IO
7/21/06
10:45 AM

Do Switching Costs Make Markets Less Competitive?

by

Jean-Pierre Dubé
Günter J. Hitsch
Peter E. Rossi

Graduate School of Business
University of Chicago

Revised June 2006

This is a Preliminary Draft

Keywords: switching costs, dynamic model, brand loyalty, state dependence
JEL Classifications: D1, D4, L1, M3

Abstract

The conventional wisdom in economic theory holds that switching costs make markets less competitive. This paper challenges this claim. We find that steady-state equilibrium prices may fall as switching costs are introduced into a simple model of dynamic price competition that allows for differentiated products and imperfect lock-in. To assess whether this finding is of empirical relevance, we consider a more general model with heterogeneous consumers. We calibrate this model with data from a frequently purchased packaged goods market where consumers exhibit inertia in their brand choices, a behavior consistent with switching costs. We estimate the level of switching costs from the brand choice behavior in this data. At switching costs of the order of magnitude found in our data, prices are lower than without switching costs.

All correspondence may be addressed to the authors at the Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637; or via e-mail: jdube@gsb.uchicago.edu, guenter.hitsch@chicagogsb.edu, or peter.rossi@chicagogsb.edu.

We are grateful for comments and suggestions from seminar participants at the University of Chicago Structural lunch, Erasmus University in Rotterdam, Northwestern University and Yale University.

1. Introduction

For some models, it has been shown that switching costs make markets less competitive in the sense that prices are higher in equilibrium with switching costs than without (c.f. Farrell and Klemperer 2005). Klemperer (1995) conjectures that this result is likely to hold across a wide array of models whenever firms are unable to price discriminate between existing and new customers. In this paper, we analyze a class of switching cost models that can be taken directly to data. In order to account for customer behavior and the observed outcomes in "real world" markets, we allow for product differentiation and the possibility of imperfect lock-in of customers. In our model, we find that there are plausible values of switching costs for which prices decline in equilibrium relative to the zero switching cost case. We estimate the level of switching costs from scanner data on frequently purchased consumer products for which customers exhibit "brand loyalty," a particular form of switching costs. The result that switching costs lead to lower, not higher prices is robust to flexible demand specifications and holds for a wide range of switching costs centered on the values obtained from our data.

The existing literature on switching costs stems from Klemperer (1987) who shows that prices increase in the second period of a two-period model: "bargain then rip-off prices." In general, markets with switching costs exhibit two forces that have opposite effects on equilibrium prices. First, firms have an incentive to invest in their market share, which induces them to lower their prices. Second, firms want to "harvest" their base of customers for whom switching is costly by raising prices. In two period models, the typical prediction is that prices are low in the first period, when firms compete for market share, and then rise in the second period to a level above that which would exist in a model without switching costs. Cabral and Villas-Boas (2005) provide sufficient conditions under which this "bargain then rip-off" pricing can lower firms' profits in equilibrium.

In models without a terminal period or with long-lived consumers, theories have been developed to assess which of the two forces, harvesting versus investing, will dominate in a long-run equilibrium. Farrell and Shapiro (1988), Padilla (1995) and Anderson et al. (2004) consider markets with undifferentiated products and overlapping generations of consumers, where price equals cost in the Bertrand-Nash equilibrium without switching costs. Each of these papers demonstrates that the presence of switching costs generates economic rents with equilibrium price above marginal cost. However, product differentiation can generate positive profits even without switching costs and, therefore, may lead to lower prices in the presence of switching costs as firms compete to "lock-in" customers.

Beggs and Klemperer (1992) consider the differentiated products case and customers that have the potential to live many periods and show that, even in this case, the equilibrium prices are higher in the presence of switching costs. The model considered in Beggs and Klemperer assumes that customers who purchase a product become perfectly locked-in (i.e. will never switch in the future), an assumption of infinite switching costs. Beggs and Klemperer state that the main results of their paper would hold for large, but finite switching costs. Viard (2003) and Doganoglu (2005) challenge this claim using slightly different models in which firms sell to overlapping generations of consumers whose tastes change over time. For sufficiently low switching costs, equilibrium prices are found to be lower than in the absence of switching costs. It remains to be seen if these levels of switching costs are similar to those found to be consistent with the brand switching behavior of consumers.

A small body of empirical research has developed which supports the qualitative conclusions formed from the theoretical literature. Viard (2005) and Shi et al. (2006) look at the effects of telephone number portability on the prices of telephone service. Both papers conclude that, after number portability is achieved, prices fall, consistent with the theoretical prediction that prices are increasing in switching costs. Stango (2002) documents a similar positive correlation between observed prices and switching costs in the credit cards market.

This paper considers models of switching costs with differentiated products that allow for the possibility of imperfect lock-in. That is, we consider the case where a consumer may switch brands in spite of the switching cost. We find that steady-state prices under a wide range of plausible switching costs often fall, and hence switching costs can make markets more competitive. We first illustrate this effect in a very simple model that allows us to explore the forces that drive this main result. We solve for a Markov perfect equilibrium using computational techniques. Computation has obvious drawbacks over analytical methods, yet progress in the extant literature on switching costs has been impeded by the difficulty of characterizing the complex behavior that arises under state-dependent demand.

We find that in our model, steady-state prices first fall and then rise as the magnitude of the switching costs increases. Hence, the counter-example to Klemperer's claim is dependent on model parameters. In order to investigate whether our result is merely a theoretical curiosity or actually a relevant prediction for some real-world markets, we extend the model and allow for a rich demand side, which generalizes much of the empirical literature on differentiated product demand systems. We estimate the demand model from data on two categories of frequently purchased consumer products (refrigerated orange juice and margarine), and then compute the price equilibrium.

Orange juice is an example of a branded, frequently purchased product. A large literature in marketing has shown that such products exhibit "brand loyalty," a special form of switching costs (Klemperer 1995, Farrell and Klemperer 2005). Brand loyalty may have "psychological" sources or it may be the rational result of "shopping costs," where consumers do not re-optimize the set of products considered and bought at each shopping occasion. In this paper, we are not concerned with the exact source of brand loyalty, but rather whether it in fact exists and can be identified from observed purchasing behavior.

Switching costs are typically not directly observed. Instead, the analyst must infer their magnitude from the observed switching behavior of consumers. Given that consumers are heterogeneous in their tastes, we face the well-known problem of separating heterogeneity and state-dependence in demand. Our approach is to use panel data with a reasonably long time dimension and considerable price variation coupled with a semi-parametric model of consumer heterogeneity. This allows us to separate heterogeneity from state dependence in demand and obtain reasonably precise estimates of the distribution of switching costs across consumers.

Our estimated switching costs are on the order of 15 to 60 per cent of the purchase price of the goods. When these switching costs are used in model simulations, equilibrium prices decrease relative to prices without switching costs. This prediction is very robust to variation in the parameter values. In particular, if switching costs are scaled up to several times those inferred from our data, we still find that prices decline in the presence of switching costs.

2. Model

In this section, we first develop a simple model of price competition in markets with switching costs. As a point of departure from much of the established literature, we allow for product differentiation and the possibility that consumers switch away from products they have previously purchased, which are features commonly present in actual markets. The simple model allows us to explore the economic forces that determine equilibrium prices. Finally, we show how the simple model can be easily extended to be suitable for an empirical application.

Model Details

We consider a market with J competing firms. Each firm sells one product. Time is discrete, $t = 0,1,\ldots$. There is exactly one consumer in the market, who chooses among the J products and an outside option in each period.

In each period, the consumer is loyal to one product, $j = s_t$. The loyalty variable $s_t \in X = \{1,\ldots,J\}$ summarizes all current-period payoff-relevant information, and describes the state of the market. Demand is derived from a discrete choice model. Conditional on price p_{jt} and her current loyalty state s_t, the consumer's utility index from the choice of j is

$$U_{jt} = \delta_j + \alpha p_{jt} + \gamma I\{s_t = j\} + \lambda \varepsilon_{jt}. \tag{1}$$

As is common with much of the empirical literature on demand estimation, we assume that the random utility component ε_{jt} is i.i.d. Type I Extreme Value distributed. λ determines the scale of the utility shock, and thus the degree of horizontal product differentiation between the products. In the limiting case of $\lambda = 0$, product differentiation is purely vertical. If the consumer is loyal to j but buys product $k \neq j$, she foregoes the utility component γ. Thus, she implicitly incurs a switching cost.

An alternative specification of the utility index is

$$U_{jt} = \delta_j + \alpha p_{jt} - \gamma I\{s_t \neq j\} + \lambda \varepsilon_{jt}. \tag{2}$$

Under this formulation, the consumer incurs an explicit switching cost γ when she chooses a product to which she is not loyal. If there is no outside alternative, the state dependence model (1) is equivalent to the switching cost formulation in (2). With an outside alternative, there are some subtle differences as discussed in the section on comparison of switching cost formulations below. Those familiar with the switching cost literature may find the pure switching cost specification in (2) more appealing as this corresponds to a literal interpretation of an explicit switching cost which might be a monetary cost, a search cost, a "hassle" cost, or a psychological barrier. However, the empirical literature has favored the state dependence specification in (1). Thus, we will carry forward both specifications and report results for both.

Let $U(j, s_t, p_t)$ denote the deterministic component of the utility index, such that $U_{jt} = U(j, s_t, p_t) + \lambda \varepsilon_{jt}$. The utility from the outside alternative is $U_{0t} = U(0, s_t, p_t) =$ $= \delta_0 + \varepsilon_{0t}$. If $\lambda > 0$, demand is given by the logit choice probabilities

$$P_j(s_t, p_t) = \frac{\exp(U(j, s_t, p_t) / \lambda)}{\sum_{k=0}^{J} \exp(U(k, s_t, p_t) / \lambda)}.$$

If there is no horizontal product differentiation ($\lambda = 0$), the consumer buys the product with the highest utility index. If there is more than one product that maximizes utility, the consumer chooses the product to which she is loyal if it is among the utility-maximizing options, and randomizes among the utility-maximizing products otherwise.

The current loyalty state of the consumer, s_t, evolves as follows. If the consumer buys product $k \neq j$ in period *t*, then $s_{t+1} = k$, i.e. she becomes loyal to *k*. If the consumer buys product $j = s_t$ or chooses the outside option, then $s_{t+1} = s_t$, i.e. her loyalty remains unchanged. The firms cannot observe the random utility component. Hence, conditional on a product price vector, p_t, the state variable follows a Markov process from the firm's point of view:

$$\Pr\{s_{t+1} = j \mid s_t, p_t\} = \begin{cases} P_j(s_t, p_t) + P_0(s_t, p_t) & \text{if } j = s_t, \\ P_j(s_t, p_t) & \text{if } j \neq s_t. \end{cases} \tag{3}$$

Below, we discuss the extension to forward-looking consumer behavior.

Conditional on all product prices and the state of the market, firm *j* receives the expected current-period profit $\pi_j(s_t, p_t) = P_j(s_t, p_t) \cdot (p_{jt} - c_j)$. c_j is the marginal cost of production, which does not vary over time. Firms compete in prices, and choose Markovian pricing strategies that depend on the current payoff-relevant information, summarized by s_t. This assumption rules out behavior that conditions current prices also on the history of past play, and thus collusive strategies in particular. We denote firm *j*'s strategy by $\sigma_j : X \to \mathbb{R}$. Firms discount the future using the factor β, $0 \leq \beta < 1$. For a given profile of strategies, $\sigma = (\sigma_1, ..., \sigma_J)$, the expected PDV of profits, $\sum_{t=0}^{\infty} \beta^t \pi_j(s_t, \sigma(s_t))$, is well-defined. Conditional on a profile of competitor's strategies, σ_{-j}, firm *j* chooses a pricing strategy that maximizes its expected value. Associated with a solution of this problem is firm *j*'s value function, which satisfies the Bellman equation

$$V_j(s) = \max_{p_j \geq 0} \left\{ \pi_j(s,p) + \beta \left(\sum_{k=1}^{J} P_k(s,p) V_j(k) + P_0(s,p) V_j(s) \right) \right\} \qquad \forall s \in X. \qquad (4)$$

In this equation, the price vector consists of firm *j*'s price and the prices prescribed by the competitor's strategies, $p = (\sigma_1(s),...,\sigma_{j-1}(s), p_j, \sigma_{j+1}(s),...,\sigma_J(s))$. Therefore, the Bellman equation (4) depends on the pricing strategies chosen by the competitors. Note that the expectation of the firm's future value is taken with respect to the transition probabilities of s_t, which are directly related to the current choice probabilities.

