Price Discrimination with Experience Goods: a Structural Econometric Analysis*

Ronald L. Goettler†
Tepper School of Business
Carnegie Mellon University

Karen Clay‡
Heinz School of Public Policy and Management
Carnegie Mellon University

February 15, 2006

*We gratefully acknowledge helpful comments from Shane Greenstein, Bob Miller, Christine Parlour, Holger Sieg, and seminar participants at the University of Arizona, Northwestern (Kellogg), Stanford (GSB), the 2003 ZEW Conference on the Economics of Information and Communication Technologies, and the 2003 CEPR Applied Industrial Organization Conference. We especially thank Eric Wolff and the online grocer for providing the data.

†Tel: (412) 268-7058, E-mail: `ronald.goettler@cmu.edu`

‡Tel: (412) 268-4197, E-mail: `karen.clay@cmu.edu`

Price Discrimination with Experience Goods: a Structural Econometric Analysis

Abstract

Firms often offer menus of two-part tariffs to price discriminate among consumers with heterogeneous preferences. In this paper we study the effectiveness of this screening mechanism when consumers are uncertain about the quality of the good and resolve this uncertainty through consumption experiences. We use consumer-level data to estimate a dynamic structural model of forward-looking consumers with heterogeneous demands, both ex-ante and ex-post, for an experience good sold by a monopolist offering a fixed menu of two-part tariffs. Our analysis highlights four elements that influence consumer behavior and affect pricing strategies: beliefs, switching costs, experiential learning, and (ex-ante) mistakes in tariff choice. Since elements of our data contradict the rational expectations assumption, we impose a slightly weaker beliefs assumption. Despite consumers having, on average, unbiased priors, their beliefs conditional on tariff choice are biased. Consumers on flat fee tariffs tend to have optimistic priors whereas consumers on per-use tariffs tend to have pessimistic priors. Combined with high switching costs, this sorting-induced bias implies that flat fee tariffs can yield high profits for the firm even after optimistic consumers revise their beliefs. Biased priors also lead to biased expectations of consumer surplus. Realized surplus is on average negative, despite expectations of surplus of \$118 per consumer. Regarding the use of menus, we find they are ineffective, yielding almost no gain over the optimal single two-part tariff.

1 Introduction

Firms often price discriminate among consumers with heterogeneous preferences by offering menus of tariff or bundle options. In many cases the tariff or subscription choice is complicated by uncertainty regarding the quality of the product. Furthermore, as consumers learn their valuations of the product, they may change which offer they select from the menu. In the terminology of Nelson (1970), uncertainty regarding an "experience" good can only be resolved through its consumption. Forward looking consumers therefore have an informational incentive to experiment with the good, which leads them to trade off current utility for future utility. In this paper we study the effectiveness of tariff menus as screening mechanisms to price discriminate among consumers of a nondurable experience good.[1]

Our empirical focus is the online grocer market in which the seller delivers groceries ordered online by consumers. Being a new service, uncertainty is particularly relevant. We use consumer-level data to estimate a dynamic structural model of forward-looking consumers with heterogeneous demands, both ex-ante and ex-post, for an experience good sold by a monopolist offering a fixed menu of a flat fee, a two-part tariff, and a per-use tariff.[2] The model we estimate highlights four elements that influence consumer behavior and affect pricing strategies: beliefs, switching costs, experiential learning, and (ex-ante) mistakes in tariff choice.

Our model is rooted in the dynamic brand choice model of Eckstein, Horsky, and Raban (1989), which introduced Bayesian learning to consumer choice in a fully dynamic context.[3] In our application, each consumer is endowed with a "match-value" (i.e., mean utility) for the online grocer's service. Each time the service is used, the consumer experiences a realized utility centered around her true match-value. This signal is used to update her belief regarding match-value. We extend this basic learning model to account for the consumer's choice of tariff upon enrollment and after each update of beliefs. Given a belief of match-value, the optimal tariff is simply the one that

[1]Courty and Hao (2000) and Miravete (1996,2002) investigate screening mechanisms for non-experience goods. In that context, there is no dynamic tradeoff since uncertainty is resolved independently of consumption choices. DellaVigna and Malmendier (2005) investigate tariff choice for health clubs (arguably an experience good) but not in the context of learning.

[2]We focus on fixed tariff menus for two reasons. First, determining optimal tariff menus with consumer learning and heterogeneous demand (ex-post and ex-ante) is difficult even when menus are fixed. Second, the tariff menu in our data was fixed over the whole seventy-week period. More generally, firms rarely play the game of unexpectedly raising prices after a learning period of low prices.

[3]Variants of this model have been estimated in the economics and marketing literatures. Erdem and Keene (1996) investigate the role of advertising in consumer learning about goods. Ackerberg (2003) also studies advertising in a dynamic context, focusing on the distinction between informative advertising and prestige or image advertising. Crawford and Shum (2005) estimate how rapidly consumers learn about the effectiveness and side-effects of anti-ulcer drugs based on their own experiences. Predating these consumer oriented applications, Miller (1984) estimates a Bayesian learning model to study the matching of workers to jobs.

maximizes expected discounted utility. This tariff choice induces a sorting of consumers based on beliefs. Consumers who expect to use the service often choose tariff plans with higher flat fees and lower per-use prices. The observed high usage of such consumers therefore reflects both the self-selection and the lower per-use price they face. By endogenizing tariff choice, our structural model disentangles these two effects on usage rates.

Our data present three puzzles that are difficult to explain if consumers are assumed to act rationally given rational expectations. First, 79 percent of consumers who initially choose the flat fee have realized usage rates below the level for which the flat fee is optimal. Under rational expectations, however, at least half of the flat fee enrollees should have realized usages high enough to justify this choice since their priors are the means of their known match-value distributions. The second puzzle is that, despite being able to change plans at any time without penalty, many consumers remain on a "wrong" plan even after their usage behavior reveals that they have sufficiently revised their beliefs. The final puzzle is that some consumers sign-up for a tariff with a flat-fee component and never use the service. These puzzles are not unique to our data: Miravete (2003) and DellaVigna and Malmendier (2005) show that consumers sometimes choose and retain the wrong tariff for telephone calling plans and health clubs, respectively.

To address the first puzzle we abandon the rational expectations assumption that consumers know the distribution of their match-value. Instead we assume each consumer receives a signal, prior to any consumption, that is centered around her actual match value. A consumer's prior is then centered around this signal's value and has a variance equal to the signal's known variance.[4] The key advantage of this assumption for explaining our data is that despite being an unbiased signal for each consumer, the signal induces a bias in beliefs when conditioning on the consumer's tariff choice. For example, consumers with optimistic priors tend to have high beliefs and therefore tend to choose tariffs with high flat fees and low per-use prices. DellaVigna and Malmendier (2005) find that biased beliefs contribute to seemingly irrational behavior by health club members. Our model offers a plausible source of such biases when consumers face tariff choices or other types of menus.

The puzzle of persistence in tariff choice despite modified usage is easily accomodated by including switching costs in the tariff choice model. We estimate that consumers would change plans only if the value of expected discounted utility were at least \$176 higher on an alternative plan. Given our estimated discount factor, this cost is equivalent to \$4.75 per week.

The first two puzzles are resolved maintaining optimal behavior. The fact that some consumers sign

[4] Implicitly, this assumes that the consumer has no knowledge (i.e., a flat prior) prior to the arrival of this signal. Otherwise, beliefs before using the service would be a combination of the signal and the (informative) belief prior to the signal. The informative signal may arise, for example, from conversations with friends, active search regarding the nature of this new good, or exposure to advertisements.

up for plans with fixed fees and never use the service, however, suggests optimization errors. Our econometric specification therefore permits mistakes in the initial plan choice.[5] Our estimates imply that nearly half of consumers enroll with the wrong tariff, sacrificing an average of \$43, which is 23 percent of the expected consumer surplus per mistaken consumer. A consequence of abandoning rational expectations, however, is that expected surplus is biased due to the bias in prior beliefs. The average consumer (accounting for mistakes) expects surplus of \$118.54 but realizes a surplus of negative \$45.45. In realized terms, the cost of mistakes is approximately zero. Also, the effect of mistakes on the firm's profits are neglible: some consumers' mistakes increase profits whereas others' mistakes reduce profits.

Using our estimated model of consumer behavior, we numerically solve for fixed tariffs that maximize discounted profits under a variety of restrictions.[6] For example, we solve for the optimal flat fee tariff, the optimal uniform price (i.e., per-use tariff), the optimal menu of two two-part tariffs, and the optimal menu of two two-part tariffs with an additional per-use only tariff. In all cases the firm is assumed to know the distribution of consumers' match-values as well as all other demand parameters.

Two features of the data and model drive the optimal tariff findings. First, consumers' tariff choices are highly persistent, resulting in the high switching costs. The second feature is the substantial uncertainty regarding match-values: the estimated standard deviation of the prior belief is \$17.60. Combining these features, our model suggests offering tariffs with high flat fees to appropriate expected consumer surplus from optimistic consumers who then continue to pay these fees even after learning that their initial beliefs were optimistic. To assess the importance of switching costs we simulate the model under the assumption that switching costs are occasionally zero for each consumer.[7] In this case we find uniform prices are substantially better than flat fees. Hence, the effectiveness of uniform pricing versus flat fees for experience goods depends crucially on the nature of switching costs.

Consistent with Miravete (2004) and Courty and Hao (2000), we find that the use of tariff menus to price discriminate is largely ineffective.[8] Adding a second two-part tariff increases discounted profits marginally and adding a third tariff (per-use only) offers no additional gain. This finding

[5]The tariff choice literature distinguishes between ex-ante mistakes in which consumers choose the wrong plan given their beliefs (i.e., expected usage) and ex-post mistakes given their realized usage. Our use of "mistakes" always refers to ex-ante mistakes.

[6]A number of theoretical studies derive optimal (uniform) price paths when quality is uncertain (e.g., Shapiro (1983), Milgrom and Roberts (1986), and Bergemann and Välimäki (2004)).

[7]The assumption that switching costs are always zero seems implausible. Most consumers, however, have periods in which the (time) cost of managing their subscriptions are lower.

[8]Miravete (2004) finds limited gains from complex tariffs when consumers learn about their demand over time, as does Courty and Hao (2000) when ex-ante consumer heterogeneity is high.

may be sensitive to the particular manner in which consumers are permitted to differ. We therefore estimate our model with random coefficients on all the parameters, using the importance sampling method of Ackerberg (2004). Despite the added consumer heterogeneity, we still find only a small increase in profits when the monopolist offers a menu of tariffs, relative to a single two-part tariff.

