel of Fig. 2 shows the estimated population distribution of switching costs. Most households display switching costs, although a small proportion—roughly 8%—are estimated to be variety-seeking (negative switching costs). An estimate of the disutility of switching costs to the consumer can be made by dividing the switching cost parameter by the price coefficient. This average of this ratio is about 2.1 cents per ounce. The modal package size for liquid detergents is 64 ounces, which suggests the disutility from switching costs are about \$1.35, roughly 30 percent of the price of a bottle of detergent.

I present another exercise which quantifies the importance of learning in Table 5. First, consider the first three rows of the table. The first column shows the predicted market shares at the estimated parameter values. To compute these shares I simulate consumer purchases under the estimated taste

²⁹When I compute the population distribution of σ_{ij} , 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, $\sigma_{ij,g}$, using $\theta_{i,g}$, θ_g , and demographics for i (Eq. 5). Then I average these $\sigma_{ij,g}$'s across draws, obtaining an estimated σ_{ij} for each household. These estimated σ_{ij} 's are shown in the plots.

³⁰This independence is a result of rational expectations—the consumer knows both γ_{ij}^0 and σ_{ij} , and therefore they cannot signal anything about the value of z_{ij} .

Table 5 Effect of removing learning on new product market share (full model)

Product	Predicted market share, learning	Predicted market share, no learning	% Change
Entire period			
Cheer	3	3.23	7.76
Surf	8.19	8.27	1.65
Dash	1.43	1.69	19.44
1st 12 weeks after intro			
Cheer	10.89	11.58	6.22
Surf	18.07	18.29	1.44
Dash	6.16	7.04	14.51

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

distributions.³¹ The next column shows the result of the simulation after setting consumer priors about the new products equal to their true tastes for the new products.³² The third column shows the percentage change in the product's market shares. The results show that removing learning increases the overall market shares of the new products. The absolute increase in share is small, but the percentage increase can be large.

Changing the amount of learning in this way also can have a different impact on the short run market shares of a new product as opposed to the long run shares due to the option value of learning. The simulated increase in share due to removing learning is shown in the bottom three rows of the table. Although the short run absolute increase in share is larger than the long run increase, it is smaller in percentage terms. To see why the long and short run response to learning can be different in Fig. 3 I plot 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 1.7 and variance of 1.4 (Table 4). The flatter one is the population distribution of true tastes for Cheer, which are

³¹For each consumer, I compute the average of her individual level posterior draws, and then simulate her purchases. I simulate each consumer's purchases ten times to reduce variance due to simulation error. Note that an alternative way to compute this counterfactual would be to simulate shares at each posterior draw, and then to take the average over these draws. When I experimented with the latter approach, I found it to be much slower computationally, but it provided very similar results.

³²Note that when this simulation is done, all random draws on variables such as the error term or coupons are set to be the same as the previous simulation, so that the increase in share is not contaminated by simulation error.