If $p_j = \sigma_j(s)$ attains the right-hand side of the Bellman equation for each state *s*, then σ_j is a *best response* to the strategy profile σ_{-j}. With these preliminaries, we can define the solution concept for the pricing game:

Definition. A strategy profile σ^* is a *Markov perfect equilibrium* if each σ_j^* is a best response to σ_{-j}^*. That is, for each firm *j* and state *s*, $p_j = \sigma_j(s)$ attains the right-hand side of the Bellman equation (4).

For this pricing game, there always exists a Markov perfect equilibrium in pure strategies. The proof for the case of horizontal product differentiation, $\lambda > 0$, relies on the quasi-concavity of a logit-based objective function, which is well known for the static case and also holds for the case of dynamic competition considered here. Quasi-concavity ensures that each player has a unique best response. The proof is presented in the Appendix. While we can show the existence of a pure-strategy equilibrium, we cannot characterize the equilibrium policies analytically. Instead, we solve the game numerically for different parameter values.

In the case of no horizontal product differentiation, $\lambda = 0$, the equilibrium can be characterized analytically. We focus on the case of symmetry across players, where all firms have the same utility intercepts and costs. We assume that $\delta > c \geq 0$.

Proposition. *Let* υ *be such that* $0 \leq \upsilon \leq (1-\beta)\gamma$ *and* $c + \upsilon \leq \delta + \gamma$. *Then under the assumptions stated above there is a symmetric Markov perfect equilibrium with pricing strategies* $\sigma_j^*(j) = c + \upsilon$ *and* $\sigma_j^*(k) = c + \upsilon - \gamma$ *for all* $k \in X,\ k \neq j$.

Proof. j denotes the product to which the customer is loyal and k denotes any other product. Because $p_j = c + \upsilon = p_k + \gamma$, the customer's utility index is the same for all products. Therefore, by assumption she will not switch from product j to k, and because $0 \leq \delta + \gamma - (c + \upsilon)$, she will not choose the outside option. The value from this strategy is $V_j(j) = (1-\beta)^{-1}\upsilon$ and $V_j(k) = 0$. In order to assess whether the proposed strategies constitute a best response for each player, we only need to consider one-period deviations. If firm j reduces its price, it will reduce its current-period profit and leave its future value unchanged. If firm j raises its price, it will loose its loyal customer and receive a payoff of zero now and in future. Hence, $p_j = c + \upsilon$ is a best response to p_k. Competitor k needs to offer a price $p_k = c + \upsilon - \gamma - \varepsilon,\ \varepsilon > 0$, in order to acquire the customer. Because $\upsilon \leq (1-\beta)\gamma$, the present value from this one-period deviation is negative:

$$\upsilon - \gamma - \varepsilon + \beta \frac{\upsilon}{1-\beta} = \frac{\upsilon}{1-\beta} - \gamma - \varepsilon < 0.$$

Alternatively, firm k cannot improve on its current outcome by raising its price, and hence, $p_k = c + \upsilon - \gamma$ is a best response to p_j.

Model Predictions

We now explore the predictions of the pricing model developed above. To keep the exposition as simple as possible, we focus on symmetric games with two firms. Each firm has the same utility intercept and marginal production cost. In a symmetric equilibrium, $\sigma_1^*(1) = \sigma_2^*(2)$ and $\sigma_1^*(2) = \sigma_1^*(1)$. We therefore only need to know firm 1's pricing policy to characterize the market equilibrium.

We first consider the case of homogenous products $(\lambda = 0)$. The proposition above states that switching costs allow firms to raise prices above the baseline Bertrand outcome, where $p = c$. In particular, there is an equilibrium where the firm that possesses the loyal customer increases its price above cost by the value $\upsilon = (1-\beta)\gamma$. υ is the flow value of the switching cost. If the firm charges an even higher price, the competitors could poach the customer by subsidizing the switching cost, incurring a loss in the current period, and recouping this loss by pricing above cost in the future. In summary, if products are not differentiated, then we find that switching costs make markets less competitive, as predicted by much of the previous literature.

We now turn to the case of differentiated products and switching in equilibrium $(\lambda > 0)$. In the case of homogenous products, the customer never switches in equilibrium, and hence the realized transaction price is the price that the customer pays for the product that she is loyal to. In the case of product differentiation, the customer sometimes switches, and therefore we characterize the equilibrium outcome by the average transaction price paid, conditional on a purchase:

$$p^a = \frac{P_1(1,\sigma^*(1)) \cdot \sigma_1^*(1) + P_2(1,\sigma^*(1)) \cdot \sigma_2^*(1)}{P_1(1,\sigma^*(1)) + P_2(1,\sigma^*(1))}.$$

That is, p^a is the expected price paid in state $s_t = 1$, which—due to symmetry—is the same as the expected price paid in state $s_t = 2$.

Figure 1 shows the relationship between the level of switching costs and the average transaction price for the case of $\delta_j = 1$, $c_j = 0.5$, $\alpha = 1$, and $\lambda = 1$. We find that prices initially fall and then rise for larger switching cost levels. Indeed, only for switching cost levels larger than 4 does the average transaction price exceed the transaction price without switching costs. For a switching cost level of $\gamma = 3$, despite the fact that the probability of staying loyal is 0.77, the average prices are lower than without switching costs[1]. Table 1 displays the average transaction price and the individual prices set by firms 1 and 2. The table also shows the purchase probabilities for each product and the probability that the customer stays loyal. In this example, the price of firm 1 increases and the price of firm 2 decreases in the level of switching costs. However, there are other cases where both firm's prices decrease for small switching cost levels (see the case of $\lambda = 0.2$ in Table 2).

In Table 2 we explore the relationship between the random utility scale factor (λ) and equilibrium prices at different switching cost levels. As λ decreases, we observed prices falling for a fixed level of switching costs as there is less horizontal differentiation. What is important to note is that prices fall and then rise in switching costs for all values of λ considered. For smaller values of λ, the switching cost level as which prices begins to rise are lower. Holding the other parameter values constant, lower values of λ imply less horizontal product differentiation and increase the probability that the customer stays loyal.

[1] This pattern of declining-then-rising equilibrium prices is robust to other parameter values. We have not been able to find an example where average transaction price always—even for small switching cost levels—increases.

This higher rate of retention increases the incentive to "milk" the loyal customer and causes the average transaction price to begin increasing at lower switching cost levels.

The result of decreasing-then-increasing equilibrium prices also persists across different levels of the outside good intercept, δ_0. In particular, this outcome arises even if there is no outside option ($\delta_0 = -\infty$).

Our results show that the conjectured effect of switching costs on prices—switching costs make markets less competitive—need not be true in a model that is simple, yet nonetheless the foundation of an important and widely used class of empirical demand models. As previously discussed, under competition with switching costs firms face two incentives that work in opposite directions. First, firms can "milk" a loyal customer by charging higher prices. Second, firms can invest into future loyalty by lowering current prices. Our results imply that either force can dominate in equilibrium. We now examine the case where only the first force ("milking") is present to illustrate this point. To exclude the investment motive, we consider competitors who do not anticipate the future benefits from lowering current prices, and are hence myopic ($\beta = 0$). Figure 1 shows the average transaction price paid under this scenario, and allows us to compare the pricing outcomes with fully rational, forward-looking firms and myopic decision makers. After eliminating the investment motive, prices always rise in the degree of switching costs—switching costs make markets less competitive. In contrast, a forward-looking firm has an incentive to lower its price to poach a customer who is loyal to its competitor. The firm with the loyal customer anticipates the competitor's incentive, and lowers its price to prevent the customer from switching. Therefore, the average transaction price under competition with forward-looking firms is always lower than the average price under myopic competition.

Forward-Looking Consumers

The demand side of the model can be extended to allow for forward-looking consumer behavior, where the customer anticipates the consequences of becoming loyal to a particular product. Demand is then similar to Rust (1987); the technical details can be found in the Appendix. In the particular case of symmetric competition, which we discussed above, demand and hence equilibrium prices are the same under both myopic and forward-looking consumer behavior. Due to symmetry, the customer's current and future payoffs are identical regardless of the identity of the product to which she is loyal. If she is currently loyal to product 1, for example, she faces the identical choice situation next period regardless of whether she switches to product 2 today or remains loyal to product 1. This argument depends on the assumption that the customer is always loyal to one of the products, and not to the outside option.

Comparison of Switching Cost Formulations

In our introductory discussion of the demand side of the model, we noted two alternative switching cost formulations. Both models imply inertia in consumer choices over time: ceteris paribus, under switching costs the customer is more likely to buy the product to which she is loyal. The difference is in the exact way that the switching cost enters the utility index: in model (1), the customer "gains" additional utility from the product to which she is loyal, while in (2) she pays a monetary or utility cost when she switches to another product. Therefore, some may consider (2) a more "natural" formulation of demand under switching costs. However, the empirical brand choice literature has routinely used model (1) to capture observed inertia in household brand choices, c.f. Erdem (1996), Roy, Chintagunta, and Haldar (1996), Keane (1997), Seetharaman, Ainslie, and Chintagunta (1999), Seetharaman (2004) and Shum (2004). We call model (1), which we analyzed before, the "state dependence model" of switching costs[2], and model (2) the "pure switching cost model." To clarify the difference between the two, we solve and compare the equilibria of both models using the same parameter values.

Figure 2 shows the average price for both the state dependence and the pure switching cost models. Prices fall and then rise in the pure switching cost model, as we previously found for the state dependence model. Furthermore, the average price is always lower in the pure switching cost model compared to the state dependence model. However, in contrast to the case analyzed previously, in the pure switching cost model prices may decline even in the case of myopic competition ($\beta = 0$), where the firms do not invest in future loyalty. In order to understand the intuition for this difference in outcomes, note that varying the switching cost level, γ, varies both the relative purchase probabilities of the products in the market and the share of the outside good. In the pure switching cost model, the share of the outside good increases in γ, holding prices constant, while in the state dependence model the share of the outside good decreases. The market becomes less attractive to all firms jointly in the former case, and more attractive in the latter. Recall that in a static Bertrand game with logit demand, the price of a product is proportional to its market share. In the pure switching cost model, the average share of the firms in the market decreases in the level of switching costs. Figure 2 shows that therefore, switching costs can lower the average transaction price under myopic competition. Under pure switching costs, prices may decline in the level of switching costs even if the investment motive is absent. In

[2] Klemperer (1995) notes that state dependence can be viewed as a form of switching costs.

the state dependence model, on the other hand, prices increase under myopic decision making[3].

Overlapping Generations Version of the Simple Model

In the theory literature on switching costs, competition in an infinite horizon setting (Beggs and Klemperer 1992 and Padilla 1995, for example) is examined using overlapping generations (OLG) models. Shortening the lifetime of a customer reduces the incentive to invest in customer loyalty and prices might therefore be higher compared to an infinite horizon setting. We now develop an OLG version of our simple state dependence model and examine the robustness of our previous finding that switching costs can lower equilibrium prices.