In section 2 we present our model of consumer learning. In section 3 we discuss the data and the implausibility of rational expectations. In section 4 we discuss econometric issues, and in section 5 we present the parameter estimates and their implications for price elasticity and consumer surplus. In section 6 we perform counterfactual experiments to decompose consumer behavior into effects due to tariff choice mistakes, switching costs, and match-value uncertainty, followed by experiments to investigate pricing strategies.

2 Model

We model the consumer's decision of whether to use the online grocer or traditional grocers.[9] One could imagine modeling this decision on each shopping occasion, as well as the endogeneity of shopping frequency. This is not possible given our data, since we do not observe the use of traditional grocers. Instead, we assume consumers buy groceries at least once per week, and we model whether they use the online grocer on at least one of these occasions. Throughout the paper, we speak as if consumers purchase groceries exactly once per week, from either the online grocer or a traditional grocer.

The consumer's decision has two dynamic aspects. First, the online grocer is a new service about which consumers have limited information. We model online grocery delivery as an "experience" good. As the consumer uses the good, she learns, in a Bayesian fashion, whether it is a good match for her. If her prior belief suggests the product is not good, she may still try it since the lower expected current utility from the online grocer may be offset by the possibility of learning that it is in fact good.

The second dynamic aspect arises from the online grocer's use of "subscription plans". Consumers are offered a fixed menu of M two-part tariffs denoted by $(F_1, p_1, \ldots, F_M, p_M)$, where F denotes the vector of flat fees (paid each week regardless of usage) and p denotes the vector of per-use prices (paid only if the service is used).[10] To be incentive compatible, the total payment when the

[9] While the model we propose applies to many products or services that are used repeatedly, we present it using language specific to our empirical application.

[10] F and p are sometimes called ex-ante and ex-post prices, respectively. Although the online grocer quotes fees on a monthly basis, consumers are permitted to change plans at any time, with fees appropriately pro-rated based on the actual time on each plan. Hence, F is a payment per week, and consumers only commit to paying one week of fees.

service is used must be decreasing in F. That is, if $F_1 > F_2$ then $F_1 + p_1 < F_2 + p_2$. Each consumer chooses the best subscription plan (i.e., tariff), given her beliefs about the value of the service to her. For example, if she believes the service is of high value then she expects to use the plan often and accordingly chooses a plan with a high fixed component.[11]

We also allow for costs to changing plans. For many products such costs are explicit financial charges. Since the online grocer does not charge consumers to change plans, we interpret this cost as a disutility from thinking about which plan is best and having to call the online grocer to request the change. Furthermore, switching costs are permitted to vary independently across time to reflect the varying nature of demands on a consumer's time and attention. Let δ_{it} denote the switching cost to consumer i in week t.

Each week consumers choose subscription plans and usage to maximize the expected discount flow of utility from grocery shopping, net of switching costs, conditional on the set of available information[12]:

$$\max_{\{s_\tau(I_{i\tau}), c_\tau(I_{i\tau}, s_\tau, u_{i\tau})\}_{\tau=t}^{\infty}} \mathrm{E}\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} \left(\alpha F_{s_\tau} + \delta_{i\tau}\mathcal{I}(s_\tau \neq s_{\tau-1}) + U_{ic_\tau\tau}(s_\tau, u_{i\tau})\right)|I_{it}\right], \tag{1}$$

where $c_t \in \{0, 1\}$ is the consumer's usage choice in period t ($c_t = 1$ corresponds to the online grocer), $s_t \in \{0, \ldots, M\}$ is the subscription (i.e., tariff) choice, u_{it} is a vector of i.i.d. shocks to utility from each of the usage choices, β is the weekly discount factor, α is negative the constant marginal utility of money, and $\mathcal{I}$ is an indicator function. Importantly, u_{it} is known by the consumer prior to the choice of c_{it} but is unknown prior to the choice of s_{it}. Hence, the F_{s_t} is the fixed fee component of the selected tariff. The notation in (1) is explicit about the fact that the consumer's maximization is over the set of functions that maps future information sets into choices since information evolves over time as the consumer processes experience signals.

The utility consumer i obtains from using the traditional grocery store in period t is simply the idiosyncratic shock:

$$U_{i0t} = u_{i0t}. \tag{2}$$

[11]The classic use of two-part tariffs is to extract surplus from consumers who face no uncertainty and demand multiple units of a good. The monopolist chooses p to induce the efficient consumption level at which the consumer's willingness to pay for the marginal unit equals the firm's marginal cost, and chooses F to extract surplus from the consumer on the inframarginal units. If the consumer instead demands either zero or one unit, and faces no uncertainty, then two-part pricing is powerless since F and p are indistinguishable. However, if consumers face uncertainty regarding their future demand—perhaps due to a demand shock yet to be realized—then two-part pricing can extract surplus even when demand is only for a single unit. In essence, consumers with uncertain demand for a single unit have a downward sloping demand curve in the probability of buying the good. With downward sloping demands, a menu of two-part tariffs may be used to segment a population of heterogenous consumers.

[12]The utility is derived from the services provided by the grocery vendor, not from the groceries themselves. We could alternatively refer to the consumer as minimizing the disutility from grocery shopping.

For example, U_{i0t} will be high if the consumer happens to be driving by the store while running other errands.

The utility consumer i obtains from using the online grocer in period t is

$$U_{i1t} = \mu_i + \epsilon_{it} + \alpha p_{s_{it}} + u_{i1t} \tag{3}$$

where u_{i1t} is the idiosyncratic shock, $p_{s_{it}}$ is the per-use component of tariff s_{it}, and $\mu_i + \epsilon_{it}$ is the "experience signal." The first part of this signal is the consumer's "match value" and the latter part is mean zero idiosyncratic variation due to uncertainties in the provision of the service (e.g., the quality of the fresh produce, or the time it took for the delivery to arrive). If μ_i is known by the consumer, then ϵ_{it} can be deduced. For new products, however, μ_i is unknown, so the consumer is unable to decompose the overall experience signal into its separate components. Nonetheless, since μ_i is the mean experience signal, each observed signal provides information that can be used to learn about the value of μ_i.

Following Eckstein et al. (1988), Erdem and Keane (1996), Ackerberg (2003), and Crawford and Shum (2005), we specify a Bayesian learning process that exploits the theory of conjugate distributions, as described in DeGroot (1970). In particular,

$$\epsilon_{it} \sim \text{i.i.d. } N(0, \sigma_\epsilon^2), \tag{4}$$

combined with an initial prior on μ_i

$$\text{Initial prior:} \quad \mu_i \sim N(m_{i0}, \sigma_{i0}^2) \tag{5}$$

yields a learning process in which the consumer's posterior on μ_i in period t after n experience signals with mean $\bar{\mu}$ is given by

$$\text{Posterior:} \quad \mu_i \sim N(m_{it}, \sigma_{it}^2), \tag{6}$$

where

$$m_{it} = \frac{\sigma_\epsilon^2 m_{i0} + n\sigma_0^2 \bar{\mu}}{\sigma_\epsilon^2 + n\sigma_0^2}, \tag{7}$$

and

$$\sigma_{it}^2 = \frac{\sigma_\epsilon^2 \sigma_0^2}{\sigma_\epsilon^2 + n\sigma_0^2}\ . \tag{8}$$

This model of Bayesian learning is tractable because all of the consumer's information regarding μ_i is captured by the posterior mean m_{it} and posterior variance σ_{it}, both of which have closed-form expressions. Furthermore, the posterior variance is a function of only the number of signals received, not the realized values for these signals. Note that if σ_ϵ is zero, all uncertainty is resolved after the first experience signal.

Since m_{it} and σ_{it} summarize all relevant information at time t, we can use them as state variables to convert the sequential maximization problem of (1) into a recursive formulation given by Bellman's equation. Recall that consumers choose their subscription plans prior to observing the idiosyncratic shocks u. The recursive formulation is therefore easier if we model the sequence of decisions within a period to first entail the usage choice for the current period, followed by the plan choice for the following period. That is, the consumer enters the period with the plan choice that was selected at the end of the previous period as a state variable. The consumer then chooses usage and considers changing plans if an experience signal causes her beliefs to change by a sufficient amount to warrant incurring the switching costs. The resulting Bellman's equation is

$$V_u(m_{it}, \sigma_{it}, s_{it}, u_{it}) = \max_{c_{it}, s_{it+1}} \mathrm{E}\left[U_{ic_{it}t} + \beta V_u(m_{it+1}, \sigma_{it+1}, s_{it+1}, u_{it+1}) | (m_{it}, \sigma_{it}, s_{it}, u_{it}), c_{it}\right]. \quad (9)$$

The expectation is over the current period experience signal $\mu_i + \epsilon_{it}$ (with an expected value of m_{it}) in addition to next period's state. Following Rust (1987) we integrate out over the i.i.d. u shocks to remove them from the state space since these shocks only affect current utility. Assuming u are i.i.d. type I extreme value, this integration has an analytic solution. The remaining expectation of the continuation value is driven by the random experience signals and the random switching costs. To make this explicit, let $\mu_{it} = \mu_i + \epsilon_{it}$ denote the realized experience signal (when $c_{it} = 1$) and G_δ denote the i.i.d. distribution of δ_{it}. In the following "integrated" value function, the second line corresponds to $c_{it} = 0$, while the subsequent lines correspond to $c_{it} = 1$:

$$\begin{aligned} V(m_{it}, \sigma_{it}, s_{it}) \quad = \quad & \gamma + \ln \Bigg[\\ & \exp\left(\beta \int \max_{s_{it+1}} \left\{ V(m_{it}, \sigma_{it}, s_{it+1}) + \alpha F_{s_{it+1}} + \delta_{it} \mathcal{I}(s_{it+1} \neq s_{it}) \right\} G_\delta(d\delta_{it}) \right) + \\ & \exp\Bigg(m_{it} + \alpha p_{s_{it}} + \\ & \quad \beta \int \max_{s_{it+1}} \left\{ V(Q_m(m_{it}, \sigma_{it}, \mu_{it}), Q_\sigma(\sigma_{it}), s_{it+1}) + \alpha F_{s_{it+1}} + \delta_{it} \mathcal{I}(s_{it+1} \neq s_{it}) \right\} \\ & \quad G_\delta(d\delta_{it}) \Phi(d\mu_{it} | m_{it}, \sigma_{it}) \Bigg) \Bigg], \end{aligned} \quad (10)$$

where γ is Euler's constant, Q_m and Q_σ govern the updating of the posterior mean and variance using (7) and (8), and Φ is the perceived distribution of experience signals.[13] Importantly, Φ accounts for uncertainty about μ_i (via σ_{it}) as well as the variance due to ϵ_{it} (via σ_ϵ).[14]

[13] The updating equations are evaluated using $n = 1$, $\bar{\mu} = \mu_{it}$, and the current beliefs (m_{it}, σ_{it}) in place of the initial prior values m_{i0} and σ_0.