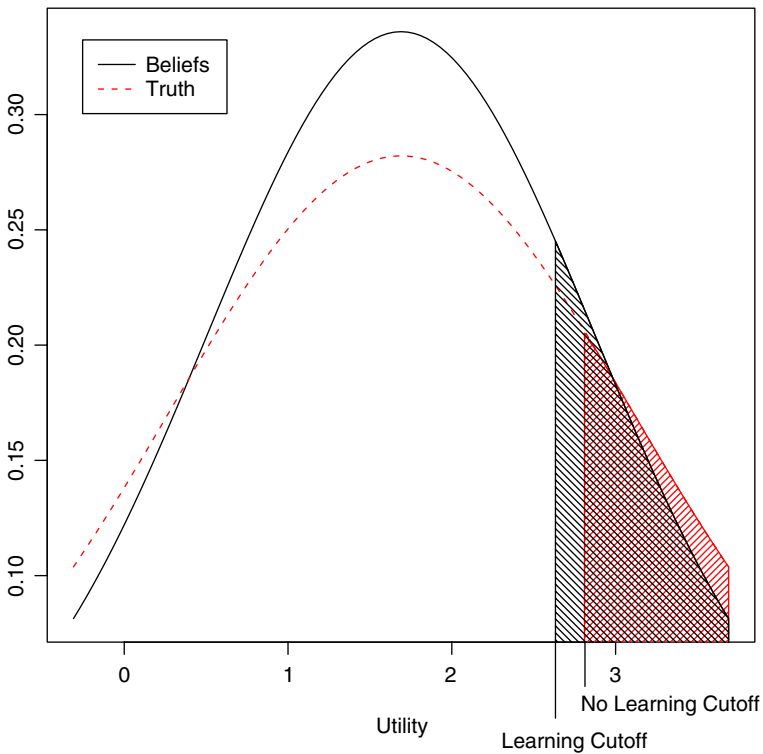


Fig. 3 Estimated taste distributions for cheer, before and after learning

tastes after learning has occurred. This distribution is normal, and has mean of 1.7, and a variance of 2, which is the variance specified in Eq. 21.

When there is no learning, a consumer will make a first purchase of Cheer when her taste draw is such that her utility from purchasing Cheer is greater than any other product. In the figure, the line labeled “No Learning Cutoff” shows the cutoff for a consumer with average values of tastes for all products, assuming no switching costs, prices for all products are the same (and consumers expect them to stay the same), and the error terms are set to zero.³³ The share of consumers who will choose Cheer will be those whose taste draw is to the right of this line. Under full information, this share will be the red shaded area, which is the area under the dotted true taste curve to the right of the “No Learning Cutoff” line.

Figuring out the share of consumers who purchase the new product when there is learning is more complicated, because consumers are forward-looking and there is an option value of learning. All else equal, a higher option value

³³Although the results in the table include switching costs, which are not incorporated into this stylized example, if I run the counterfactual with switching costs set to zero I get qualitatively similar findings.

of learning will result in more experimentation. The reason for this is that the cutoff in the presence of learning is equal to the no learning cutoff plus the option value of learning—it is always to the left of the no learning cutoff. This new cutoff is shown in the figure with the line labeled “Learning Cutoff”. The mass of consumers who experiment with the new product will be the area under the expected tastes distribution, which is shaded in black on the figure. If this black shaded area is smaller than the red shaded area, then informing consumers of their true match values will cause an increase in the product’s short run market share, even when accounting for the future value of learning. This is why removing learning increases the short run shares.

In the intermediate run, the effect of giving consumers their true taste draws can 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 similar to that of consumers do not learn, which means that the mass of consumers who purchase the new product when there is learning will be approximately those consumers to the left of the “No Learning Cutoff” line under the expected taste distribution. When consumers are informed of their true tastes, the increase will be the difference between this probability and the mass of consumers to the left of the line under the true taste distribution. This increase is greater than the difference between the mass of consumers who experiment when there is value to learning, and the mass of consumers who purchase when they are informed of their true taste. This explains why the percentage increase in long run share is larger than that of the short run share.

5.2 Managerial implications

This section examines two questions: what is the impact of introductory price promotions and introductory advertising on product revenues and market shares in my model of learning and switching costs? Second, if managers were to ignore the presence of switching costs or learning, how would their predictions about the impact of these marketing variables differ? In order to answer the second question, I estimated two restricted models: one where switching costs were set to zero, and one where it was assumed that consumers knew their tastes for all the new products beforehand.³⁴ For brevity, parameter estimates for these restricted models are not shown in the body of the paper, however, they are shown in the Technical Appendix, Section 6. There are a few qualitative differences between the estimates which are important to discuss.

³⁴In the model of switching costs only, it is assumed that each consumer’s expected taste is equal to their actual taste; that is, $\gamma_{ij}^0 = \gamma_{ij}$, and $\sigma_{ij}^2 = 0$. In the model of learning only, it is assumed that $\eta_i = 0$ for all consumers.

The estimated taste coefficients and coefficients on switching costs look broadly similar for the full model and the switching costs only. This is consistent with the argument made in Section 3.3 regarding the identification of these parameters. I argued that consumer taste distributions and switching cost coefficients will be identified from longer run behavior. In the no learning model, the first few periods after the new product introduction will have some impact on the estimates, but most of the model's identification will come from longer run behavior since we observe consumers for a long period of time after the product introductions. Although the estimates of these model parameters are similar, these two models make very different predictions about the impact of introductory price discounts and long term price elasticities, which will be discussed in Section 5.3.

In contrast to the switching costs only model, the model with only learning produces very different parameter estimates than the full model. In particular, the taste variances produced by the model are larger, often by a factor of two or three. The reason this result is produced is that when switching costs are not modeled, all observed tenure in brand choice is attributed to taste heterogeneity.