In each period, a new customer is born. The customer lives for two periods and, hence, there are always a "young" and an "old" customer in the market. A customer can be loyal to one of the *J* products, or she can be unattached, i.e. loyal to the outside alternative. If a customer is loyal to the outside alternative, she does not incur a switching cost for any product choice. Otherwise, her demand is as in the model analyzed before. When the young customer is born, she is unattached. If she chooses the outside alternative, she stays unattached in the next period, when she is old. Otherwise, if she buys product *j* she becomes loyal to *j*. The state of the market is now described by $s_t \in \{0,1,...,J\}$, the choice that the currently old customer made in the previous period, $t-1$.

Table 3 shows the average transaction prices paid by the young and the old customer for different switching cost levels. The model was solved with forward-looking consumers. Due to lock-in, the old customer always pays a higher average price than the young

[3] For this reason, we chose the state dependence formulation to illustrate the basic economics of pricing under switching costs.

customer. Unless switching cost levels are sufficiently large, however, both the young and the old customer pay a lower price, on average, than in the case without switching costs. The young customer, in particular, generally pays a lower price. Similar to the basic model examined above, reducing the magnitude of the random utility scale factor, λ, causes the average transaction price of the old customer to begin rising at lower switching cost levels.

Thus, our main conclusion that switching costs do not necessarily lead to higher prices is robust to a different model formulation as well as a wide range of parameter values.

3. Empirical Model

In both the infinite horizon and OLG versions of the simple model, we observe that equilibrium prices are *lower* with switching costs than without for a wide range of parameter values. Switching costs are rarely directly observed (some components may be known, but the "hassle" costs of switching are not). For this reason, we must turn to data on the purchase histories of customers to infer switching costs from the observed patterns of switching between brands in the face of price variation. Consumer panel data on the purchases of packaged goods are ideal for estimating switching costs as the panel length is long relative to the average inter-purchase times and there is extensive price variation. If households are observed to forego large utility increases afforded by a temporary price cut or sale, we can infer that there must be a relatively high level of switching cost.

To infer switching costs from consumer panel data, we must enrich our model to consider multiple differentiated products as well as multiple consumer types or consumer heterogeneity. It is well documented (c.f. Allenby and Rossi 1999) that consumers exhibit a very high degree of heterogeneity with differing product preferences (intercepts) as well as price sensitivities (price coefficient). It is also entirely possible that households will exhibit differing degrees of switching costs. Our approach will be to specify a very flexible model of consumer heterogeneity around a standard logit specification.

Extending the Simple Model

For these reasons, we consider a market populated by many heterogeneous customers. We allow for N different types and assume that there is a continuum of consumers with mass μ_n for each type *n*. The latter assumption is for convenience. As we will see below, it makes the evolution of the state vector deterministic. Demand at the consumer level is identical to the simple model described previously, but the utility parameters are now type-specific. We thus index the choice probability of a consumer by her type. The probability of buying product *j* by a consumer of type *n* in state s_t, for example, is denoted by $P_j(s_t, p_t; n)$.

To summarize the overall state of the market, we need to know the distribution of consumers of different types over loyalty states. Let x_{jt}^n be the fraction of consumers of type *n* who are loyal to product *j*. The vector $x_t^n = (x_{1t}^n, ..., x_{Jt}^n)$ summarizes the distribution over loyalty states for all consumers of type *n*, and $x_t = (x_t^1, ..., x_t^N)$ summarizes the state of the whole market. As before, we denote the state space by X. We write each state as an NJ dimensional vector. Note, however, that by definition, $\sum_{j=1}^{J} x_{jt}^n = 1$ for all types *n*. Hence, the information contained in the state can be described more parsimoniously by a vector that has only $N(J\text{-}1)$ dimensions. This is not theoretically important, but allows us to simplify the solution algorithm for the model on a computer.

Aggregate demand is obtained by summing household level demand over consumer types and loyalty states:

$$D_j\left(x_t, p_t\right) = \sum_{n=1}^{N} \mu_n \left(\sum_{k=1}^{J} x_{kt}^n P_j\left(k, p_t; n\right) \right).$$

In contrast to the simple model, demand is now deterministic. This is a consequence of the assumption that there is a continuum of consumers for each type.

In the discussion of the simple model, we described the law of motion of the state variable at the individual level. The transition of the aggregate state can be easily derived from the transition probabilities of the individual states, as shown in (3). Conditional on a price vector p_t, we can define a Markov transition matrix $Q(p_t; n)$ with elements

$$Q_{kj}(p_t; n) = \Pr\{s_{t+1} = j \mid s_t, p_t; n\}.$$

$Q_{kj}(p_t; n)$ denotes the probability that a household of type *n* who is currently loyal to *j* will become loyal to product *k*. The whole state vector for type *n* then evolves according to the Markov chain

$$x_{t+1}^n = Q(p_t; n) x_t^n.$$

Households can change loyalty states but not types such that the overall market state vector x_t also evolves according to a Markov Chain with a block diagonal transition matrix. The evolution of the state vector is deterministic, and we denote the transition function by *f*, $x_{t+1} = f(x_t, p_t)$.

Firm *j*'s current-period profit function is $\pi_j(x_t, p_t) = D_j(x_t, p_t) \cdot (p_j - c_j)$. As in the case of the simple model, firms compete in Markovian strategies, $\sigma_j : X \to \mathbb{R}$. The best response to a profile of competitor's strategies, σ_{-j}, is found from the Bellman equation:

$$V_j(x) = \max_{p_j \geq 0} \left\{ \pi_j(x, p) + \beta V_j(f(x, p)) \right\} \qquad \forall x \in X.$$

Here, $p=(\sigma_1(x),...,\sigma_{j-1}(x),p_j,\sigma_{j+1}(x),...,\sigma_J(x))$.

The definition of the Markov equilibrium concept is the same as in the simple model. In contrast to the simple model, we cannot prove that a pure strategy equilibrium generally exists. Even in static games of price competition, restrictions on the distribution of consumer tastes typically need to be imposed to establish the existence of a pure strategy equilibrium (Caplin and Nalebuff 1991). In general, the "non-parametric" distribution of tastes that our model allows for does not obey these restrictions. In our empirical application, we can therefore only establish the existence of a pure strategy equilibrium computationally on a case-by-case approach.

Econometric Specification

We have extended the simple model by using a standard multinomial logit model conditional on consumer/household type. The probability that household h chooses alternative j given loyalty to product k is given by

$$P\left(j\mid s=k;\theta^h\right)=\frac{\exp\left(\delta_j^h+\alpha^h p_j+\gamma^h I\{s=j\}\right)}{1+\sum_{k=1}^J \exp\left(\delta_j^h+\alpha^h p_j+\gamma^h I\{s=k\}\right)}.$$

To accommodate differences across households, we use a potentially large number of household types and a continuum of households of each type. A literal interpretation of this assumption is that the distribution of demand parameters is discrete but with a very large number of mass points. In the consumer heterogeneity literature (c.f. Allenby et al 1999), continuous models of heterogeneity have gained favor over models with a small number of mass points. The distinction between continuous models of heterogeneity and discrete models with a very large number of mass points is largely semantic. In fact, some non-parametric methods rely on discrete approximations. Our approach will be to specify a very flexible, but continuous model of heterogeneity and then exploit recent developments in Bayesian inference and computation to use draws from the posterior of this model as "representative" of the large number of consumer types. Each household in our data will be viewed as "representative" of a type. We will use MCMC methods to construct a Bayes estimate of each household's coefficient vector.

It is well known (c.f. Heckman 1981 and Keane 1997) that state dependence and heterogeneity can be confounded in the sense that mis-specified tightly parametric models of heterogeneity can lead to spurious findings of state dependence. The state-of-the-art in this literature (cf. Keane 1997 and Seetharaman et al. 1999) is to use normal models of heterogeneity. There is good reason to believe that there may be substantial departures from normality for the distribution of choice model parameters across households. For example, there may be sub-populations of households with different preferences for different brands. This might lead to multimodality in the distribution of the intercepts.

Our approach is to use a mixture of normals as the distribution of heterogeneity in a hierarchical Bayesian model. As with sufficient components in the mixture, we will be able to accommodate deviations from normality such as multi-modality, skewness, and fat tails. Let θ^h be the vector of choice model parameters for household *h*. The mixture of normals model specifies the distribution of θ^h across households as follows:

$$\begin{aligned} &\theta^h \sim N\left(\mu_{ind}, \Sigma_{ind}\right) \\ &ind \sim \text{multinomial}\left(\pi\right) \end{aligned}$$

π is a vector giving the mixture probabilities for each the K components. We complete the model specification with priors over the mixture probabilities and the mean and covariance matrices:

$$
\begin{aligned}
&\pi \sim Dirichlet(\alpha) \\
&\mu_k | \Sigma_k \sim N\left(\bar{\mu}, \Sigma_k \times a_\mu^{-1}\right) \\
&\Sigma_k \sim IW(\upsilon, V) \\
&\{\mu_k, \Sigma_k\} \text{independent}
\end{aligned}
$$

We implement posterior inference for the mixture of normals model of heterogeneity and the multinomial logit base model along the lines of Rossi et al. (2005). We use a hybrid Metropolis method that uses customized Metropolis candidate densities for each household. Conditional on the draws of θ^h, we use an unconstrained Gibbs sampler. Since our goal is to estimate the distribution of model parameters over households, we do not have to impose constraints on this Gibbs sampler to ensure identification. The density of model parameters is identified even if there is label switching.[4] Moreover, it has been noted (Frühwirth-Schnatter (2001)) that the unconstrained Gibbs sampler has superior mixing properties relative to Gibbs Samplers that are constrained in hopes of achieving identification of each component parameters.

Our MCMC algorithm will provide draws of the mixture probabilities as well as the normal component parameters. Thus, each MCMC draw of the mixture parameters provides a draw of the entire multivariate density of household parameters. We can average these densities to provide a Bayes estimate of the household parameter density. We can also

[4] In mixture models, there is a generic identification problem which has been dubbed "label switching." That is, the likelihood is unchanged if the labels for components are interchanged. This is only a problem if inference is desired for the mixture component parameters. In our application, we are interested in estimating individual household parameters and the distribution of parameters across households. These quantities are identified even in the presence of label switching.

construct Bayesian credibility regions for any given density ordinate to gauge the level of uncertainty in the estimation of the household distribution.

Some might argue that you do not have a truly non-parametric method unless you can claim that your procedure consistently recovers the true density of parameters in the population of all possible households. In the mixture of normals model, this requires that the number of mixture components (K) increase with the sample size. There are several ways to achieve this. One could put a prior over models with differing numbers of mixture components and use a reversible jump MCMC algorithm to navigate this space of models. However, there are no reliable reversible jump MCMC methods for multivariate mixtures of normals. Our approach is to fit models with successively larger numbers of components and gauge the adequacy of the number components by examining the fitted density as well as the Bayes factor associated with each number of components. What is important to note is that our improved MCMC algorithm is capable of fitting models with a large number of components at relatively low computational cost.

Description of the Data

For our empirical analysis, we estimate the logit demand model described above using household panel data containing all purchase behavior for the refrigerated orange juice and the 16 oz tub margarine categories. The panel data were collected by AC Nielsen for 2,100 households in a large Midwestern city between 1993 and 1995. In each category, we focus only on those households that purchase a brand at least twice during our sample period. Hence we use 354 households to estimate orange juice, and 444 households to estimate margarine demand. Table 4 lists the products considered in each category as well as the purchase incidence, product shares and average retail and wholesale prices. Over 85 per cent of the trips to the store recorded in our panel data do not involve purchases in the product category. This means that the outside good share is very large as is typical in many product categories and analyses of scanner data. In addition, households who adopt a pattern of purchasing a product on a regular cycle will be perceived as relatively price insensitive as the changes in price of the category relative to the outside good will have little influence on purchase incidence for these households.