[14] The dependence of V on the model's fixed parameters (σ_ϵ and α) is supressed.

Note that the maximization that determines s_{it+1} occurs inside the integral since consumers make this choice after observing the random switching cost and, in the case of $c_{it} = 1$, the experience signal. Also note that when $c_{it} = 0$, the consumers beliefs μ_{it} and σ_{it} do not change since no experience signal is received.

Although the above dynamic model does not have an analytical solution, we can solve it numerically. Since Φ is $N(\mu_{it}, \sigma_{it}^2 + \sigma_\epsilon^2)$, the integral over μ_{it} is efficiently evaluated using Gauss-Hermite quadrature (Judd 1998). In our econometric model and counterfactual simulations, G_δ is specified to have either one or two mass points, which is computationally trivial. We discretize the posterior mean (a continuous state variable) and use linear interpolation to evaluate V at points off the grid.[15] The posterior variance is a deterministic function of the number of experience signals processed, which we set sufficiently high that the incentive to learn beyond the grid is very small.[16]

The learning aspect of the model has important implications for predicted usage patterns. As consumers learn about their match-values over time, the frequency with which they use the online grocer will change. If experience signals are relatively informative (i.e., σ_ϵ is low relative to σ_{i0}), usage patterns will stabilize quickly.

The forward-looking nature of the model provides an incentive to learn about one's match-value, which also has implications for usage patterns. In particular, a consumer may be willing to sacrifice current (expected) utility to experiment with the online-grocer to attain information that will be useful for making better future decisions. This incentive to experiment is increasing in β and σ_{it} and decreasing in σ_ϵ. Since σ_{it} weakly declines over time, a decline in usage over time does not necessarily reflect a decline in the mean perceived match-value μ_{it}: the decline may instead reflect the decreased incentive to experiment.

The model also has important implications regarding the distribution of beliefs across consumers subscribing to each of the tariff plans. Consumers on tariffs with high flat fees (F) will tend to have high beliefs μ_{it}. Hence, their frequent use of the online grocer reflects both their high μ_{it} and the fact that they face low per-use prices (p). The presence of switching costs dampens this sorting: a consumer may initially choose a high F tariff but then experience a sequence of bad signals thereby lowering μ_{it}, although not by enough to incur the cost of switching plans. That is, the consumer may make the correct choice ex-ante, but end up on the "wrong" plan ex-post.

Additional implications of the model appear below in our discussions of parameter identification and the counterfactual experiments.

[15] In our counterfactual experiments, we find that with no parameter heterogeneity across consumers, profits are quite discontinuous unless the grid is sufficiently fine (around .05). The jumps reflect the fact that optimal subscription plan thresholds (in beliefs) are common across consumers.

[16] In the estimation we allow for 70 uses since the data cover 70 weeks.

3 Data

We use consumer-level data on grocery deliveries to 5310 households in a single metropolitan market during the 70 weeks from September 16, 1997 to January 23, 1999. The earlier date is the online grocer's commencement of service. The online grocer teamed up with an existing local grocery chain to supply the groceries. Online prices and discounts were the same as offered in the chain's stores. Consumers learned about the service through advertising, in the form of mass mailings, media stories, print and radio advertising, in-store advertising in the partner chain, and displays on the delivery trucks. Most consumers signed up while shopping in the partner-chain's stores. Once enrolled, consumers placed orders from their computers using installed software or a web-based interface. Consumers selected a two-hour delivery window, typically the next day, during which someone would be home to accept the delivery.

We observe each consumer's enrollment date and initial tariff choice, the date of each of her orders, her subscription plan (i.e., tariff) at the time of each order, and the dollar amount of each grocery order (net of delivery costs).[17] We do not, however, observe the set of consumers who considered enrolling but chose not to do so. The fact that we observe a steady stream of new enrollees throughout the 70 weeks suggests that consumers became aware of this new service slowly over time.[18] Hence we would be uncomfortable using an assumption about market size to infer the proportion of consumers who deliberately chose not to signup. Our estimates are therefore conditional on the set of consumers who enrolled. For example, our estimate of the distribution of match-values is the distribution of match-values across enrollees, not the general population.

Fortunately, for many policy experiments this does not create a problem. For example, the effect of a price increase can be predicted without concern. Usage and enrollment predictions for price decreases, however, will be lower bounds since lower prices would induce participation by consumers outside the population of our observed enrollees. Nonetheless, a conclusion that lower prices would increase profits from our conditional population would necessarily imply that profits from the general population would also increase. We take this issue into account when performing such counterfactuals.

Recall that the model in the previous section treats the consumer's decision on a weekly basis.

[17] The average purchase amount was approximately \$119. Boatwright, Borle, and Kadane (2003) study the joint distribution of purchase quantity and timing.

[18] We treat each consumer's enrollment date as exogenous. This timing is clearly endogenous when consumers strategically delay adoption of a new good in order to learn from other consumers' experiences. McFadden and Train (1996) and Bolton and Harris (1999) provide theoretical models in which consumers trade-off ex-post learning from one's own experiences versus learning from ex-ante sources, such as others' experiences or advertising. Our results suggest that consumers in our data learned primarily from their own experiences.

Week t=1 for each consumer is the week beginning with her enrollment date. The usage variable c_{it} is set to 1 for each week (since enrollment) in which the consumer ordered from the online grocer one or more times.[19] Otherwise, c_{it} is set to 0 to indicate that only the traditional grocer was used.

For weeks with orders we set s_{it} to be the index value associated with the recorded subscription plan. Unfortunately, we do not observe a consumer's subscription plan during weeks beyond the enrollment week unless an order is placed. Hence, our econometric model will need to account for this censoring. For weeks between orders in which the subscription plan is the same, we can safely set s_{it} to be the index of this plan.[20] Somewhat surprisingly, in our data we never observe consumers who switch plans and order after the switch.[21] Hence, we only encounter a censoring issue for the "trailing weeks" between a consumer's last order and the end of our sample period (January 23, 1999). As detailed in the next section, we integrate over the censored subscription to compute the probability of observing no usage during these trailing weeks.

Thus far this section has provided information about the structure of the data we use to estimate the model. The remainder of this section describes the tariff menu, characteristics of consumers who enrolled, and interesting moments in the data.

3.1 Evidence of suboptimal behavior or irrational beliefs

The online grocer in our market offered consumers a menu of three tariffs.[22] In Table 1 we describe each of these tariffs and provide relevant summary statistics. Plan 1 is a fee only tariff with a weekly fee of \$5.76; Plan 2 is a two-part tariff with a flat fee of \$1.14 and a per-use price of \$6.95; and Plan 3 is a uniform price of \$11.95 per use.[23] These weekly fees are derived by dividing the quoted monthly fees (\$24.95, \$5.00, \$0) by 4.33 weeks per month. Using the notation of the model in section 2, $F = (5.76, 1.14, 0)$ and $p = (0, 5, 11.95)$.

Only 12 percent of consumers signed up for plan 1, compared to 32 percent for plan 2 and 56 percent for plan 3. The mean usage rate was .56 for plan 1 enrollees, .36 for plan 2 enrollees, and .20 for plan 3 enrollees. Each consumer's usage rate is computed from the weeks spanning enrollment and

[19]We rarely observe a consumer ordering more than once in a given week. Of weeks with orders (at the consumer level), fewer than 1 in 22 have multiple orders.

[20]The probability of switching plans and then switching back without receiving any experience signals is zero.

[21]The online grocer claims that customers even ignored letters explaining that they should switch to a different plan.

[22]Most online grocers offer a single fee structure, choosing either delivery fees or monthly charges. Some offer delivery charges declining in the size of the order.

[23]On plan 1 (the high fee plan), orders less than \$60 incur a \$3.95 delivery charge whereas larger orders have no marginal delivery charge. Although our model does not account for this charge on small orders, we are not concerned since only 3 percent of orders from plan 1 consumers incur this charge and only 12 percent of orders from consumers on plans 2 and 3 are smaller than \$60.

the last observed order. This measure is an upper bound since it ignores weeks beyond the last order during which the consumer may have remained on the plan but did not order. In Table 1 we also report the range of usage rates for which each plan minimizes the expected cost per order. Figure 1 plots this expected cost for each plan as a function of expected usage. Plan 1 minimizes this cost for usage rates above .67, plan 3 minimizes costs for usage rates below .23, and plan 2 minimizes costs elsewhere.

The observed usage rates for consumers on plan 1, however, indicate that some consumers are not minimizing expected costs per order. The average usage rate for these consumers is only .56. The average usage rate for consumers on plan 1 who order at least once is .64, which is also below the minimum usage rate for which plan 1 minimizes expected costs per order. This implies that some consumers are either behaving suboptimally given rational beliefs or behaving optimally given beliefs that are "irrational" in some sense, or both.[24] Since consumers are maximizing expected discounted utility, not minimizing costs per order, the reasoning behind this claim may not be obvious. For example, a usage rate of plan 3 consumers exceeding .23 would not necessarily imply suboptimal behavior or irrational beliefs. If switching costs are substantial, consumers who expect to use less often than .23 (in the long run) may sign up with plan 3 but initially use more often than .23 as they "experiment" to sharpen their beliefs.[25] Since experimental learning raises usage, similar reasoning cannot, however, reconcile the observed plan 1 consumers with usage rates less than .67.

The above evidence of suboptimal behavior or irrational beliefs is based on averages across consumers. Not surprisingly, we have identified individual consumers, on all plans, with usage patterns that suggest suboptimal behavior or irrational beliefs. For example, we report in the last column of Table 1 that 12 percent of consumers on plan 1 and 18 percent of consumers on plan 2 never order. Such outcomes are unlikely since we only use consumers who enroll at least 12 weeks before the end of our data.

We present additional evidence of irrational beliefs, based on the changes in usage over time, in Figure 2. The top solid line represents the mean usage rate over time for consumers who initially signed up on plan 1. For example, the plotted usage for week t is the mean c_{it} across consumers who initially signed up for plan 1, even if they may have already quit the service (by switching to plan 3 and never ordering again). Nearly 67 percent of the plan 1 enrollees used the online grocer during their first week. Their usage declined steadily to below .20 by 60 weeks after enrollment.[26]

[24]Here, we use the term "irrational" to refer to any beliefs that do not satisfy the strict rational expectations assumption that consumers know the distribution(s) from which their match values are drawn.