We now turn to the first counterfactual, which measures the impact of a temporary price cut on current and future market shares. Estimates of this impact, expressed in terms of elasticities, are shown in Table 6. To compute this elasticity, I hold prices fixed over time at their average values (outside the period in which the promotion occurs) and assume that no features, displays or couponing occurs. I then solve for each consumer's value function, assuming beliefs are that prices will be fixed at their average values forever. I then simulate purchases for ten periods, where each consumer makes a single purchase every period. Since the modal interpurchase time is 8 weeks, each period can be thought of occurring approximately every two months.

The first row and first column of the table shows the impact of a one period price cut that occurs in period 1 on Cheer's period one market share.³⁵ Note that this is a partial equilibrium counterfactual: I assume that competitor prices do not respond to the temporary price cut. There is a large response in the period the price cut occurs: a 1% price cut generates a 6% increase in the quantity of Cheer sold. The next row shows how this price cut impacts period 2 quantity. The impact on period 2 quantity is still substantial: a 1% percent price cut generates a 0.84% increase in second period quantity. The intertemporal impact dies out quickly, however: after five or six periods, quantity sold is very close to that with no price cut. The next two columns of the table show the impact of the price cut if it occurs in period 3 or period 5, respectively. When I compute this counterfactual, I simulate purchases for the periods prior to the price cut, and I assume that the price cut occurs unexpectedly.

³⁵When I simulate the response to the price cut, I assume that all package sizes of Cheer are available. In the simulation, I cut the price of all package sizes by 20%.

Table 6 Counterfactual: impact of price cuts in periods 1, 3 and 5

Model	Full model			Switching costs only			Learning only		
	1	3	5	1	3	5	1	3	5
Cheer									
1	-6.7			-3.74			-4.35		
2	-0.84			-0.3			-0.09		
3	-0.34	-5.34		-0.11	-4.11		-0.06	-4.47	
4	-0.18	-0.56		-0.06	-0.48		-0.04	-0.03	
5	-0.12	-0.24	-4.9	-0.04	-0.2	-4.18	-0.04	-0.02	-4.41
6	-0.08	-0.12	-0.5	-0.04	-0.12	-0.46	-0.04	-0.02	0
7	-0.05	-0.05	-0.16	-0.02	-0.1	-0.09	-0.04	-0.02	0
8	-0.04	-0.05	-0.08	0	-0.09	-0.08	-0.03	-0.01	0
9	-0.03	-0.02	-0.04	0	-0.05	-0.05	-0.03	-0.01	0
10	-0.03	-0.01	-0.04	0	-0.04	-0.05	-0.03	0	0
Surf									
1	-7.16			-4.24			-6.38		
2	-0.63			-0.49			-0.09		
3	-0.17	-6.15		-0.13	-4.77		-0.08	-6.75	
4	-0.08	-0.68		-0.05	-0.54		-0.07	-0.04	
5	-0.06	-0.24	-5.79	-0.04	-0.19	-4.9	-0.07	-0.04	-6.67
6	-0.03	-0.14	-0.56	-0.03	-0.08	-0.54	-0.06	-0.03	-0.04
7	-0.03	-0.06	-0.19	-0.03	-0.03	-0.2	-0.05	-0.03	-0.03
8	-0.04	-0.03	-0.08	0	-0.02	-0.12	-0.04	-0.03	-0.02
9	-0.04	-0.03	-0.05	0	-0.02	-0.08	-0.02	-0.02	-0.02
10	-0.03	-0.03	-0.05	0	-0.02	-0.04	-0.02	-0.01	-0.02
Dash									
1	-6			-5.81			-6.42		
2	-0.83			-0.79			-0.15		
3	-0.43	-5.84		-0.35	-5.7		-0.14	-6.61	
4	-0.25	-0.82		-0.26	-0.85		-0.13	-0.06	
5	-0.17	-0.39	-5.69	-0.15	-0.32	-5.8	-0.12	-0.06	-6.59
6	-0.1	-0.2	-0.64	-0.1	-0.2	-0.65	-0.11	-0.06	-0.04
7	-0.11	-0.15	-0.21	-0.08	-0.12	-0.3	-0.11	-0.06	-0.04
8	-0.09	-0.11	-0.11	-0.05	-0.07	-0.14	-0.1	-0.06	-0.03
9	-0.07	-0.12	-0.09	-0.05	-0.05	-0.15	-0.1	-0.05	-0.03
10	-0.06	-0.1	-0.05	-0.05	-0.05	-0.07	-0.09	-0.05	-0.04