In our econometric specification, we have been careful to control for heterogeneity as flexibly as possible to avoid confounding state dependence with unobserved heterogeneity. Even with these controls in place, it is still important to ask which patterns in our consumer shopping panel give rise to the identification of a "switching cost." Table 5 indicates that for each of the brands in the two categories, the marginal purchase probability is considerably smaller than the re-purchase probability. While this evidence is consistent with state dependence, it could also be a reflection of heterogeneity in consumer tastes for brands. Identification of state dependence in our context relies on the frequent temporary price changes typically observed in supermarket scanner data. If there is sufficient price variation, we will observe consumers switching away from their preferred products. The detection of state dependence relies on spells during which the consumer purchases these less-preferred alternatives on successive visits, even after prices return to their "typical" levels.

Demand Estimates

We now report the empirical estimates of demand from the orange juice and margarine data. In Table 6, we report the log-marginal density for several alternative model specifications and for each category. The posterior probability of a model specification is monotone in the log-marginal density, so that by choosing the model with the largest log marginal density we are picking the model with the highest posterior probability. It should be noted that the log-marginal density includes an automatic penalty for adding additional parameters (c.f. Rossi et al. 2005). By comparing models with and without switching costs and with varying degrees of heterogeneity, we can assess the importance of incorporating switching costs and non-normality. We assess the non-normality of the distribution of heterogeneity by comparing the log-marginal density for mixture models with varying numbers of components.

The results in Table 6 indicate several important features of the model. First, heterogeneity clearly leads to a substantial improvement in fit in both categories. Adding a switching cost term to the model also leads to an improvement in fit, albeit smaller. However, the usual state dependence specification appears to generate a better fit than the pure switching costs specification. These results confirm the well-established belief that consumer demand for frequently-purchased CPG products exhibits state dependence. For the remainder of this section, we will focus on the results from the state dependence specification. In the next section, we will contrast the equilibrium implications of state dependence versus switching costs to assess whether the choice of specification alters the substantive predictions we make for pricing.

An interesting finding is the extent to which flexibility in the heterogeneity distribution may be required to "fit" the data. In the orange juice category, a model with a single mixing component (the usual normal random coefficients model) performs relatively well. However, in the margarine category, we observe considerable improvement in fit by adding more components to the mixture. The improved fit from including five components in the mixture confirms the non-normality of the distribution of tastes in this category.

We now examine the model estimates to assess the non-normality of the fitted distributions of taste parameters. Ultimately, our goal is to estimate the distribution of tastes across households, not to attach any meaning or substantive significance to the parameters of the mixture components. Rather than report parameter estimates for the moments of each of the normal components, we instead plot the fitted marginal densities for several taste coefficients.

In Figures 3 and 4, we plot several fitted densities from the 1, 2 and 5 component mixture models for the margarine data under the state dependence specification. We also report the 95% posterior credibility region for the 5-component mixture model. This region provides point-wise evidence for the non-normality of the population marginal density for a given coefficient. Figure 3 provides compelling evidence of the need for a flexible model capable of addressing non-normality. In the upper panel, the Shedd's brand intercepts from the 5-component model exhibit bimodality that cannot be captured by the 1 or 2 component models. The bimodality implies that there are households who differ markedly in their quality perceptions for margarines (note: the outside good is purchased most often so that the intercepts for all margarine brands are typically negative). In general, the results suggest that one would recover a very misleading description of the data-generating process if the usual symmetric normal (1-component) prior were used to fit these data.

In Figure 4, the price coefficient (upper panel) for the 5-component model leads to a slightly asymmetric density with fat tails. In contrast, a symmetric 1-component model has both a mode and tails lying outside the credibility region for the 5-component model. For the state dependence estimates (lower panel), the 1 component model has a higher mean and thinner tails than the 5-component model.

In Figure 5, we report fitted densities from the orange juice category. These plots illustrate why we do not get the same improvement from more mixing components as we did in the margarine data. In the upper plot (96 oz MM), the marginal densities from the 1 and 2 component models are completely contained within the credibility region around the 5-component model. For the orange juice data, the one component normal approximation seems adequate.

Figures 6 and 7 display the fitted densities of the state dependence premium (i.e. switching costs) in dollar terms for each category. The inclusion of the outside option in the model enables us to assign money-metric values to our model parameters simply by re-scaling them by the price parameter (i.e. the marginal utility of income). For the switching cost parameter reported in the figures, this ratio represents the dollar cost foregone when a consumer switches to another brand than the one purchased previously. In the graphs, the point-estimate of switching costs from the homogeneous logit specification is denoted by a vertical red line.

Figures 6 and 7 display an entire distribution of switching costs across the population of households. Some of the values on which this distribution puts substantial mass are rather large values, others are small. To provide some sense of the magnitudes of these values, we compute the ratio of the dollar switching cost to the average price of the products. The ratio of the mean dollar switching cost to average price is 0.13 for margarine and 0.19 for orange juice. It should be emphasized that the entire distribution of switching costs will be used in computation of equilibrium prices. The distribution of dollar switching costs puts mass on some very large values. For example, the ratio of the 95th percentile of dollar switching costs to average prices is 0.61 for margarine and 0.60 for orange juice. In the computations in Section 4 below, we will use this distribution of switching costs as the center point. We will also explore magnifying this distribution by scaling it by a factor of 4.

4. Pricing Implications of the Demand Estimates

In this section, we use the estimated demand systems to explore the implications of switching costs for pricing. For each of the categories, we compute the steady-state Markov perfect equilibrium prices corresponding to the demand estimates. We then examine the sensitivity of these steady-state price levels to specific parameter values.

To compute prices, we need to simplify the demand estimates to reduce the dimension of the state space of the model to a feasible range. For the orange juice data with 355 consumer "types" and 6 products, one would literally need to solve a dynamic programming problem with a $355 \cdot 5 = 1,775$ dimensional state space. We simplify the problem as follows. For the orange juice category, we focus only on 64 oz Tropicana and Minute Maid. We also take each household's posterior mean taste vector and cluster them into 5 consumer "types." Then our state space is $5 \cdot 1 = 5$ dimensional. Similarly, in margarine we focus on all 4 products, and we cluster consumers into 2 "types." This clustering reduces the state space to $2 \cdot 3 = 6$ dimensions. Results from the clustering are reported in Table 7 for each of the categories. While these simplifications eliminate some of the richness of the true product category, they should not detract from our main objective, which is to examine the pricing implications of the estimated switching costs.

We begin with the pricing results for the refrigerated orange juice category. In Figure 8, we report the optimal pricing policy functions for the 64 oz Minute Maid and Tropicana brands respectively. The pricing policy is shown as a function of the loyalty states of two consumer segments holding the loyalty levels of the other three segments constant.

Optimal Minute Maid prices are rising with the fraction of loyals to Minute Maid whereas Tropicana prices are falling with the fraction of Minute Maid loyals. Not surprisingly, prices are barely affected by the loyalty of consumers in cluster two, who—despite having a fairly high loyalty premium in dollars—represent a small share among all customers (3%). These policy functions are then used to compute the steady-state price levels by simulating 10,000 weeks of competition. Figure 9 shows the convergence of prices and states to the steady state from a starting point where every consumer is loyal to Tropicana. Although not reported, we obtained convergence to the same steady state from any randomly chosen starting value.

In Table 8, we report our results relating steady state price and profits levels to the magnitude of the state dependence premium/switching costs. We compute equilibrium prices for a range of switching costs achieved by scaling the distribution of cluster parameters. That is, we multiply the loyalty or state dependence parameter in each cluster by a scale factor reported in the left margin of Table 8. We see that prices decline as state dependence increases from the zero order case of zero state dependence. We are able to compute equilibrium prices for not just the level of state dependence found in our data but for much higher levels corresponding to scale factors greater than one. We find that even with state dependence levels twice that revealed in out data, equilibrium prices are lower in the presence of state dependence. At scale factors of 3, OJ prices start to rise above the zero state dependence levels. State dependence is a source of additional profits to the firms as consumer utility increases for any fixed level of price.

We also compute the steady-state prices and profits for the pure switching costs model estimates in Table 9. In both the margarine and orange juice categories, prices fall as switching costs increase away from zero. By comparing the prices for scale factors of zero and one, we see that the estimated level of switching costs in this data result in a fall in equilibrium prices. As expected, this fall is larger in the margarine category than in the orange juice category as the estimated dollar value of switching costs relative to prices is smaller in the orange juice category. Even large scale factors of 4 do not reverse this finding. We still observe equilibrium prices below the levels without switching costs. The bottom half of Table 9 provides equilibrium profit calculations and show that firms are worse off with switching costs and without.

6. Conclusions

We have demonstrated that equilibrium prices fall as switching costs increase for a variety of stylized and more realistic models. This finding holds for a wide range of switching costs or state dependence centered on those obtained from consumer panel data. Very high levels of switching costs must prevail in order to obtain results similar to those conjectured by Klemperer, i.e. that switching costs make markets less competitive and provide a source of economic rent. Our switching cost estimates are based on consumer panel data for two categories of consumer products, margarine and orange juice. These switching costs are important from a statistical point of view in the sense that models with switching costs account for observed behavior better than those without. Our switching costs distribution puts mass on switching costs in the range of 15 to 60 per cent of purchase price. In addition, we have scaled this distribution up by a factor of four and still observe lower prices with switching costs. This means that our basic result applies to situations where switching costs are as much as four times purchase price. We would argue that many classic examples of switching costs such as cellular service carriers or airline frequent flyer programs have associated switching costs in this region.

Our results can be reversed if switching costs reach very high levels or if, indeed, they are infinite as assumed in Beggs and Klemperer. In a world with the levels of switching costs envisaged by much of the theoretical literature, we would not see consumers switching brands very often. The empirical fact that consumers are observed to switch brands in many product categories implies that our results are relevant in many situations.

References

Allenby, G. and P. E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics* 89, 57-78.

Anderson, E.T., N. Kumar and S. Rajiv (2004), "A Comment on: 'Revisiting dynamic duopoly with consumer switching costs,'" *Journal of Economic Theory*, 116, 177-186.

Beggs, A. and P. Klemperer (1992), "Multi-Period Competition with Switching Costs," *Econometrica* 60, 651-666.

Besanko, D., U. Doraszelski, Y. Kryukov, and M. Satterthwaite (2005), "Learning-by-Doing, Organizational Forgetting, and Industry Dynamics," manuscript, Harvard University.

Cabral, L.M.B. and M. Villas-Boas (2005), "Bertrand Supertraps," *Management Science*, 51, 599-613.

Doganoglu, T. (2005), "Switching Costs, Experience Goods and Dynamic Price Competition," Working Paper, University of Munich.

Doraszelski, U. and M. Satterthwaite (2005), "Foundations of Markov-Perfect Industry Dynamics: Existence, Purification, and Multiplicity," working paper, Harvard University.

Dubé, J. P., G. Hitsch, P. E. Rossi, M. Vitorino (2006), "Category Management with State Dependent Utility," working paper, University of Chicago.

Erdem, T. (1996), "A Dynamic Analysis of Market Structure Based on Panel Data." *Marketing Science*, 15, 359-378.

Farrell, J. and P. Klemperer (2004), "Coordination and Lock-In: Competition with Switching Costs and Network Effects," forthcoming, Handbook of Industrial Organization.

J. Farrell and C. Shapiro (1988), Dynamic competition with switching costs, *RAND Journal of Economics*, 19, 123–137.

Früwirth-Schnatter, S. (2001), "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models," *JASA* 96, 194-209.

Heckman, J. J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating A Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence" in C. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data With Econometric Applications*, Cambridge: MIT Press.