[25]In the absence of switching costs, a consumer would always choose the tariff that minimized expected costs given next period's expected usage.

[26]The usage measures become noisy as "weeks since enrollment" increases since few consumers signed up early

This pattern suggests these consumers initially had overly optimistic beliefs. If their beliefs were unbiased (i.e., correct on average), half the consumers would have revised their beliefs up and half would have revised their beliefs down as they experienced the service. Usage would have declined slightly (due to the reduced incentive to learn as beliefs get more accurate), but not nearly as much as we observe.

The decline in usage for consumers on plan 3 in Figure 2 provides evidence of consumption-based learning by forward-looking consumers. Although we just argued that the steep decline in plan 1 consumers' usage reflects their revision of biased beliefs, such an argument is less plausible for plan 3 consumers. Plan 3 has no flat fee and therefore appeals to consumers who have low expectations for their match value. Hence the (relatively) high usage rates during the first few weeks of enrollment is likely a response to the incentive to acquire information regarding their match values.

The dotted lines in Figure 2 are usage rates conditional on consuming some time beyond the week for which the usage rate is being computed. Since this conditioning event selects from consumers who are learning that their match-value is (relatively) high, the rates are higher than the unconditional usage rates and are increasing over time. This conditional usage rate goes as low as .53 for consumers on plan 1, despite the fact that this plan only makes sense for consumers who expect to use the service at least 2 out of every 3 weeks. Our specification of beliefs and initial plan choice, described in the next section, is designed to explain these elements of the data.

3.2 Characteristics of enrollees

More than half the enrollees voluntarily provided demographic information. This demographic information includes household structure, age of the subscriber, and income. Comparing households with and without demographic data, we find the two groups do not differ significantly on dimensions such as enrollment date, plan choice, and usage. Thus, the consumers that provided demographic data appear to be representative of all enrollees.

Table 2 shows the demographic characteristics of consumers across plans. Households on plans 1 and 2 tend to have more children and have higher income than households on plan 3.

4 Estimation

In this section we address issues that arise when estimating our model. First, we need to specify initial beliefs. Second, we consider the initial tariff choice when the consumer signs up. We then

enough to provide such data.

discuss identification of the model's parameters. In particular, the price coefficient's identification is non-standard since the menu of tariffs is fixed. We conclude the section by discussing our use of maximum likelihood.

4.1 Initial prior beliefs

In the previous section we presented evidence that consumers who choose the fee only tariff (plan 1) are, on average, overly optimistic regarding their match value. However, the standard rational expectations assumption—that consumers know the distribution from which match-values are drawn—implies consumers have unbiased beliefs. Hence an alternative specification is needed. Rather than basing the initial prior on knowledge of the actual distribution of match-values, we assume each consumer receives an unbiased signal prior to initially choosing a tariff.

Specifically, let $G_\mu(\mu_i)$ denote the distribution of match-values and let

$$m_{i0} \sim N(\mu_i, \sigma_0^2) \tag{11}$$

denote the consumer's signal of her match-value. We assume that the consumer has no knowledge about G_μ (i.e., her "prior" before receiving the signal has infinite variance). Hence, the consumer's updated prior for μ_i is $\mathrm{N}(m_{i0}, \sigma_0^2)$ and the Bayesian learning model of section 2 applies, with $\sigma_{i0} = \sigma_0$ for all consumers. Note that this specification provides two sources of ex-ante heterogeneity among consumers—match-values and signals.

Since signals are unbiased, beliefs are unbiased when averaging over all consumers. Conditional on m_{i0} being high relative to the mean of G_μ, however, beliefs are optimistic (i.e., biased upwards). Consumers who choose plan 1 have high m_{i0} and therefore base this choice in part on their biased beliefs. Of course, given our assumption that consumers do not know G_μ, they are still acting rationally: they are not aware that their m_{i0} are high relative to the mean of G_μ. In short, our signal-based specification of beliefs yields an "endogenous bias" that can explain the aspects of the data depicted in Figure 2 and Table 1, as discussed in the previous section.

4.2 Initial tariff choice

Tariff choice s_{it} is a state variable in the model of section 2, but it is undefined prior to enrollment. We therefore need a separate model of the intitial tariff choice. Recall the model of tariff choice given s_{it} that is embedded in the continuation value in (10). We can ignore the switching cost component (since all tariffs entail a "switch" from no tariff) and model consumers as choosing s_{i0}

to solve

$$\max_{s_{i0}\in\{1,2,3\}} V(m_{i0},\sigma_{i0},s_{i0})+\alpha F_{s_{i0}}. \tag{12}$$

This model implies a threshold level of m_{i0} separating all plan 3 enrollees from plan 2 enrollees and a higher threshold separating all plan 2 enrollees from plan 1 enrollees. To gain an appreciation for the effect of this segmentation on usage, consider a simplified version of our model in which consumers are myopic. In this case $V(m_{i0},\sigma_{i0},s_{i0})$ is simply $\log(1+\exp(m_{i0}+\alpha p_{s_{i0}}))$, given that the idiosyncratic utility shocks u are type I extreme value. In Table 3 we report, for various values of α, the belief thresholds at which the optimal plan changes as well as the expected usage rates at these thresholds. When α is high plan 2 is sub-optimal for all values of m, and usage jumps from nearly 0 to nearly 1 when m crosses the threshold separating plan 1 consumers from plan 3 consumers. As α declines the range of m for which plan 2 is optimal widens and the discrete jumps in usage at the thresholds decline. As α goes to zero, the differences in utility across plans for a given m goes to zero. Nonetheless, segmentation (based on minute differences in utility) still occurs, leading to distinct usage rates for consumers on each plan, despite the fact that these plans are essentially identical in utility terms.

Unfortunately, this strict segmentation makes it difficult for the model to explain the high usage we observe for many consumers on plan 3 and the low usage of many consumers on plans 1 and 2. In particular, the model struggles to explain why many consumers sign up for plans 1 and 2 and never order. We therefore modify the initial tariff choice model to include idiosyncratic preferences for each of the tariffs:

$$\max_{s_{i0}\in\{1,2,3\}} \lambda_{s_{i0}}+\Lambda(V(m_{i0},\sigma_{i0},s_{i0})+\alpha F_{s_{i0}})+\xi_{i,s_{i0}}, \tag{13}$$

where $\xi_{i,s_{i0}}$ are type I extreme value.[27] This logit model may be interpreted as allowing consumers to make mistakes, the probability of which is decreasing in the magnitude of the differences in V across plans. This model also avoids the undesirable segmentation as α goes to zero discussed above. Also note that as Λ increases, tariff choices converge to the solution to (12).

4.3 Identification

At the end of section 2, we discussed implications of the learning model for predicted usage patterns and in section 3 we presented aspects of the data that are consistent with these implications. We

[27] In principle, we could also add idiosyncratic preferences for each plan in the post-enrollment tariff choices. This would somewhat complicate the problem of censored subscription plans between usages. Furthermore, idiosyncratic preferences for particular plans probably persist over time. Such persistence would be difficult to identify separately from the switching costs already in the model. Our estimated switching costs likely reflect some persistence in preferences for particular plans.

now provide more detail regarding the identification of particular parameters.

The distribution of match-values, G_μ, is identified by consumers' usage rates at the end of their histories—after experience signals have eliminated much of the uncertainty.[28] The degree of initial uncertainty, σ_0, is identified by differences in "pre-information" behavior (before many experiences) from "post-information" behavior (after many experiences). The speed with which behavior adjusts identifies σ_ϵ, the informativeness of the experience signals.

Given the absence of observed plan switches (i.e., switches followed by usage), one might expect switching costs to be estimated to be infinite. High switching costs, however, reduce the model's ability to explain the "trailing weeks" between a consumer's last usage and the end of our dataset's 70-week timeframe. Long trailing periods are more likely due to consumer's "quitting" by switching to plan 3 and never ordering again. Such events only receive significant weight in the integration over the unobserved plan during the trailing weeks if switching costs are not too high.

The tariff menu is constant throughout our data, which leads one to wonder how we identify α, the coefficient on prices. In essence, α is identified from cross-sectional variation in prices and usage across similar households on different plans. As documented, usage across households differs for two reasons—different marginal prices and different beliefs about match values.

First consider the case in which consumers strictly segment themselves by choosing tariffs according to (12). This sorting of consumers according to m_{i0}, however, does not translate into a strict sorting of the true match values, μ_i. In Figure 2 we use simulated data (for which we know μ_i and μ_{i0}) to depict this identification strategy.[29] Consumers with initial beliefs (on the vertical axis) less than -1.676 initially subscribe to plan 3. Consumers with initial beliefs greater than -.175 initially choose plan 1. The remaining consumers with initial beliefs between -1.676 and -.175 choose plan 2. Notice, however, that the distributions of true match qualities (on the horizontal axis) conditional on each initial plan choice are dispersed over the whole range.[30]

If we were to estimate this model with only one week of data, then we could not identify α since differences in usage rates across plans could be exactly explained by shifting the distributions of initial beliefs. However, with multiple weeks of data, consumers' beliefs evolve. If switching costs are substantial (as suggested by the lack of observed switches), then the distribution of beliefs begins to resemble the distribution of true match qualities as consumers revise their beliefs towards μ_i. Hence, similar consumers face different prices, which enables the identification of price sensitivity. In the absence of switching costs, α is still identified by the fact that the distributions of beliefs

[28] G_μ is parameterized in the next sub-section.

[29] The data is simulated using the estimates presented in the next section.

[30] As expected, the means are increasing in the fee of the plan: -4.082 for plan 3, -2.542 for plan 2, and -.961 for plan 1. That is, consumers with high m_{i0} tend to have high μ_i.

for each plan cannot be shifted arbitrarily in each period: the Bayesian learning component of the model specifies how these distributions shift over time.

When we allow consumers to make mistakes regarding their initial plan choice, as in (13), consumers with similar m_{i0} face different marginal prices from the outset. This weakening of sorting via tariff choice strengthens our ability to identify α despite having a fixed tariff menu.

4.4 Estimation

We estimate three econometric specifications of the model. The base specification assumes no parameter heterogeneity beyond what has already been discussed. That is, consumers differ only through their μ_i and beliefs. For comparison, we also estimate this specification with myopic consumers. Finally, we estimate the forward-looking model with all parameters varying across consumers as random coefficients. This latter estimation uses the importance sampling methodology proposed by Ackerberg (2002).