This table shows the simulated intertemporal elasticity of period *i* quantity with respect to a change in the period *j* price which lasts for a single period. The columns labeled “Period” denote the period *j* when the price cut takes place, and the row denotes the change in period *i*’s market share. The first three columns show the impact of the price cut under the estimates of the full model, while the next three show those for the model with no learning, and the last for the model with no switching costs. For example, the third row of the first column shows that a ten percent increase in the period 1 price of Cheer lowers the period 3 quantity of Cheer by 3.4%

This means that consumers have time to purchase the new product and learn about it prior to the price cut. My results indicate that the price cut becomes less effective at building both present and future market shares the later it is introduced. This is a result of the fact that consumers learn over time, and learning and switching costs reinforce each other. When consumers learn and have switching costs, some consumers who purchase the product and receive a low match value with it continue to purchase it due to the switching cost. As time goes on and the amount of learning occurring in the population decreases,

the intertemporal impact of the price cut decreases as well. The impact of the price cut in the period during which it occurs drops because the option value of learning can rise in the presence of switching costs.³⁶ The intuition behind this is that if a consumer receives a good match value for the new product and has switching costs, it is less likely she will have to switch in the future, making the option value of learning larger. This provides an alternative explanation for why we observe introductory pricing of the new products to the idea that firms recognize dynamics and are pricing with them in mind. It may be that firms are myopic and simply observe that short run demand elasticity is falling, and raise their prices in response. The rows below Cheer show the impact of a similar price cut for Surf and Dash. The results are similar for these two products.

It is interesting to contrast the bias that occurs from using model estimates which do not account for learning or switching costs. The fourth, fifth and sixth columns show the impact of the price cut under the model parameters produced by the switching costs only model. The impact of the temporary price cut on current shares is smaller when there is no learning. This occurs because when there is no learning, there is no option value of learning. The option value of learning makes the new product more attractive to consumers, and the price cut more effective at building current share. The intertemporal effect is also smaller: as noted in the previous paragraph, this is due to the fact that learning combined with switching costs can reinforce each other when temporary price cuts occur. Note that the bias from not modeling learning can be significant: for Cheer, the intertemporal elasticity is underestimated by as much as 65%. The own (short term) elasticity is underestimated by as much as 55%. Similar results are obtained for Surf; for Dash, the intertemporal elasticities are often smaller, although the differences are less stark.

The last three columns show the results for the model with no switching costs. The impact of a price cut during the period it occurs can be greater or smaller than the model with learning and switching costs. This ambiguity is likely driven by two effects which work against each other. First, taste variances are estimated to be higher in the model without switching costs, which will tend to drive own price elasticities towards zero. However, the lack of switching costs will tend to drive price elasticities away from zero, counteracting this effect. The overall impact of leaving out switching costs on the short term effect of the price cut is therefore ambiguous. In contrast, the effect on intertemporal elasticities is clear. They are significantly lower, due to the lack of switching costs. The impact of the price cut on future market shares is not zero, however, due to the presence of learning. The price cut induces some consumers to purchase the new product, and those who have good matches with it continue to purchase it for a long time. However, it appears that the effect of learning alone on intertemporal elasticities is small.

³⁶See Osborne (2006).

The second counterfactual, shown in Table 7, 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 $\sigma_{a_j}^2$. 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):

$$\gamma_{uij}^0 = \frac{\frac{\gamma_{ij}^0}{\sigma_{ij}^2} + \frac{a_{ij}}{\sigma_{a_j}^2}}{\frac{1}{\sigma_{ij}^2} + \frac{1}{\sigma_{a_j}^2}} \tag{21}$$

$$\sigma_{uij}^2 = \frac{1}{\frac{1}{\sigma_{ij}^2} + \frac{1}{\sigma_{a_j}^2}}$$

For each product, I assume that the signal variance $\sigma_{a_j}^2$ is one half of the estimated population variance. This counterfactual is simulated using the results of the full model and the model with no switching costs.