Judd, K. L. (1998), *Numerical Methods in Economics*, Cambridge: MIT Press.

Keane, M.P. (1997), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics*, 15, 310-27.

Klemperer, P. (1987), "The Competitiveness of Markets with Switching Costs," *Rand Journal of Economics* 18, 138-150.

Klemperer, Paul (1995), "Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade," *Review of Economic Studies*, 62 (4), 515-539

Padilla, A. J. (1992), "Revisiting Dynamic Duopoly with Consumer Switching Costs," *JET*, 67, 520-530.

Rossi, P. E., G. M. Allenby and R. McCulloch (2005), *Bayesian Statistics and Marketing*, New York: John Wiley and Sons.

Roy, R., P. K. Chintagunta, and S. Haldar (1996). "A Framework for Investigating Habits, `The Hand of the Past,' and Heterogeneity in Dynamic Brand Choice," *Marketing Science*, 15 (3), 280-299.

Rust, J. (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica,* 55 (5), 999-1033.

Seetharaman, P.B., A. Ainslie and P.K. Chintagunta (1999), "Investigating Household State Dependence Effects Across Categories," *Journal of Marketing Research*, 36, 4, 488-500.

Seetharaman, P. B. (2004), "Modeling Multiple Sources of State Dependence in Random Utility Models: a distributed lag approach," *Marketing Science* 23, 263-271.

Shi, M., J. Chiang, and B. Rhee (2006), "Price Competition with Reduced Consumer Switching Costs: The Case of "Wireless Number Portability" in the Cellular Phone Industry," *Management Science*.

Shum, M. (2004), "Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast Cereals Market," *Journal of Economics and Management Strategy*, 13, 241-71.

Stango, V. (2002), "Pricing with Consumer Switching Costs: evidence from the credit card market," *The Journal of Industrial Economics*, 50, 475-92.

Stokey, N. L. and R. E. Lucas (1989), *Recursive Methods in Economics,* Cambridge: Harvard University Press.

Viard, B. (2003), "Do Switching Costs Make Markets More or Less Competitive?: the case of 800-number Portability," working paper, Stanford University.

Viard, B. (2006), "Do Switching Costs Make Markets More or Less Competitive?: the case of 800-number Portability," forthcoming *Rand Journal of Economics*.

Wernerfelt, B. (1991), "Brand Loyalty and Market Equilibrium," *Marketing Science*, 10(3), 229-245.

Whitt, W. (1980), "Representation and Approximation of Non-cooperative Sequential Games," *SIAM Journal of Control and Optimization* 18, 33-48.

Appendix A. Existence of Equilibrium in the Simple Model

The existence of a Markov perfect equilibrium in our model follows from arguments given in Whitt (1980) and Doraszelski and Satterthwaite (2005). In order to show that the equilibrium is in pure strategies, we need to show that the best-reply correspondence is single-valued. Our strategy is to show that in the case of one consumer with logit demand, the right-hand side of the Bellman equation is strictly quasi-concave, and hence has a unique maximizer. This strategy has been employed previously by Besanko et al. (1995).

Recall the Bellman equation in the simple model:

$$V_j(s) = \max_{p_j \geq 0} \left\{ \pi_j(s,p) + \beta \left(\sum_{k=1}^{J} P_k(s,p) V_j(k) + P_0(s,p) V_j(s) \right) \right\} \qquad \forall s \in X.$$

Denote the right-hand side of this functional equation by $\Psi_j(s, p_j, p_{-j})$, such that

$$V_j(s) = \max_{p_j \geq 0} \Psi_j(s, p_j, p_{-j}), \qquad \forall s \in X.$$

This maximization problem has the following first-order condition:

$$\begin{aligned} \frac{\partial \Psi_j}{\partial p_j} &= \alpha P_j (1 - P_j)(p_j - c_j) + P_j + \beta \left(\sum_{k=1}^{J} (-\alpha) P_j P_k V_j(k) + \alpha P_j V_j(j) - \alpha P_0 P_j V_j(s) \right) \\ &= \alpha P_j \left(-\Psi_j + (p_j - c_j) + \frac{1}{\alpha} + \beta V_j(j) \right). \end{aligned}$$

Here, P_j is shorthand for $P_j(s, p_j, p_{-j})$. Evaluating the second-order condition at a price where $\partial \Psi_j / \partial p_j = 0$, we find that

$$\begin{aligned} \frac{\partial^2 \Psi_j}{\partial p_j^2} &= \alpha^2 P_j (1 - P_j) \left(-\Psi_j + (p_j - c_j) + \frac{1}{\alpha} + \beta V_j(j) \right) + \alpha P_j \left(-\frac{\partial \Psi_j}{\partial p_j} + 1 \right) \\ &= \alpha (1 - P_j) \frac{\partial \Psi_j}{\partial p_j} + \alpha P_j \left(-\frac{\partial \Psi_j}{\partial p_j} + 1 \right) \\ &= \alpha P_j < 0. \end{aligned}$$

Hence, Ψ_j is strictly quasi-concave in p_j, and hence it follows that there is a unique price that maximizes the right-hand side of the Bellman equation for any state *s* and price profile $p_{-j} = \sigma_{-j}(s)$.

Appendix B. Forward-Looking Consumers

We now extend the model to allow for forward-looking consumers who anticipate the consequences of becoming loyal to product *j*. In general, the presence of forward-looking consumers can complicate the computation of an equilibrium. For example, Anderson, Kumar and Rajiv (2004) show the equilibrium proposed by Padilla (1995) does not in fact constitute a Markov perfect equilibrium under forward-looking consumer behavior.

As before, the current-period utility from choosing product *j* is $U_j = U(j,s,p) + \lambda\varepsilon_j$. But, now consumers maximize the PDV of current and future utilities. For simplicity, we assume that consumers discount future utilities at the same rate as firms, β. Define the state transition function $s' = \phi(s,j) = j$ if $j \neq 0$ and $s' = \phi(s,0) = s$. The value function of the consumer given state *s* and idiosyncratic utility draws $\varepsilon = (\varepsilon_0,...,\varepsilon_J)$ is

$$\upsilon(s,\varepsilon) = \max_{j=0,...,J} \left\{ U(s,\sigma(s),j) + \lambda\varepsilon_j + \beta\int \upsilon(\phi(s,j),\varepsilon')f(\varepsilon')d\varepsilon' \right\} \qquad \text{(A.1)}$$

Note that this value function depends on the consumer's expectation that the firms choose prices according to $p_j = \sigma_j(s)$. Following arguments given in Rust (1987), the consumer's decision problem can be reformulated in the following way. Let the expected future value from choosing alternative *j* in state *s* be

$$W(s,j)=\int \max_{k=0,\dots,J}\left\{U(s',\sigma(s'),k)+\lambda\varepsilon_k+\beta W(s',k)\right\}f\left(\varepsilon\right)d\varepsilon,$$

where $s'=\phi(s,j)$. Since ε has the Type I extreme value distribution, $W(s,j)$ has the closed form expression

$$W\left(s,j\right)=\lambda\left(\gamma+\log\left[\sum_{k=0}^{J}\exp\left(\frac{1}{\lambda}\left(U(s',\sigma(s'),k)+\beta W(s',k)\right)\right)\right]\right). \tag{A.2}$$

Here, $\gamma\approx 0.57722$ is Euler's constant. The consumer then chooses the alternative $j=0,\dots,J$ that yields the highest utility index

$$U(s,\sigma(s),j)+\beta W(s,\phi(s,j))+\lambda\varepsilon_j.$$

Conditional on the consumer's choice behavior, which is now also described by the consumer's value function, *W,* the firm's problem remains the same under forward-looking consumer behavior. A Markov perfect equilibrium now consists of pricing strategies and value functions for each firm *j* and the consumer's consumption strategy, which is fully described by the value function *W,* such that (i) each firm's pricing strategy is optimal given the consumer's strategy and given the competitors' strategies, and (ii) given the firms' pricing strategies, the consumers value function satisfies equation (A.2).

In Section 2, we explored the predictions of the simple model for the symmetric case with a symmetric equilibrium. In this case, myopic and forward-looking consumer behavior is identical. This can be seen from equation (A.2): *W* actually depends only on $s'=\phi(s,j)$, the product that the consumer is loyal to in the next period. Due to symmetry, the identity of this product does not matter. Therefore, *W* is exactly the same for all $s'\in X$, and therefore adds the same constant to each utility index. Thus, the choice probabilities are not affected by the presence of *W*.

Appendix C. Numerical Solution to the Dynamic Program

We use numerical methods to solve for the equilibrium of the pricing game. We first discretize each axis of the state space using a finite number of points, $0 < x_{i0} < x_{i1} < ... < x_{iL} = 1$. We then form a grid representing the whole state space from the Cartesian product of these points. For each point in the grid, we store the value and policy functions of each competitor in the computer memory. For states outside the grid, we calculate the value and policy functions using bilinear interpolation. To solve for the equilibrium, we employ the following algorithm, which is an adaptation of policy iteration applied to the case of the games: start with some initial guess of the strategy profile, $\sigma^0 = (\sigma_1^0, ..., \sigma_J^0)$, and then proceed along the following steps:

1. For the strategy profile σ^n, calculate the corresponding value functions for each of the J firms. These value function are defined by the Bellman equation <equation reference>, where the right hand side of the Bellman equation is not maximized, but instead evaluated using the current strategy profile σ^n.
2. If $n > 0$, check whether the value functions and policy functions satisfy the convergence criteria, $\left\| V_j^n - V_j^{n-1} \right\| < \varepsilon_V$ and $\left\| \sigma_j^n - \sigma_j^{n-1} \right\| < \varepsilon_\sigma$ for all firms j. If so, stop.

Update each firm's strategy using the Bellman equation <reference>. In contrast to step 1, the maximization on the right hand side is now carried out. Denote the resulting new policies and value functions by σ_j^{n+1} and V_j^{n+1}, and return to step 1.

Table 1
Equilibrium prices under different switching cost levels

Switching Cost	p_1	p_2	p^a	Purchase Prob. 1	Purchase Prob. 2	Prob. Stay Loyal
0.00	1.808	1.808	1.808	0.236	0.236	0.764
0.25	1.825	1.679	1.756	0.272	0.245	0.755
0.50	1.840	1.542	1.706	0.310	0.254	0.746
0.75	1.855	1.397	1.659	0.350	0.261	0.739
1.00	1.867	1.244	1.613	0.390	0.268	0.732
1.25	1.877	1.083	1.569	0.431	0.273	0.727
1.50	1.883	0.915	1.525	0.470	0.276	0.724
1.75	1.887	0.741	1.482	0.508	0.278	0.722
2.00	1.887	0.560	1.438	0.544	0.278	0.723
3.00	1.931	0.000	1.437	0.681	0.234	0.766
4.00	2.153	0.000	1.860	0.823	0.130	0.870
5.00	2.335	0.000	2.183	0.913	0.064	0.936
6.00	2.475	0.000	2.404	0.961	0.028	0.972
7.00	2.616	0.000	2.584	0.983	0.012	0.988
8.00	2.829	0.000	2.813	0.992	0.006	0.994

Note: The results were calculated for product intercepts = 1.0, price coefficient = 1.0, and mean outside good utility = 0.0. The discount factor is $\beta = 0.998$. The table shows the prices of firm 1 and firm 2 in state 1, and the average transaction price paid by the customer. The table also shows the purchase probabilities for the products in state 1, and the probability that the customer stays loyal.