In all specifications we assume the distribution of switching costs is degenerate at a level denoted δ. The absence of plan switches (other than the censored switches to plan 3 by quitters) prevents us from estimating the random switching cost specification. We use random switching costs in the counterfactual exercises to assess whether our results regarding price discrimination are sensitive to the estimated model's prediction that many consumers remain on the wrong plan forever.

Since consumers' match-values and beliefs are not observed we integrate over μ_i, m_{i0}, and ϵ_{it} (the experience signals) to obtain a likelihood function. We assume μ_i are i.i.d. $N(\mu_{G_\mu}, \sigma_{G_\mu})$, although the learning model permits any choice of G_μ.

Let θ denote the vector of parameters to estimate. In the base specification $\theta = (\mu_{G_\mu}, \sigma_{G_\mu}, \sigma_0, \sigma_\epsilon, \beta, \alpha, \delta, \lambda, \Lambda)$. For each draw of unobservables over a consumer's entire history, we compute the likelihood of the observed sequence of c_{it} and s_{it} over the T_i weeks between the consumer's enrollment and the end of the data set. Since s_{it} is unobserved in weeks after the last usage, the likelihood for these "trailing weeks" is based only on the observed $c_{it} = 0$. Let τ_i denote the i^{th} consumer's last week with $c_{it} = 1$. Integrating over the unobserved match-value and beliefs, the likelihood for consumer i is then

$$
\begin{aligned}
L_i(\theta) \quad = \quad \int \Bigg[& \prod_{t=0}^{\tau_i} Pr(s_{it}|m_{it}, \sigma_{it}, s_{it-1}; \theta) Pr(c_{it}|m_{it}, \sigma_{it}, s_{it}; \theta) \\
& \prod_{t=\tau_i+1}^{T_i} \sum_{s_{it}} Pr(s_{it}|m_{it}, \sigma_{it}, s_{it-1}; \theta) Pr(c_{it}|m_{it}, \sigma_{it}, s_{it}; \theta) \Bigg] \\
& \Phi(d\{m_i\}_{t=0}^{T_i}|\mu_i; \theta) G_\mu(d\mu_i)
\end{aligned} \tag{14}
$$

where $\Phi(d\{m_i\}_{t=0}^{T_i}|\mu_i;\theta)$ integrates over the entire sequence of beliefs conditional on the match value, and $G_\mu(d\mu_i)$ integrates over the match value. The summation in the second line integrates over the censored tariff choice after the last usage.

As shown in Miller (1984) and Rust (1987), $Pr(c_{it}|m_{it},\sigma_{it},s_{it};\theta)$ has the familiar logit formula. Net of the idiosyncratic u_{it}, the value of choosing $c_{it}=0$ is

$$V_{0it} = \beta \max_{s_{it+1}} \Big\{ V(m_{it},\sigma_{it},s_{it+1}) + \alpha F_{s_{it+1}} + \delta \mathcal{I}(s_{it+1} \neq s_{it}) \Big\} \tag{15}$$

and the value of choosing $c_{it}=1$ is

$$\begin{aligned} V_{1it} \quad = \quad & m_{it} + \alpha p_{s_{it}} + \beta \int \max_{s_{it+1}} \Big\{ V(Q_m(m_{it},\sigma_{it},\mu_{it}), Q_\sigma(\sigma_{it}), s_{it+1}) + \\ & \alpha F_{s_{it+1}} + \delta \mathcal{I}(s_{it+1} \neq s_{it}) \Big\} \Phi(d\mu_{it}|m_{it},\sigma_{it}) \ . \end{aligned} \tag{16}$$

Both equations modify the expressions for continuation values in (10) to account for the fixed switching costs of δ. We can then write

$$Pr(c_{it}|m_{it},\sigma_{it},s_{it};\theta) = \frac{\exp(V_{c_{it}it})}{\exp(V_{0it}) + \exp(V_{1it})} \ . \tag{17}$$

Discrete choice models require two normalizations since neither the absolute level nor the variance of utility are identified. The absence of an estimated mean utility for $c_{it}=0$ is the additive normalization, and the fixed variance of the u_{it} is the scale normalization.

The probability of the initial plan choice s_{i0} is the logit probability based on (13). The subsequent plan choices are deterministic given beliefs m_{it} and σ_{it}. That is, $Pr(s_{it}|m_{it},\sigma_{it},s_{it-1};\theta)$ equals one if s_{it} is optimal given $(m_{it},\sigma_{it},s_{it-1})$ and equals zero otherwise.[31] Mathematically,

$$Pr(s_{it}|m_{it},\sigma_{it},s_{it-1};\theta) = \mathcal{I}\{s_{it} = s(m_{it},\sigma_{it},s_{it-1};\theta)\}, \tag{18}$$

where $s(\cdot) \equiv \text{argmax}_s\{V(m_{it},\sigma_{it},s) + \alpha F_s + \delta\mathcal{I}(s \neq s_{it})\}$ denotes the model's predicted choice.

Now consider the integration over the censored plan choice for the trailing weeks $t > \tau$. With fixed switching costs, plan changes are predicted to occur only immediately after beliefs are updated. Since beliefs are fixed after the last usage in period τ, the censored plan is simply $s(m_{i\tau+1},\sigma_{i\tau+1},s_{i\tau};\theta)$.[32] Integration over the censored plan is therefore automatically handled by the integration over unobserved beliefs.

[31]From the econometrician's perspective, however, s_{it} is probabilistic since beliefs are unobserved.

[32]With random switching costs, a consumer may update beliefs but postpone switching plans until a period of low switching costs is encountered. For example, with two possible switching costs, $\delta_0 < \delta_1$, the censored plan will be $s(m_{i\tau+1},\sigma_{i\tau+1},s_{i\tau},\delta_1;\theta)$ until the consumer first encounters low switching costs. At this time the plan becomes $s(m_{i\tau+1},\sigma_{i\tau+1},s_{i\tau},\delta_0;\theta)$. For $t > \tau$ the probability of having drawn high switching costs each period since the last usage is $Pr(\delta_1)^{t-\tau}$. The probability of having encountered low switching costs at least once is therefore $1 - Pr(\delta_1)^{t-\tau}$.

We use monte carlo simulation with 500 draws to evaluate $L_i(\theta)$ for each consumer.[33] Our estimator is obtained by maximizing the product of the consumers' simulated likelihoods, using the nested fixed-point algorithm of Rust (1987).

4.5 Parameter heterogeneity

Thus far we have avoided using observable consumer characteristics. Allowing θ to vary across consumers requires finding the fixed point V for each possible consumer type each time we evaluate the likelihood. For example, using three binary demographic variables increases computation time for the likelihood by a factor of eight. Given the number of consumer characteristics at our disposal, simply interacting them with model parameters is not computationally feasible. Furthermore, we only observe these characteristics for the subset of consumers who volunteered such information.

Instead, we use the importance sampling methodology of Ackerberg (2002) to allow θ to vary across consumers as random coefficients. Integrating over random coefficients typically involves averaging $L_i(\theta_i)$ over many draws of θ_i for each consumer. This leads to an infeasible number of fixed points to compute since each θ_i for each i requires an associated V. Furthermore, these V must be recomputed each time the likelihood function is called during the nonlinear maximization over distributions of random coefficients. The idea behind Ackerberg (2002) is to compute and retain $L_i(\theta_i)$ for a set of θ_i. The likelihood under an alternative distribution of random coefficients is obtained not by redrawing θ_i from this new distribution and recomputing conditional likelihoods, but by changing the weights in the averaging of the retained $L_i(\theta_i)$.

To be more precise, let $g(\theta|\rho)$ be the probability density function of random coefficients parameterized by ρ and let $h(\theta)$ be an arbitrary distribution (independent of ρ). Then,

$$L_i(\rho) = \int L_i(\theta_i) g(\theta_i|\rho) d\theta_i = \int L_i(\theta_i) \frac{g(\theta_i|\rho)}{h(\theta_i)} h(\theta_i) d\theta_i \tag{19}$$

We draw $(\theta_i^1, \ldots, \theta_i^{NS})$ from h and compute the simulated likelihood

$$\tilde{L}_i^{NS}(\rho) = \frac{1}{NS} \sum_{ns=1}^{NS} L_i(\theta_i^{ns}) \frac{g(\theta_i^{ns}|\rho)}{h(\theta_i^{ns})} \tag{20}$$

In practice, we initially choose h to be centered around the estimates from the model without random coefficients and to have sufficiently high variance that the reweighting (with h in the

[33] Simulated maximum likelihood has been shown to yield an inconsistent estimator for a fixed number of draws (Hajivassiliou and Ruud,1994). To increase the efficiency and smoothness of the simulation estimator, we draw experience signals from the (truncated normal) distribution for which the observed plan choice is indeed optimal, and reweight the likelihood accordingly. Due to switching costs the set of experience signals for which a given plan is optimal often contains two non-contiguous regions.

denominator) does not explode. We then iterate two or three times by redrawing from an h set to the previous iteration's g. We stop iterating when the estimated ρ implies g is similar to h. Restrictions of parameters like $\alpha > 0$ and $0 < \beta < 1$ are imposed by using truncated normals for g and h.[34]

If demographics were available for all consumers, we could directly condition g on them. Unfortunately, some consumers do not report their demographics. For those who do provide demographic data, we regress their posterior means of θ_i (given the model, g, and choices c_{it} and s_{it}) on their demographics to determine the degree to which preferences are related to observables.

5 Results

Table 4 presents estimates for three specifications: myopic consumers, forward-looking consumers, and forward-looking consumers with random coefficients. Estimates are precisely estimated and are quite similar across all three models, with a few exceptions. Our discussion of the estimates and their implications focuses on the dynamic (i.e., forward-looking) model without random coefficients. Of the three models, this is the most parsimonious one that captures the dynamic trade-offs of interest. Figure 3 depicts simulated versions of the same usage moments plotted in Figure 2. Comparing the two figures suggests that this model is indeed able to replicate the key dynamic features of the data.

The price coefficient is $-.284$ which implies a price elasticity of -1.21 over the 70-week period. The elasticity for week 1, however, is only $-.41$ compared to -1.49 for week 70. The expected discounted value of revenues declines by 4.8 percent when all prices and fees of the tariff menu are increased by 10 percent to compute these elasticities. This distinction in short-term versus long-term behavior is of central importance in the counterfactuals of the next section.

The high estimate of σ_0 (4.998) suggests that consumers were very uncertain about the suitability of this service for them. Indeed, the amount of uncertainty is more than twice the degree of heterogeneity in actual match values ($\sigma_{G_\mu} = 2.136$). Furthermore, the high estimate of σ_ϵ (5.388) implies that the uncertainty persists even after several usage experiences.