The simulated market shares are shown in the first ten rows of Table 7. Informative advertising increases the market shares of all three products. The increase is largest for Dash and smallest for Cheer, which is consistent with the estimated learning parameters being largest for Dash and smallest for Cheer (Fig. 2). Because consumers are very uncertain about Dash, which was a

Table 7 Counterfactual: effect of informative advertising

Share increase in period	Full model			Learning only		
	Cheer	Surf	Dash	Cheer	Surf	Dash
1	4.63	21.36	75.05	14.42	15.26	25.79
2	6.8	28.1	83.97	12.34	9.97	15.71
3	8.25	34.14	92.41	12.3	11.15	17.16
4	6.61	34.63	91.27	12.57	11.85	19.03
5	6.76	38.91	89.18	12.66	12.61	19.92
6	3.59	35.24	90.43	12.75	12.95	19.9
7	5.57	37.29	87.56	13.04	13.54	20.22
8	5.84	35.96	85.77	13.03	13.74	21.7
9	8.21	35.99	85.37	13.1	13.67	21.79
10	9.67	35.53	80.04	12.99	13.61	22.46
Revenue increase	52.82	331.42	446.52	98.96	122.55	96.35

This table shows the impact of informative advertising in the first period on product market shares, in percentage terms, and on product revenues, which are in dollars. Consumers are given a signal on the product’s quality which has half the variance of their uncertainty about their match value with the product. The first three columns show the simulated percentage increase in market shares for each new product, as implied by the full model, while the second show the results for the model with no switching costs. The final row, labeled “Revenue Increase”, shows the total increase in revenues from the advertising, in dollars

niche produce, informative advertising is very effective at increasing its market share. Revenues for all three products also increase.

The results of the learning only model are shown in the fourth, fifth, and sixth columns. The impact of the advertising is different, in part because this model produces different estimates of consumer uncertainty. In the no learning model, the average uncertainty for Cheer is 1.43, for Surf it is 0.44, and for Dash it is 1.59. The impact of the advertising on the time series behavior of market shares is also different, and this is due to the fact that switching costs are zero. In the model with switching costs, the impact of the informative advertising is smaller in the initial periods as opposed to later periods. This happens because, before the new product is introduced, consumers have been using one of the older products, and it takes time for them to overcome this switching cost and purchase the new product. In contrast, in the model with no switching costs, the impact of the advertising is large at first, but then it drops over time as consumers who purchase the new product find they dislike it and switch away from it.

5.3 Long term price elasticities

In Table 8 I compute the empirical price elasticities implied by the full model, the switching costs only model, and the learning only model. The price elasticities are computed in a method similar to how I computed the impacts of price cuts in the counterfactual exercises from the previous section. First, I compute the market shares for ten periods, holding the prices of all products fixed over time at their averages. Then, I cut the price of all sizes of one of the products by ten percent for all ten periods, and compute the percentage change in the product's market share, divided by ten. Note that

Table 8 Counterfactual: long term empirical price elasticities

	Full model				Switching costs only				Learning Only			
	Cheer	Surf	Dash	Tide L	Cheer	Surf	Dash	Tide L	Cheer	Surf	Dash	Tide L
Cheer	-4.93	0.29	0.15	0.29	-4.27	0.33	0.12	0.26	-3.49	0.26	0.14	0.22
Surf	-	-5.76	0.18	0.46	-	-5.74	0.24	0.34	-	-5.16	0.16	0.31
Dash	-	-	-6.27	0.43	-	-	-6.23	0.35	-	-	-5.13	0.35
Wisk L	0.2	0.25	0.13	0.21	0.11	0.22	0.14	0.22	0.11	0.18	0.09	0.2
Tide L	0.19	0.27	0.11	-3.31	0.12	0.24	0.13	-3.53	0.11	0.17	0.09	-2.5
Tide P	0.13	0.18	0.1	0.12	0.14	0.34	0.12	0.16	0.1	0.16	0.09	0.14

This table shows the empirical price elasticities implied by my model estimates, $\frac{\partial Q_i}{\partial P_j} \frac{P_j}{Q_i}$, for $t = 1$. The row labels show the product i that is being affected, and the period of interest. The column labels show the product j whose price is being changed. The price change is assumed to occur from period 1 onwards, and consumers are assumed to understand this, which is why this is a long term elasticity. Furthermore, when this elasticity is computed, the prices of all products are set to be constant over time. Thus the number in the first row and second column, 0.29, shows the impact of a permanent price cut in Cheer from period 1 onwards on Surf's period 1 market share. The first four columns show the elasticities implied by the full model, the next four show the elasticities for the no learning model, and the next four for the no switching costs model

this type of price elasticity is not simply the impact of a 1 period price cut on the product's market share. When I compute the market shares at the lower price, consumers expect the lower price to last forever, and adjust their value functions accordingly.