Table 2
Equilibrium prices versus random utility scale factors

	Lambda (Random Utility Component Scale Factor)											
	0.2			0.4			0.6			1.0		
Switching Cost	p_1	p_2	p^a	p_1	p_2	p^a	p_1	p_2	p^a	p_1	p_2	p^a
0.00	0.84	0.84	0.84	1.08	1.08	1.08	1.32	1.32	1.32	1.81	1.81	1.81
0.25	0.82	0.65	0.75	1.09	0.93	1.02	1.34	1.19	1.27	1.83	1.68	1.76
0.50	0.79	0.45	0.68	1.09	0.77	0.96	1.35	1.04	1.22	1.84	1.54	1.71
0.75	0.77	0.24	0.63	1.08	0.58	0.91	1.35	0.87	1.17	1.86	1.40	1.66
1.00	0.75	0.02	0.60	1.07	0.38	0.85	1.35	0.70	1.12	1.87	1.24	1.61
1.25	0.82	0.00	0.73	1.05	0.18	0.81	1.35	0.51	1.07	1.88	1.08	1.57
1.50	0.87	0.00	0.84	1.04	0.00	0.79	1.34	0.32	1.02	1.88	0.92	1.53
1.75	0.91	0.00	0.90	1.11	0.00	0.92	1.33	0.12	0.98	1.89	0.74	1.48
2.00	0.96	0.00	0.96	1.17	0.00	1.04	1.34	0.00	1.01	1.89	0.56	1.44
3.00	1.53	0.00	1.53	1.33	0.00	1.31	1.57	0.00	1.44	1.93	0.00	1.44
4.00	2.36	0.00	2.36	1.61	0.00	1.61	1.71	0.00	1.67	2.15	0.00	1.86
5.00	3.26	0.00	3.26	2.19	0.00	2.19	1.90	0.00	1.89	2.34	0.00	2.18
6.00	4.20	0.00	4.20	2.95	0.00	2.95	2.28	0.00	2.28	2.48	0.00	2.40
7.00	5.15	0.00	5.15	3.78	0.00	3.78	2.87	0.00	2.87	2.62	0.00	2.58
8.00	6.11	0.00	6.11	4.66	0.00	4.66	3.58	0.00	3.58	2.83	0.00	2.81

Note: The results were calculated for product intercepts = 1.0, price coefficient = 1.0, and mean outside good utility = 0.0. The discount factor is $\beta = 0.998$. The table shows the prices of firm 1 and firm 2 in state 1, and the average transaction price paid by the customer.

Table 3
Equilibrium prices in the OLG model

Switching Cost	Lambda (Random Utility Component Scale Factor) 0.2		0.4		0.6		0.8		1.0	
	p^a_{young}	p^a_{old}	p^a_{young}	p^a_{old}	p^a_{young}	p^a_{old}	p^a_{young}	p^a_{old}	p^a_{young}	p^a_{old}
0.00	0.84	0.84	1.08	1.08	1.32	1.32	1.56	1.56	1.81	1.81
0.25	0.80	0.82	1.05	1.06	1.29	1.30	1.54	1.54	1.78	1.78
0.50	0.76	0.86	1.02	1.05	1.27	1.28	1.51	1.52	1.76	1.76
0.75	0.69	0.88	0.99	1.06	1.24	1.28	1.49	1.51	1.74	1.75
1.00	0.62	0.90	0.95	1.08	1.21	1.28	1.46	1.51	1.71	1.74
1.25	0.54	0.93	0.89	1.11	1.18	1.30	1.43	1.51	1.69	1.74
1.50	0.47	0.97	0.83	1.13	1.13	1.31	1.40	1.52	1.66	1.74
1.75	0.39	1.02	0.76	1.16	1.08	1.33	1.37	1.53	1.63	1.74
2.00	0.31	1.08	0.70	1.18	1.03	1.35	1.32	1.54	1.59	1.75
3.00	0.07	1.21	0.42	1.34	0.77	1.46	1.10	1.62	1.41	1.80
4.00	0.00	1.30	0.17	1.55	0.50	1.59	0.85	1.69	1.19	1.83
5.00	0.00	1.31	0.13	1.57	0.31	1.73	0.66	1.76	1.01	1.85
6.00	0.00	1.31	0.13	1.58	0.30	1.75	0.56	1.82	0.90	1.89
7.00	0.00	1.31	0.13	1.58	0.29	1.76	0.51	1.84	0.83	1.91
8.00	0.00	1.31	0.13	1.58	0.29	1.76	0.50	1.85	0.81	1.92

Note: The results were calculated for product intercepts = 1.0, price coefficient = 1.0, and mean outside good utility = 0.0. The customers are forward-looking, and discount the future at the same rate as the firm ($\beta = 0.998$). The table shows the average transaction prices paid by the "young" and the "old" customer.

Table 4
Description of Data

Refrigerated Orange Juice

Product	Retail Price	Wholesale Price	% trips
64 oz MM	2.21	1.36	1.52
premium 64oz MM	2.62	1.88	0.96
96 oz MM	3.41	2.12	2.01
premium 64oz TR	2.73	2.07	3.96
64 oz TR	2.26	1.29	0.93
premium 96 oz TR	4.27	2.73	1.09
no-purchase (% trips)	89.53		
# households	355		
# trips per household	89.51		
# purchases per household	9.37		

Margarine

Product	Retail Price	Wholesale Price	% trips
Promise	1.69	1.22	2.93
Parkay	1.63	1.02	1.11
Shedd's	1.07	0.83	2.83
ICBINB	1.55	1.11	5.26
no-purchase (% trips)	87.86		
# households	429		
# trips per household	81.02		
# purchases per household	9.89		

Table 5
Purchase versus re-purchase rates

Category	Refrigerated Orange Juice		Tub Margarine			
Brand	Minute Maid	Tropicana	Promise	Parkay	Shedd's	ICBINB
Sample purchase frequencies	0.449	0.551	0.241	0.091	0.233	0.433
Sample re-purchase frequencies	0.841	0.837	0.735	0.83	0.719	0.8001

Table 6
Fit and the Role of Heterogeneity and State-dependence

		Log Marginal Density	
Model	**# Components**	**Margarine**	**Orange Juice**
w/o Switching Costs	No heterogeneity	-18302.52	-15888.78
	5-component	-13126.00	-11622.32
with Switching Costs	No heterogeneity	-16170.00	-14330.43
	1-component	-13237.49	-11503.13
	2-component	-13217.10	-11510.86
	5-component	-13107.69	-11505.56
	10-component	-13099.00	-11478.65
with State Dependence	No heterogeneity	-15735.04	-14330.43
	1-component	-13182.76	-13284.55
	2-component	-13155.41	-13250.65
	5-component	-13027.58	-13147.53
	10-component	-12994.88	-13147.33

Table 7
Clusters Used In Equilibrium Pricing Computations

Refrigerated Orange Juice

segment	64 oz MM	premium 64oz MM	96 oz MM	premium 64oz TR	64 oz TR	premium 96 oz TR	price	loyalty	loyalty ($)	size
1	-2.88	-2.57	-2.50	-0.25	-2.59	-0.31	-1.19	0.69	0.59	0.26
2	-2.62	-3.79	-1.79	-2.88	-3.72	-3.59	-0.91	1.23	1.36	0.25
3	-13.09	-12.20	-9.54	-1.22	-9.53	-3.19	-0.31	-0.03	-0.10	0.02
4	-0.37	0.32	0.01	1.53	-0.43	1.73	-2.08	0.23	0.11	0.18
5	-1.30	-1.59	-0.50	-0.71	-1.92	-0.82	-1.65	0.61	0.37	0.29

16 oz Tub Margarine

segment	Promise	Parkay	Shedd's	ICBINB	price	loyalty	loyalty ($)	size
1	-1.95	-3.47	-1.22	-2.67	-2.46	0.17	0.07	0.50
2	-2.88	-6.87	-6.49	-2.97	-0.87	0.19	0.22	0.50

Table 8
Equilibrium Prices and Profits for the State Dependence Model

Steady State Prices

Scale Factor	Prices					
	16-oz Tub Margarine				Refrigerated Orange Juice	
	Promise	Parkay	Shedd's	IBNINB	Minute Maid	Tropicana
0	1.44	0.91	0.82	1.42	1.51	1.79
1	1.42	0.91	0.81	1.39	1.49	1.77
2	1.41	0.91	0.81	1.37	1.48	1.79
3	1.39	0.91	0.81	1.35	1.50	1.84
4					1.54	1.94

Steady State per Period Profits

Scale Factor	Profits					
	16-oz Tub Margarine				Refrigerated Orange Juice	
	Promise	Parkay	Shedd's	IBNINB	Minute Maid	Tropicana
0	3.74	0.16	3.39	3.50	1.42	9.11
1	4.07	0.16	3.93	3.81	1.74	14.8
2	4.47	0.16	4.57	4.17	2.58	26.63
3	4.94	0.16	5.31	4.61	4.90	51.71
4					11.03	101.1

Table 9

Equilibrium Prices and Profits for the Pure Switching Cost Model

Steady State Prices

Scale Factor	Prices					
	16-oz Tub Margarine				Refrigerated Orange Juice	
	Promise	Parkay	Shedd's	IBNINB	Minute Maid	Tropicana
0	1.16	0.92	0.82	1.20	1.53	1.82
1	1.11	0.88	0.76	1.16	1.49	1.80
2	0.95	0.84	0.70	1.08	1.47	1.79
3	0.93	0.77	0.72	1.13	1.45	1.79
4	0.94	0.72	0.72	1.18	1.44	1.79

Steady State per Period Profits

Scale Factor	Profits					
	16-oz Tub Margarine				Refrigerated Orange Juice	
	Promise	Parkay	Shedd's	IBNINB	Minute Maid	Tropicana
0	4.23	0.38	2.90	10.09	1.66	15.48
1	2.31	0.18	1.91	8.08	1.33	15.01
2	1.56	0.08	1.29	6.88	1.12	14.68
3	1.36	0.05	1.13	6.84	0.98	14.45
4	1.31	0.03	1.08	7.01	0.89	14.29

Figure 1

Average transaction price paid for different switching cost levels

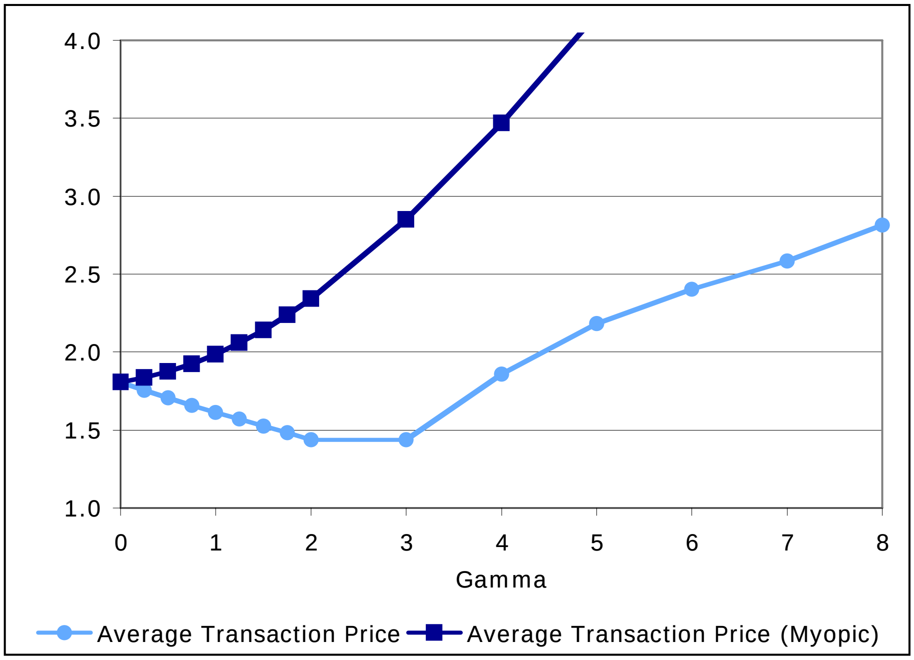


Figure 2
Transaction prices under the "pure switching cost" and "state dependence" versions of the model