The learning rates are further detailed via the evolution of posterior variances and means as usage increases, in Table 5. For example, the standard deviation of beliefs declines by 27 percent after

[34] For each consumer we use 36 draws of the parameters that enter the consumer's dynamic program (i.e., $\sigma_0, \sigma_\epsilon, \beta, \alpha, \delta$). dynamic parameters, we solve the dynamic program and evaluate the likelihood for 100 draws of the other parameters (i.e., $\mu_{G_m u}, \lambda_1, \lambda_2, \Lambda$), thereby reducing simulation error related to parameters that do not enter the fixed point computation.

one use, 39 percent after two uses, and 47 percent after three uses. Less than one-third of the uncertainty remains after ten uses. The third column of Table 5 reveals the evolution of posterior mean beliefs for a hypothetical consumer with a match-value of zero (i.e., $\mu_i = 0$) who received an initial signal optimistic by one standard deviation (i.e., $m_{i0} = \sigma_0 = 4.998$), and received experience signals equal to the match-value of zero.[35] This consumer's belief declines quickly from the initial prior of 4.998 to 2.686 after one experience and 1.125 after four signals.

The discount factor is estimated to be .973, which seems low given that each period is only one week.[36] Nonetheless, the hypothesis that consumers are myopic (i.e., $\beta = 0$) is easily rejected. Furthermore, this degree of discounting still provides a substantial incentive for consumers to sacrifice current utility to obtain information about their match value. The upper graph in Figure 4 plots the model's predicted probability of usage in week 1 as a function of the initial belief and the subscribed plan. The solid lines correspond to forward-looking consumers and the dashed lines depict myopic consumers.[37] For a given belief m_{i0} and initial plan, the probability of using the online grocer is higher when consumers are forward-looking. The lower graph plots the difference in these usage rates. The differences are substantial over a broad range of beliefs, with a maximim difference of .56.

The effect of per-use prices on usage for given beliefs is also evident in Figure 4. Consumers on plan 1 face zero per-use prices and therefore consume at higher rates than consumers on plans 2 and 3.

Not suprisingly given the lack of plan switching, our estimate of δ is high. Dividing the estimate of 50 utils by the price coefficient (.284 dollars per util) implies that the switching cost is equivalent to \$176. That is, consumers would switch plans only if the value of expected discounted utility were at least \$176 higher on an alternative plan. While \$176 may seem high, this cost is equivalent to \$4.75 per week (using the estimated discount factor of .973).

The estimated β is sufficiently high that switching is still predicted to occur: 61.4 percent of consumers who initially choose plan 1 are predicted to switch to plan 3 within 70 weeks of enrolling. None of the plan 2 subscribers, however, are predicted to switch.[38]

In the myopic model the switching costs are substantially lower at 1.778 utils, which is equivalent to \$6.20 when dividing by the myopic model's price coefficient. To predict the same degree of

[35] The posterior mean when the signal equals μ_i is equivalent to the expected posterior mean with random experience signals centered around μ_i.

[36] Ackerberg (2003) estimates the consumer discount factor to be .981 between roughly weekly shopping trips.

[37] The myopic consumers use the same parameters (from the base dynamic model) but ignore the future when making current consumption choices.

[38] In the random coefficients model, some plan 2 subscribers do switch.

switching (in the censored trailing weeks), the myopic model must have lower switching costs since the future benefits of being on the correct plan are ignored. Note that the value of utility sacrificed per week is similar to the \$4.75 computed for the forward-looking model.

5.1 Ex-ante mistakes in tariff choice

The estimates pertaining to the initial tariff choice imply that consumers frequently choose the wrong plan. Since the coefficient on $V - \alpha F_{s_{i0}}$ is relatively low ($\Lambda = .036$), consumers near the threshold beliefs obtain similar expected discounted utility from the plans on each side of the threshold. Such consumers are the most likely to choose the wrong plan. The plan 1 intercept λ_1 is positive and the plan 2 intercept λ_2 is negative indicating preferences for plan 1 and against plan 2, relative to plan 3, after accounting for the expected utility from each plan.

The initial plan choice parameters are most easily interpreted by inspecting the predicted choices of (100,000) simulated consumers. The highest m_{i0} for which plan 3 is optimal is $-.98$, at which V is 35.30, and the lowest m_{i0} for which plan 1 is optimal is 2.65, at which V is 92.54. Plan 2 is optimal for initial beliefs between these thresholds. Given the distribution of initial priors, 18.4 percent of consumers should have initially chosen plan 1, 22.2 percent should have chosen plan 2, and 59.3 percent should have chosen plan 1. Of those who should have chosen plan 3, nearly two-thirds did so, whereas 29 percent chose plan 2 and 5.7 percent chose plan 1. Of those who should have chosen plan 2, 36.1 percent did so, whereas 15.6 percent chose plan 1 and 48.3 percent chose plan 3. Of those who should have chosen plan 1, 32.5 percent did so, whereas 36.3 percent chose plan 2 and 31.1 percent chose plan 3. Overall, only 52.5 percent of consumers chose the optimal initial plan given their initial beliefs.

How costly are consumers' tariff choice mistakes? The expected discounted utility in the absence of the online grocer is simply Euler's constant divided by $1-\beta$ (since the traditional store is chosen in every period) yielding 21 utils. The expected discounted utility, averaged across all consumers, when consumers optimally choose s_{i0} is 60.5 utils. The expected consumer surplus (CS), assuming optimal behavior, is therefore $(60.5 - 21)/\alpha$, which equals \$139.08. The expected CS given the simulated s_{i0}, which allows for mistakes, is only \$118.54. Hence, mistakes cost consumers an average of \$20.54, or 14.8 percent of potential CS. Conditional on making a mistake, the lost CS is \$43.29 (23.3 percent of attainable CS). In the next section we point out that the cost of mistakes differs if measured according to realized utility instead of expected utility.

The differences in V across plans are large enough to induce substantial sorting of beliefs and match values across plans, despite the seemingly low Λ estimate of .036 and the associated mistakes. The mean m_{i0} of consumers on plans 1, 2, and 3 are, respectively, 1.81, -1.67, and -3.51. The corresponding mean match-values μ_i are -1.56, -2.11, and -2.41. The biases, measured as mean $(m_{i0} - \mu_i)$ across consumers on each plan, are therefore 3.37, .44, and -1.10, respectively. That is, consumers on plans 1 and 2 received, on average, optimistic prior signals, whereas consumers choosing plan 3 received pessimistic prior signals.

5.2 Consumer surplus

When consumers have rational expectations, the expected discounted utility of a consumer with initial belief m_{i0} matches the average realized discounted utilities (over long simulations) of many consumers simulated with initial belief m_{i0}. As discussed in sections 3 and 4, however, our consumers do not have rational expectations. Hence, their expected discounted utilities do not match their average realized discounted utilities. We reported above that consumers expected consumer surplus upon enrolling was \$118.54. The average realized consumer surplus, however, was negative \$45.45 with only 19.8 percent of consumers realizing positive surplus.[39]

Initially, we were surprised that consumers with unbiased prior signals generate negative average CS despite an expectation of CS over \$100. We expected overestimates of expected utility by optimistic consumers (primarily on plans 1 and 2) to be partially offset by underestimates by pessimistic consumers on plan 3. We were wrong: consumers on plan 3 also overestimate their discounted utility (and hence surplus). For consumers with high initial beliefs, the discrepancy between expected and realized utility is easily explained. Such consumers tend to choose plan 1 and then learn that their match-values are actually quite low. The high switching costs, however, deter them from switching plans to avoid the flat fees of plans 1 and 2.

The story is different for consumers with low prior beliefs. These consumers tend to have pessimistic prior signals ($m_{i0} < \mu_i$), which suggests they would underestimate their discounted utility. Their expected utility, however, is based on the prior belief that μ_i has standard deviation of $\sigma_0 = 4.998$, which is much higher than the actual dispersion of $\sigma_{G_\mu} = 2.136$. Hence, the perceived probability of having a sufficiently high μ_i to earn high surplus exceeds the actual probability of this outcome, despite the pessimistic first moment belief.

[39]Perhaps not surprisingly, the online grocer has since exited this market. Bear in mind that individual consumers with rational expectations often realize negative surplus: this is the nature of uncertainty.

5.3 Parameter heterogeneity

The estimates reported in Table 4 reveal that the random coefficients model is quite similar to the base dynamic model. The means of the random coefficients are all similar to the corresponding estimates from the model without random coefficients. Estimating this more complicated model would not be worth the effort, merely to achieve the improved log likelihood. The payoff for us is two-fold. We can now assess the degree to which preferences are correlated with observables, which can potentially be useful for marketing purposes. We can also provide a more robust analysis (in the next section) of the benefits of price discrimination via tariff menus. Price discrimination hinges on consumer heterogeneity. Restricting consumers to differ only according to beliefs and match-values may inhibit our ability to use tariff choice to segment consumers.

To determine the relationship between observables and preferences, we could condition the density function of random coefficients, g, directly on observable characteristics. The large number of observables available makes this a computationally intensive approach, even with the importance sampling technique we use to avoid repeated computations of the fixed point V. This approach would also require separate treatment of consumers who do not report their demographic characteristics.

Although not a demographic characteristic, we include each consumer's week of enrollment as a regressor. The fact that this variable is insignificant in all the regressions suggests that consumers learned primarily from their own experiences, versus word-of-mouth and other ex-ante forms of learning. If consumers had indeed learned from others, then consumers who enrolled late in the sample should have had lower σ_0 than those who enrolled early.

6 Policy Experiments

We conduct two sets of policy experiments by simulating the model given the parameter estimates, as well as under alternative parameter values. The first set of counterfactuals is designed to decompose consumer behavior, and the generated revenues, into the separate effects of (ex-ante) tariff choice mistakes, switching costs, and match-value uncertainty. The second set of counterfactuals investigates the effectiveness of various price discrimination techniques for experience goods. In particular, flat fee pricing dominates per-use pricing when switching costs are high, and vice-versa when switching costs are low (at least occasionally). We also find that a single two-part tariff is nearly as effective at generating revenue as a menu of two-part tariffs.

6.1 Measuring the effects of mistakes, switching costs, and uncertainty

Table 6 summarizes consumer behavior and surplus, as well as the firm's discounted revenues, for various specifications of the consumer model. The values were generated by simulating the model for 100,000 consumers over 100 weeks facing the actual tariff menu. This period is long enough for behavior and revenues to near their steady state by its end. Discounted revenue and surplus values are reported in dollars per consumer. The firm's annual discount factor is assumed to be .9, which is .997976 on a weekly basis.