The first column of the table shows the impact of a long term price cut for Cheer Liquid on Cheer Liquid, Wisk Liquid, Tide Liquid, and Tide Powder. I show the simulated for the first period only, but the findings I will discuss below hold in later periods. The own price elasticity of Cheer for a permanent price cut in Cheer is -4.93 , and cross-price elasticity of a price cut on Cheer for Wisk's quantity in period 1 is 0.20 . Note that when I computed the elasticities in the first column, I assumed that Surf and Dash were not available, and that there was learning in Cheer. The second, third, and fourth columns show the impacts of long term price cuts in Surf Liquid, Dash Liquid, and Tide Liquid respectively. For the Tide Liquid price cut, I assume that all products are available, and that there is learning in Dash. The own price elasticity of the new products are estimated to be significantly larger than in the case of temporary price cuts. This is likely due to the fact that the price cuts are affecting consumers' future utility through the switching costs and the learning, in addition to their current utility. The cross-elasticities of demand are roughly between 10% and 46%.

The results for the no learning model are shown in the fifth through eighth columns. The own price elasticities for the new products are somewhat lower, although the difference is small. The impact on cross-price elasticities is larger: when learning is ignored, the cross elasticities between new and established products generally drops, the magnitude of the drop being as large as 45%. This happens because the option value of learning makes new products more attractive; a price cut on a new product will be more likely to draw customers away from old products in the presence of learning.

The impact of ignoring switching costs is shown in the final three columns. Interestingly, the own price elasticities of all products are estimated to be much smaller, and the cross price elasticities are smaller as well. This finding seems counterintuitive—because switching costs make it more difficult to switch between products, one might expect own and cross price elasticities to rise when the switching costs are removed. There are two forces that will tend to move the elasticities in the opposite direction. First, the estimated taste variances are larger in the no switching costs model, which compensates for the lack of switching costs somewhat. Second, there is a dynamic effect, which is a result of the fact that this price cut is a long term price cut. When there is a long term price cut on a product with switching costs, that product becomes more attractive to consumers because they realize that the price of the product will be low in the future, and therefore they will be less likely to incur the switching cost in the future. The fact that the dynamics are driving this finding becomes obvious when one compares this finding to the impact of the one period price cut in Table 6. There, the impact of a one period price cut on current demand was sometimes larger in the no switching costs model, and sometimes smaller, and there was no systematic pattern. Because the price cut only occurred for

a single period, the dynamic aspect did not come into play. The magnitude of the bias in long term elasticities is significant: it is as much as 30% for the own elasticities, and as much as 45% on the cross elasticities.

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

6 Conclusions and extensions

This paper estimates a novel model of learning and switching costs, finding evidence for both types of dynamics. By estimating restricted version of the model and computing counterfactuals that are relevant to managers and policymakers, I demonstrate that a model of learning and switching costs has different economic implications than a model with only switching costs and a model with only learning. In my 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.

There are a number of extensions for this research that would be useful. First, this paper abstracts from the supply side: for example, the counterfactuals I compute are partial equilibrium counterfactuals and do not account for competitor responses. It would be interesting to examine the model's supply side implications: for example, we might be interested in knowing the impact of learning on the ease of new product entry, or on equilibrium pricing behavior. It would be interesting to know how the different models explored in these papers might lead researchers to different conclusions. One way to perform this kind of exercise would be to take the demand systems as given, solve for the market equilibria implied by each model, and compute comparative statics. This sort of exercise has been performed in markets with switching costs in Dubé et al. (2008). In this paper, the computation of the market equilibrium

is tractable because switching costs are the only source of dynamics, and consumers are not forward-looking (Dubé et al. (2008) argue that the problem with forward-looking consumers is similar to the problem with myopic ones, when firms prices follow a Markov process.). Solving the model with forward-looking consumers who learn their match values for new products is a more difficult task.