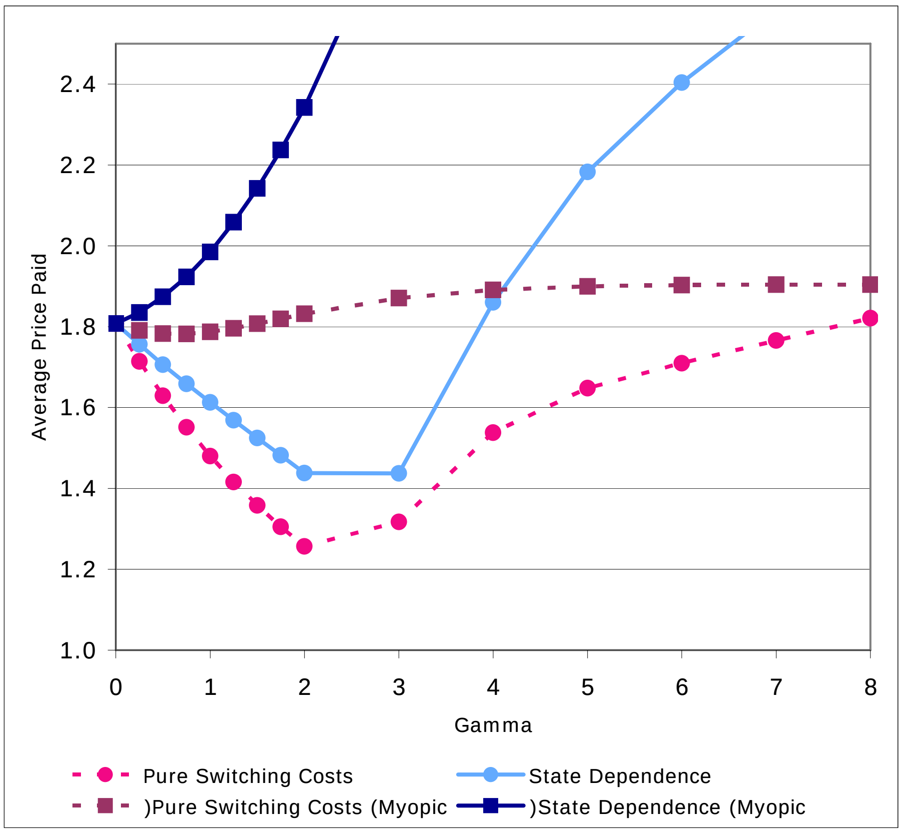


Figure 3

Fitted Densities for Shedd's and ICBINB Brand coefficients (Margarine)

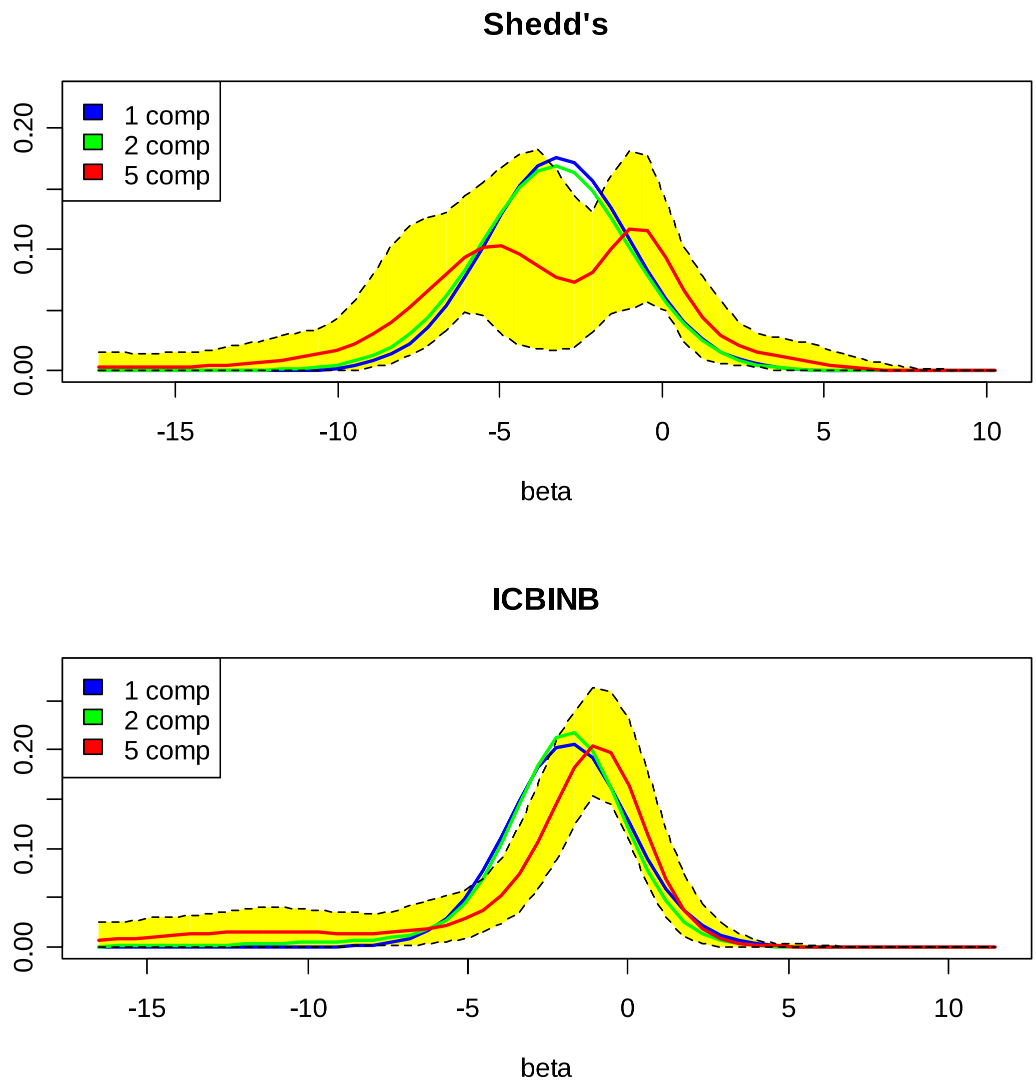


Figure 4
Fitted Densities for Price and State Dependence Coefficients (Margarine)

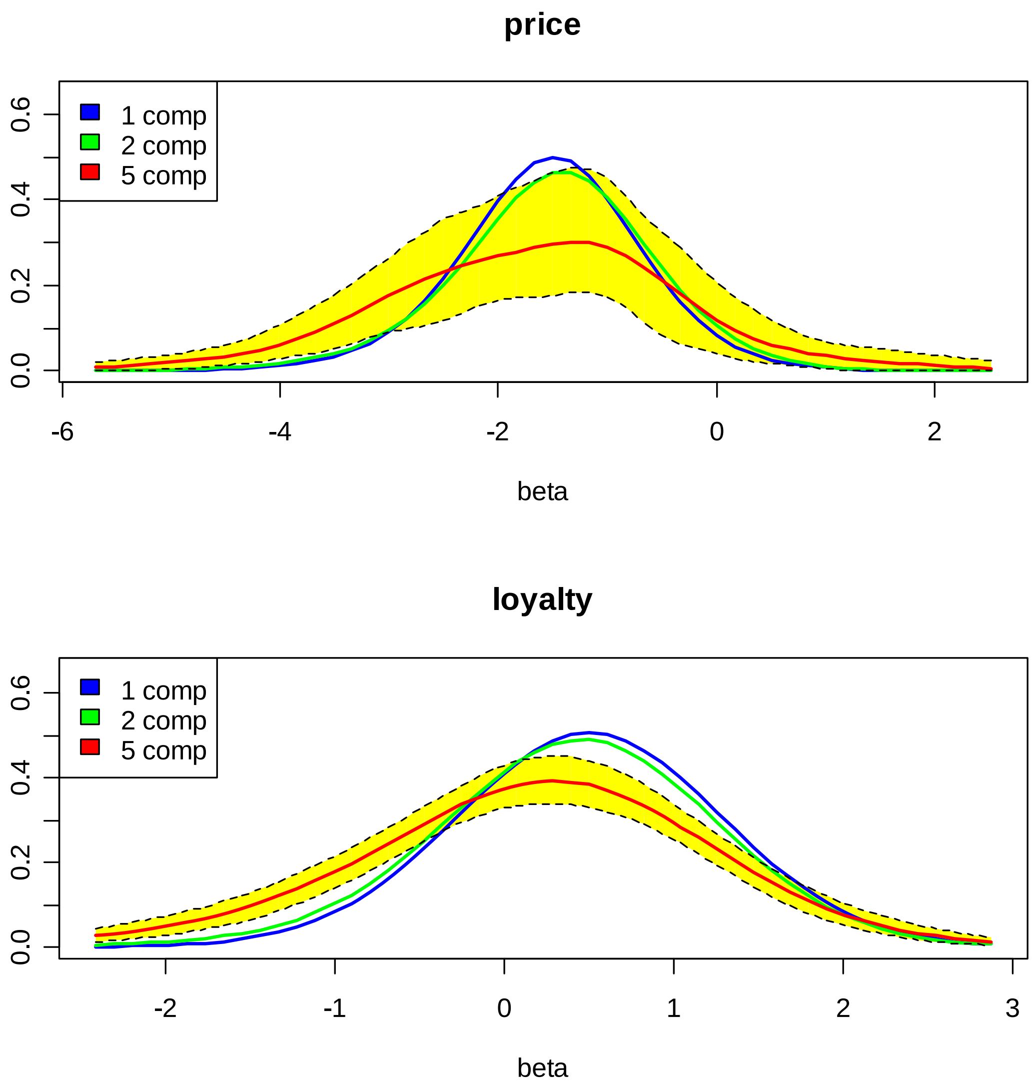
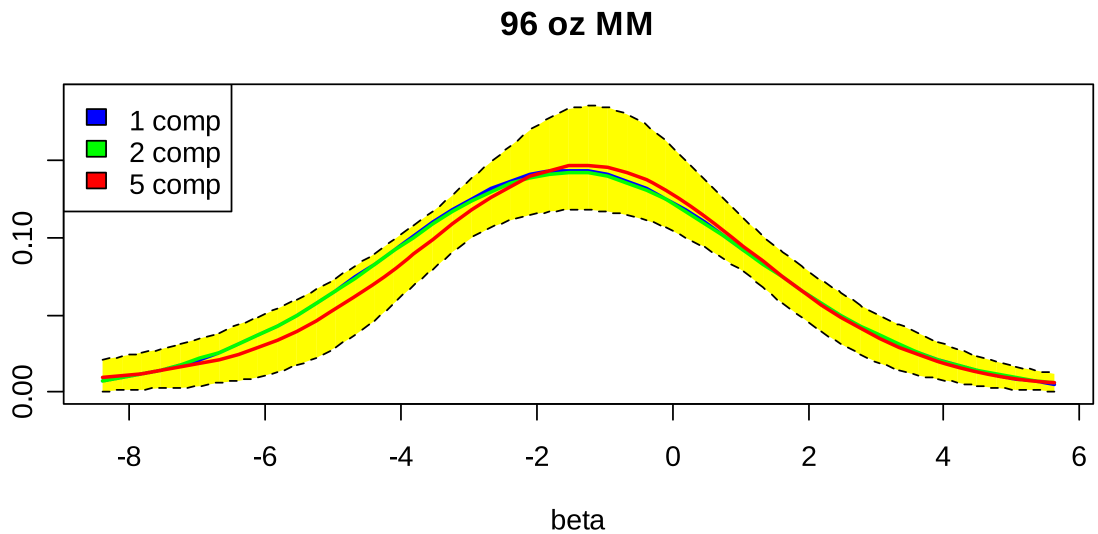