The first model uses the base dynamic model and its estimated parameters. As explained in section 5.2, consumers' realized surplus is negative ($-45.9) despite an expected surplus of $118.5. In the long run (i.e., in week 100) the firm receives $.9075 per consumer, yielding a steady-state discounted revenue of $453.8 per consumer. The steady-state revenue is generally lower than the discounted revenue along the transition path since the firm earns revenues from consumers' experiential consumption.

The second model assumes consumers optimally choose their initial tariff s_{i0}. As expected, the usage rates across plans differ by more than their differences in the first model since the belief-based sorting is muted by mistakes. Expected CS increases by $20.5, but realized CS only increases by $.5. Interestingly, the firm's revenue is $26.8 higher without mistakes.

The third model removes switching costs by setting δ to zero. Optimal s_{i0} choices are maintained, since consumers would simply fix initial mistakes by switching immediately. Comparisons with the previous model reveal that switching costs are responsible for much of the negative CS and much of the firm's revenues. Realized CS increases to negative $19.5 and revenue falls by more than half to $193.2. The influence of switching costs on initial tariff choice is illustrated by 54 percent of consumers initially choosing plan 1 without switching costs, compared to only 18.4 percent with the estimated switching costs (and optimal s_{i0}).

The final model in Table 6 removes uncertainty about match-values by setting $\sigma_0 = 0$. Optimal s_{i0} choices are maintained, and switching costs are inconsequential. Realized CS is finally positive, although only $6 per consumer (same as expected CS) since so few consumers actually use the service. The firm's discounted revenues are higher with uncertainty (in the model with no switching costs) since revenues are received from experimenting consumers. The steady-state discounted revenues, however, are higher with no uncertainty ($182.7 compared to $171.0) since all consumers who ought to use the service indeed do so.

In summary, consumers' behavior and surplus and the firm's revenues are driven largely by the combination of uncertainty and switching costs. In particular, the optimistic beliefs of consumers on plans with flat fees are eventually corrected by learning, but the switching costs lead many of them to continue paying the fees in perpetuity.

6.2 Price discrimination

Two-part tariffs are frequently used to extract surplus from individuals who consume multiple units of a given good. In our model consumers either use the service once or not at all. Nonetheless, since the fee component of the two-part tariff is paid prior to the consumer's observing an idiosyncratic shock, surplus can still be extracted. In essence, the "unit" of consumption is the probability of using the service. When a firm faces consumers with (unobserved) heterogeneous preferences, offering a menu of two-part tariffs can induce them to reveal their preferences through their tariff choices.

In this section we evaluate the use of two-part tariffs and menus of two-part tariffs to price discriminate when the firm sells an experience good. As is evident from the previous section, the role of flat fees in a dynamic context is influenced by switching costs. Not only do flat fees extract surplus in the traditional (static) sense, they also lock consumers into paying these fees even when their expected usage falls in the future. The firm's tariff offerings also determine the degree to which consumers learn. If the offered menu prevents some consumers from experimenting, the firm's long-run profits will suffer as some consumers who should use the product regularly will never discover this fact. On the other hand, the firm may earn substantial revenue from consumers who are willing to pay high prices during the learning period, given their initial beliefs.

We compute optimal tariffs under a variety of scenarios. The optimal tariff varies considerably across these scenarios, which suggests that the effectiveness of different types of tariffs is an empirical question specific to the particular market or good of interest. For our market we find that flat fees are effective only when switching costs are high. When switching costs are low, even if only occasionally, uniform (per-use) pricing is more effective.

Across all scenarios, however, we find that menus of two-part tariffs are ineffective screening devices for price discrimination. That is, menus of two-part tariffs are unable to segment consumers without letting high usage consumers retain too much surplus: the incentive compatability constraints are too costly to satisfy.

We assume that the firm's objective is to maximize expected discounted profits. In addition to the revenue from the tariffs, the online grocer also received a "kickback" from the partner chain of 15 percent of each grocery order. The average kickback is nearly \$18 per order since the average order size is \$119. We do not observe the firm's costs of delivering groceries. An industry analyst who estimates "picking and delivery costs" for a number of online grocers estimates that our firm's costs were \$25.41 per delivery.[40] Marginal costs are presumably lower than this average cost since the delivery truck is already delivering orders to other customers. To simplify matters, we therefore treat the kickback amount as exactly offsetting the marginal costs of delivery and instead maximize discounted revenues from the delivery tariffs only.

Counterfactuals that entail changing the tariff menu require an assumption regarding the intercepts (λ_1, λ_2) and the error terms (ξ) in the model of initial plan choice. That is, the estimated intercepts and distribution of ξ are specific to the actual plans offered. To obtain the intercept for an arbitrary two-part tariff (F, p), we linearly interpolate between the estimated intercepts using $F/(p+F)$ to construct the interpolation weights. Note that this ratio is one for plan 1, .14 for plan 2, and zero for plan 3. We also use this ratio to determine the correlation structure of ξ. For each simulated consumer we draw i.i.d. ξ_{i1}, ξ_{i2}, and ξ_{i3} that correspond to the idiosyncratic utilities from plans with $F/(p+F)$ of zero, .14, and one (as in the actual tariff menu). The idiosyncratic utility for an arbitrary two-part tariff is then determined by the same linear interpolation as used to obtain the tariff's λ intercept.

Table 7 presents the optimal values for various types of tariffs when consumers behave according to the estimated base dynamic model. Given the high switching costs of this model, the optimal flat fee tariff of $F = \$4.85$ generates over four times the discounted revenue of the optimal uniform price of $p = \$6.12$. Surprisingly, the optimal single two-part tariff has the same F as the flat fee tariff, but adds a low per-use price of \$.85. Revenue from this two-part tariff is less than 1 percent higher than revenue from the fee only tariff. Adding a second two-part tariff increases revenue by less than .5 percent. Furthermore, this added tariff is only chosen by consumers who make a mistake in their initial plan choice.

The role of uncertainty in determining whether flat tariffs dominate per-use tariffs is illustrated by the comparative dynamic in Figure 5. For various levels of σ_0 ranging from .1 to 5 utils, we compute the optimal flat tariff and the optimal per-use tariff.[41] The per-use tariff is \$5.50 for $\sigma_0 = .1$ and steadily increases to \$6.53 when $\sigma_0 = 5$. Discounted revenues with per-use tariffs are slightly concave starting at \$237, peaking at \$239.3 at $\sigma_0 = 1.25$, and dropping to \$230.8. The initial gain is due to experiential learning whereas the ultimate decline reflects the increasing number of consumers with high match-values who also have low initial beliefs and never try the service. The flat tariff, on the other hand, rises quickly from \$2.40 at $\sigma_0 = .1$ to \$4.85 at $\sigma_0 = 2.5$ but remains unchanged for higher uncertainty. Discounted revenues with flat tariffs are convex

[40] Reported by David Wellman in "Are we on?" *Supermarket Business*, New York: Dec 15, 1999. Vol. 54, Issue 12, p. 35.

[41] The other parameters are set at their estimated values.

starting at \$182.1 and rising to \$998.5 at $\sigma_0 = 5$. The optimal flat tariffs for σ_0 of 2.5 and higher are the highest fees that can be charged without inducing consumers to change plans after realizing their match-values are low enough that they will almost never use the service. This value equals $(1-\beta)\delta/\alpha$ since $F/(1-\beta)$ is the present value of weekly payments of F.

To avoid the difficulties of determining interecepts and ξ for hypothetical tariffs, the next model assumes consumers do not make mistakes when choosing tariffs. As presented in Table 8, optimal tariffs for this model are very similar to the base model. The only substantive difference is that the second two-part tariff offers no advantage at all, compared to the single two-part tariff.

To assess the importance of switching costs, Table 9 uses the base model augmented with random switching costs. With probability .9 switching costs are the estimated value, otherwise the costs are zero. Even with switching costs being high 90 percent of the time, the effectiveness of flat fees is drastically reduced. The optimal per-use price of \$6.12 generates about 50 percent more revenue than the optimal flat fee of \$3.09. A two-part tariff increases revenue by 1.3 percent, compared to the per-use tariff, and adding a second two-part tariff to the menu increases revenue by an additional 2.3 percent. Adding yet another tariff (a per-use only one) further increases revenue by almost 1 percent. Hence, we again find that using menus of tariffs is of little benefit to the firm.[42]

An alternative method for assessing the effect of switching costs is presented by the comparative dynamic depicted in Figure 6. We compute optimal flat tariffs and per-use tariffs for various levels of switching costs. Regardless of switching costs, the optimal per-use tariff is \$6.51, which yields discounted revenue of \$230.9.[43] The optimal flat tariff is \$2.56 with no switching costs and declines to \$1.54 for switching costs of 7 utils.[44] For switching costs of 8 and higher, the optimal flat tariff is again determined by $(1-\beta)\delta/\alpha$, for which consumers who never use the service continue to pay the fee F. Discounted revenues are higher with per-use pricing for switching costs less than or equal to approximately 7 utils (interpolating between the computed values at 6 and 8).

We have demonstrated that the effectiveness of flat tariffs relative to per-use tariffs is sensitive to the nature of switching costs and degree of uncertainty. The ineffectiveness of offering multiple tariffs to screen consumers, however, is apparent in all of the scenarios analyzed thus far. One reason we find so little benefit from using tariff menus may be that consumers differ in more ways than their match-values. To check whether this finding is robust to additional heterogeneity, we compute optimal tariffs for the random coefficients model. The values in Table 10 are generated by

[42]Results for the model with no uncertainty and no tariff choice mistakes are similar to this model.

[43]Switching costs are irrelevant with only per-use pricing since only one plan is available. When the only offered tariff is a flat fee, a "not participating" plan is implicitly available.

[44]The noisy decline in the optimal flat fee for $\delta < 7$ reflects numerical approximation of the optimal tariff. The revenue is relatively flat near these optimal levels.

simulating 5000 consumers over 100 weeks for each of 100 draws of the random coefficient vector θ. The results are similar to the base model without random coefficients.[45] The flat tariff generates discounted revenue of \$587.2 compared to only \$206.8 for the per-use tariff. The single two-part tariff generates revenue of \$601.9, which is 2.5 percent higher than the flat tariff. Adding a second two-part tariff increases revenue to only \$602.7. Hence, our finding that menus of two-part tariffs are ineffective price discriminating tools in this market is robust to the inclusion of additional consumer heterogeneity via random coefficients.