References

- Ackerberg, D. (2003). Advertising, learning, and consumer choice in experience goods markets: A structural empirical examination. *International Economic Review*, *44*(3), 1007–1040.
- Ackerberg, D. (2009). A new use of importance sampling to reduce computational burden in simulation estimation. *Quantitative Marketing and Economics*, *7*, 343–376.
- Cannon, D., Watzman, A., Lazorko, L., Bishop, J., Shafers, T., Kemezis, P., et al. (1987). Soaps and detergents: A basketful of high-tech products. *Chemical Week*, *140*, 22–33.
- Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data* (pp. 3–38). Cambridge, New York and Sidney: Cambridge University Press.
- Chintagunta, P. (1999). Variety seeking, purchase timing, and the lightning bolt brand choice model. *Management Science*, *45*(4), 486–498.
- Chintagunta, P., Kyriazidou, E., & Perktold, J. (2001). Panel data analysis of household brand choices. *Journal of Econometrics*, *103*(1–2), 111–153.
- Crawford, G., & Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica*, *37*, 1137–1174.
- DeGroot, M. (1970). *Optimal statistical decisions*. McGraw-Hill, Inc.
- Dubé, J., Hitsch, G., & Rossi, P. (2008). Do switching costs make markets less competitive? Working Paper.
- Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing Science*, *15*(4), 359–378.
- Erdem, T., Imai, S., & Keane, M. (2003). A model of consumer brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, *1*(1), 5–64.
- 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.
- Erdem, T., Keane, M., Öncü, T., & Strebler, J. (2005). Learning about computers: An analysis of information search and technology. *Quantitative Marketing and Economics*, *3*, 207–247.
- 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.
- Farrell, J., & Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. In M. Armstrong & R. Porter (Eds.), *Handbook of industrial organization* (Vol. 3). North-Holland: Elsevier.
- Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.
- Givon, M., & Horsky, D. (1979). Application of a composite stochastic model of brand choice. *Journal of Marketing Research*, *16*, 258–267.
- Goettler, R., & Clay, K. (2010). *Tariff choice with consumer learning and switching costs*. Working Paper.
- Hartmann, W. (2006). Intertemporal effects of consumption and their implications for demand elasticity estimates. *Quantitative Marketing and Economics*, *4*(4), 325–349.
- Hendel, I., & Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica*, *74*, 1637–1673.
- Imai, S., Jain, N., & Ching, A. (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica*, *77*(6), 1865–1899.

- Israel, M. (2005). Services as experience goods: An empirical examination of consumer learning in automobile insurance. *American Economic Review*, 95(5), 1444–1463.
- Johnson, N., & Kotz, S. (1970). *Continuous multivariate distributions I*. New York: Wiley.
- Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies*, 62, 515–539.
- McAlister, L., & Pessemier, E. (1982). Variety-seeking behavior: An interdisciplinary review. *The Journal of Consumer Research*, 9(3), 311–322.
- Moshkin, N., & Shachar, R. (2002). The asymmetric information model of state dependence. *Marketing Science*, 21, 435–454.
- Norets, A. (2007). *Estimation of dynamic discrete choice models using artificial neural network approximations*. Working Paper.
- Norets, A. (2009a). *Implementation of bayesian inference in dynamic discrete choice models*. Working Paper.
- Norets, A. (2009b). Inference in dynamic discrete choice models with serially correlated unobserved state variables. *Econometrica*, 77, 1665–1682.
- Osborne, M. (2006). *Consumer learning and habit formation in packaged goods markets*. Ph.D. thesis, Stanford University.
- Pollack, R. (1970). Habit formation and dynamic demand functions. *Journal of Political Economy*, 78(4), 745–763.
- Rossi, P., Allenby, G., & McCulloch, R. (2005). *Bayesian statistics and marketing*. Wiley.
- Roy, R., Chintagunta, P., & Haldar, S. (1996). A framework for investigating habits, the hand of the past, and heterogeneity in dynamic brand choice. *Marketing Science*, 15(3), 280–289.
- Shum, M. (2004). Does advertising overcome brand loyalty? Evidence from the breakfast-cereals market. *Journal of Economics and Management Strategy*, 13, 241–272.
- Spinnewyn, F. (1981). Rational habit formation. *European Economic Review*, 15, 91–109.
- Train, K. (2003). *Discrete choice methods with simulation*. New York: Cambridge University Press.