Figure 5
Fitted Densities for 96 oz Minute Maid and 64 oz Tropicana Brand coefficients (Orange Juice)



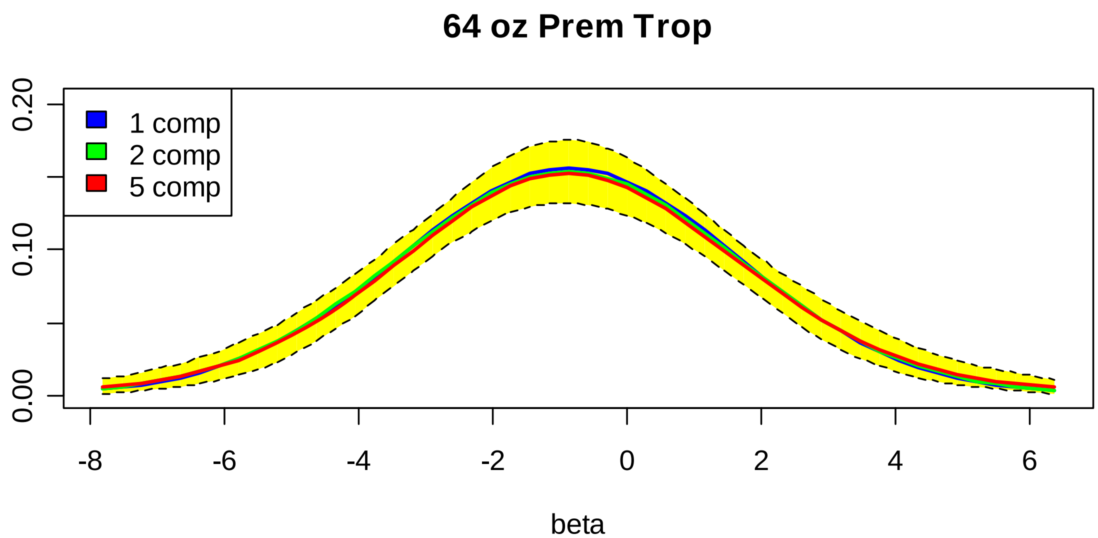


Figure 6

Fitted Densities and 95% Posterior Credibility Regions for the Money-metric State Dependence Premium in dollars (Margarine)

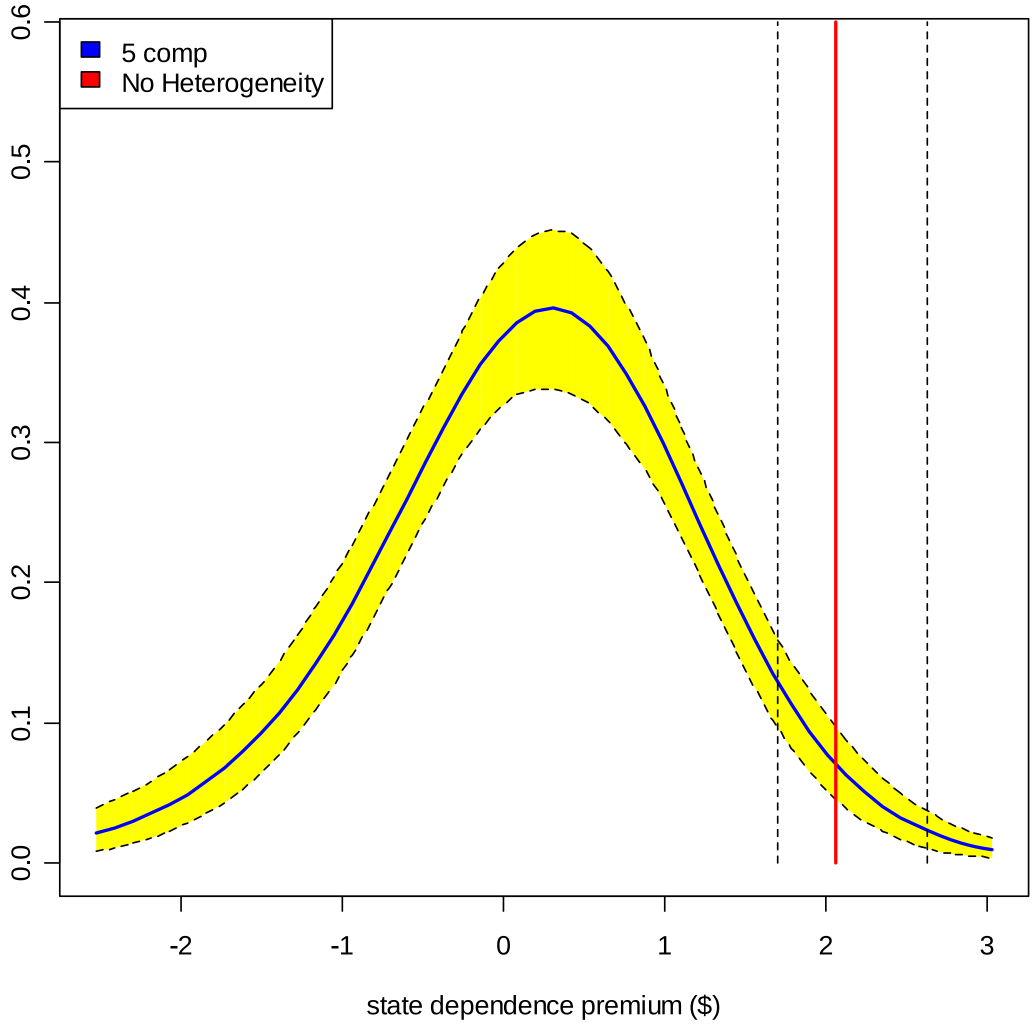


Figure 7
Fitted Densities and 95% Posterior Credibility Regions for the Money-metric State Dependence Premium in dollars (Orange Juice)

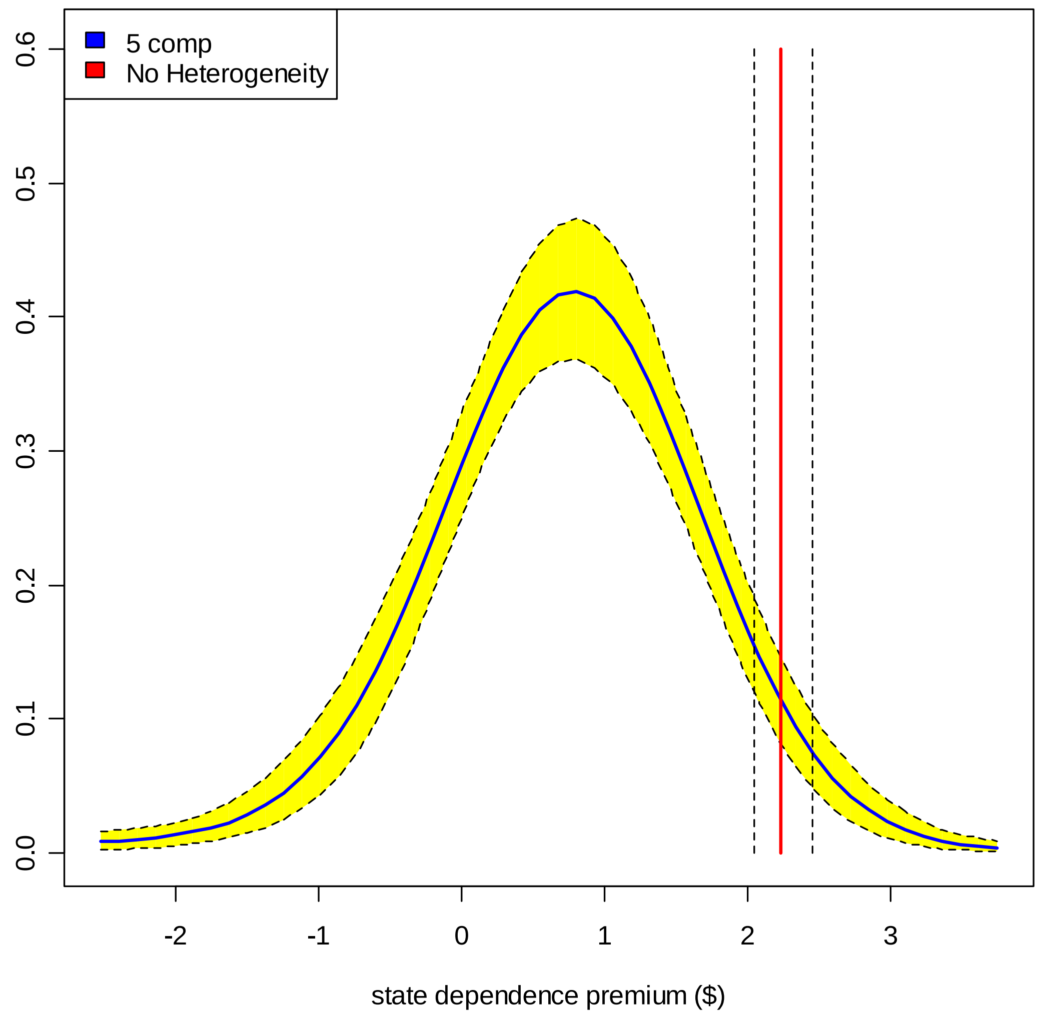


Figure 8
Optimal Pricing Policy Functions for Refrigerated Orange Juice

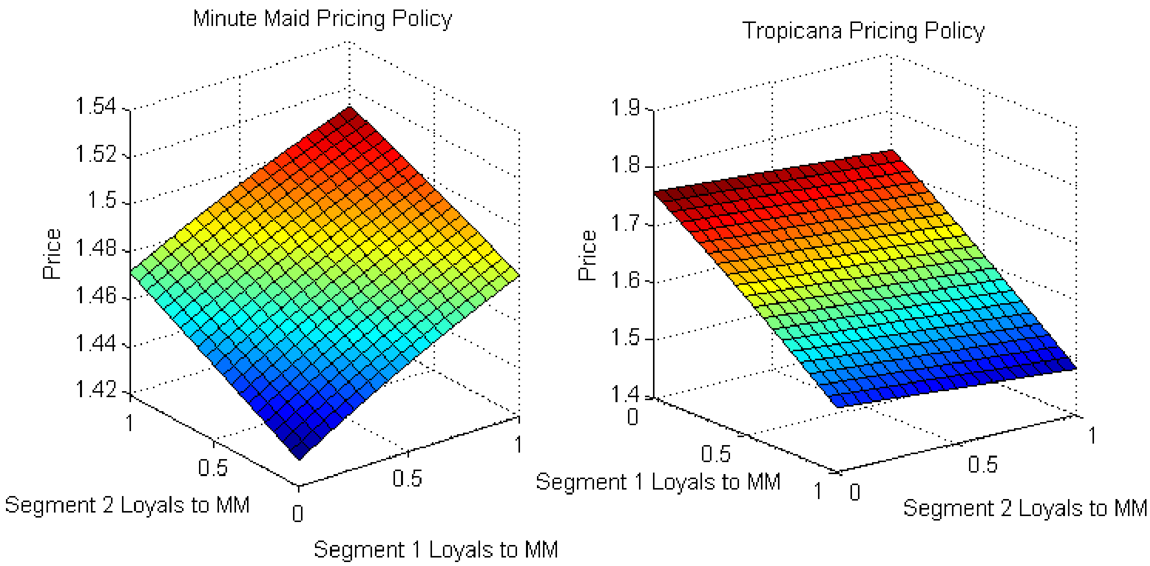


Figure 9
Convergence of Orange Juice Prices to their Stationary Levels