7 Conclusion

Comparing Table 7 to the first row of Table 6 indicates that predicted revenues using the parameter estimates are much higher with the optimal flat tariff than with the actual menu offered to consumers (\$998.1 compared to \$472.9). We would not, however, necessarily suggest that the firm should have offered this flat tariff, nor the slightly better two-part tariff with its high fee component. As discussed in the previous section, the optimal tariff implied by the base model is sensitive to the nature of switching costs. In particular, even if switching costs are only occasionally zero (or very low), per-use pricing dominates a flat fee.

The parameter heterogeneity of the random coefficients model, particularly pertaining to switching costs, results in more realistic predictions than the base model. For example, the base model implies optimal flat fees that are just low enough that no consumers quit, whereas many consumers are predicted to quit upon revising beliefs downward in the random coefficients model. We therefore put more stock in the optimal tariffs reported in Table 10 for the random coefficients model. In particular, this model suggests a flat tariff of $F =$\$3.28 or an optimal single two-part tariff with $F =$\$3.18 and $p =$\$1.70. Either of these tariffs are predicted to yield approximately 25 percent more revenue than the actual tariff offered—a reasonable gain given our advantage of hindsight.

Interestingly, Peapod, the largest online grocer in the United States, now serves 250,000 customers and offers a per-use price of \$6.95 per delivery of orders over \$100. Our analysis supports its decision not to offer a menu of two-part tariffs. Furthermore, its price of \$6.95 is consistent with our estimates, which imply an optimal per-use price ranging from \$6.12 to \$6.78 across various specifications. The fact that it does not use flat fees may reflect reduced uncertainty in this market relative to the pre-2000 period that our data covers.

[45] The flat fees are notably lower for the random coefficients model compared to the base model. Since switching costs vary across consumers, the flat tariff's F is no longer simply $(1-\beta)\delta/\alpha$. These heterogeneous switching costs are primarily responsible for the lower revenue as well.

Regarding future work, our finding that tariff menus are ineffective for experience goods suggests that extending the model to allow menus to vary over time would be fruitless. Allowing a single two-part tariff to be dynamic when consumers' demands differ both ex-ante and ex-post may, however, be worthwhile.

References

Ackerberg, D. (2003), "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44(3), 1007–1040.

Ackerberg, D. (2002), "A new use of importance sampling to reduce computational burden in simulation estimation," manuscript UCLA.

Bergemann, D. and J. Välimäki (2004), "Monopoly Pricing of Experience Goods", Cowles Foundation Discussion Paper No. 1463.

Boatwright, P., S. Borle, and J. Kadane (2003) "A Model of the Joint Distribution of Purchase Quantity and Timing," *Journal of American Statistical Association*, 98, 564–572.

Bolton, P. and C. Harris (1999), "Strategic Experimentation," *Econometrica*, 67, 349–374.

Crawford, G. and M. Shum (2005), "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, forthcoming.

DeGroot, M. (1970), *Optimal Statistical Decisions*, New York, McGraw Hill.

DellaVigna, S. and U. Malmendier (2005) "Paying Not to Go to the Gym,", *American Economic Review*, forthcoming.

Eckstein, Z., D. Horsky, and Y. Raban (1988), "An Empirical Dynamic Model of Optimal Brand Choice," Working Paper, Tel-Aviv University.

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

Hajivassiliou, V.A. and Ruud, P.A. (1994), "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics, Volume IV*, ed. by R.F. Engel and D.L. McFadden, Amsterdam: Elsevier Science.

Judd, K. (1998), "Numerical Methods in Economics," Cambridge, Massachusetts, MIT Press.

McFadden, D. and K. Train (1996), "Consumers' Evaluation of New Products: Learning from Self and Others," *Journal of Political Economy*, 104(4), 683–703.

Milgrom, P. and J. Roberts (1986), "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, 94(4), 796–821.

Miller, R. A. (1984), "Job Matching and Occupational Choice," *Journal of Political Economy*, 92(6), 1086–1120..

Miravete, E. (1996), "Screening Consumers through Alternative Pricing Mechanisms," *Journal of Regulatory Economics*, 9, 111–132.

Miravete, E. (2002), "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans," *The Review of Economic Studies*, 69, 943–971.

Miravete, E. (2003), "Choosing the Wrong Calling Plan? Ignorance and Learning," *American Economic Review*, 93, 297–310.

Miravete, E. (2004), "The Limited Gains from Complex Tariffs," manuscript, CEPR DP No. 4237.

Nelson, P. (1970) "Information and Consumer Behavior," *Journal of Political Economy*, 78, 311–329.

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

Shapiro, C. (1983), "Optimal Pricing of Experience Goods," *Bell Journal of Economics*, 14(2), 497–507.

Table 1: Menu of Subscription Plans (i.e., Tariff Choices)

Plan #	flat fee F	per-use price p	initial plan shares	usage rates for which plan is minimum cost	mean observed usage	share never order
Plan 1	\$5.76	\$0	.12	.67–1	.56	.12
Plan 2	\$1.14	\$6.95	.32	.23–.67	.36	.18
Plan 3	\$0	\$11.95	.56	0–.23	.20	.57

Weekly F is the quoted monthly fee (\$24.95, \$5.00, \$0) divided by 4.33.
Expected cost per use is $p + F/(\text{usage rate})$.
Each consumer's observed usage rate ignores weeks beyond the last order.

Table 2: Demographic Characteristics, by Plan

characteristic	Plan 1	Plan 2	Plan 3
share all demographics missing	27.3	33.5	66.5
share no demographics missing	8.9	5.6	2.4
share income missing	60.3	61.3	80.4
share income > 90k	38.2	30.7	23.2
share income 50–90k	45.2	42.4	49.2
share income < 50k	16.6	26.9	27.5
mean # adults	2.1	2.0	2.0
mean # children	1.9	1.4	1.3
mean week enrolled	24.0	23.2	21.3
share female	75.4	70.8	68.5
share married	89.5	79.4	76.1
share co-habit	3.1	5.9	5.5
share single	7.4	14.7	18.3
share age 18–24	0.3	3.1	2.6
share age 25–44	35.5	38.6	37.0
share age 35–49	58.5	49.3	50.0
share age 50–64	5.7	7.3	8.4
share age 65+	0.0	1.7	2.0
share some HS	0.3	0.3	1.1
share graduate HS	6.6	10.1	10.8
share some College	19.7	25.0	31.2
share graduate College	49.6	43.1	36.7
share some Grad School	23.8	21.4	20.1
share fulltime out	66.8	70.2	72.0
share parttime out	14.5	10.5	11.0
share fulltime in home	14.5	13.5	10.6
share student	0.9	1.8	0.9
share retired/other	3.4	4.0	5.6
share full out spouse	89.0	87.5	86.9
share part out spouse	3.4	4.3	3.6
share full home spouse	3.7	4.1	2.6
share student spouse	0.6	0.8	1.9
share retired/other spouse	3.4	3.3	5.0

Table 3: Tariff choice thresholds and usage by myopic consumers

α	Plan 3 max m	Plan 1 min m	Plan 3 max usage	Plan 2 min usage	Plan 2 max usage	Plan 1 min usage
1.0000	5.76	5.76	0.00			1.00
0.9000	5.18	5.18	0.00			0.99
0.8000	4.61	4.61	0.01			0.99
0.7000	4.03	4.03	0.01			0.98
0.6000	3.45	3.45	0.02			0.97
0.5000	2.87	2.87	0.04			0.95
0.4000	2.29	2.29	0.08			0.91
0.3000	1.57	1.78	0.12	0.37	0.43	0.86
0.2000	0.65	1.41	0.15	0.32	0.50	0.80
0.1000	-0.28	1.04	0.19	0.27	0.58	0.74
0.0100	-1.12	0.72	0.22	0.23	0.66	0.67
0.0010	-1.21	0.69	0.23	0.23	0.66	0.67
0.0001	-1.22	0.68	0.23	0.23	0.66	0.66

"Plan 3 max m" is the highest belief m for which plan 3 is optimal.
"Plan 1 min m" is the lowest belief m for which plan 1 is optimal.
Plan 2 usage is not reported when plan 2 is sub-optimal for all m.

Table 4: Parameter Estimates

Parameter		Myopic Model	Dynamic Model	Dynamic w/ Random θ_i mean θ_i	std.dev. θ_i
μ_{G_μ}	(mean	-0.473	-2.190	-2.180	1.765
	match quality)	(.025)	(0.021)	(0.075)	(0.068)
σ_{G_μ}	(std. dev.	1.146	2.136		
	match quality)	(0.017)	(0.018)		
σ_0	(initial uncertainty)	6.664	4.998	5.253	1.736
		(0.096)	(0.031)	(0.061)	(0.057)
σ_ϵ	(experience	5.200	5.388	5.639	1.938
	signal precision)	(0.054)	(0.035)	(0.075)	(0.055)
β	(weekly	0	0.973	0.965	0.012
	discount factor)		(0.001)	(0.001)	(0.001)
α	(price coefficient)	-0.287	-0.284	-0.292	0.106
		(0.001)	(0.001)	(0.005)	(0.003)
δ	(switching cost)	1.778	50.030	34.897	11.555
		(0.003)	(0.029)	(0.602)	(0.461)
Λ	(initial tariff,	0.546	0.036	0.083	0.050
	$V - \alpha F_{s_i 0}$ coeff.)	(0.035)	(0.002)	(0.008)	(0.011)
λ_1	(initial tariff,	-0.828	0.582	0.313	0.261
	plan 1 intercept)	(0.079)	(0.051)	(0.041)	(0.034)
λ_2	(initial tariff,	-0.516	-0.178	-0.043	0.139
	plan 2 intercept)	(0.036)	(0.032)	(0.025)	(0.027)
Log likelihood		–55768.4	–54689.1	–54264.4	

σ_{G_μ} is redundant with random coefficients since μ_{G_μ} already varies across consumers.
Asymptotic standard errors are in parentheses.

Table 5: Posterior Beliefs ($\sigma_0 = 4.998$, $\sigma_\epsilon = 5.388$)

cumulative usage	Posterior standard deviation	Posterior mean w/ $+\sigma_0$ bias
0	4.998	4.998
1	3.664	2.686
2	3.030	1.837
3	2.641	1.396
4	2.371	1.125
5	2.170	0.943
6	2.013	0.811
7	1.886	0.712
8	1.780	0.634
9	1.690	0.572
10	1.613	0.521
11	1.545	0.478
12	1.485	0.442
13	1.432	0.410
14	1.384	0.383
15	1.340	0.360
16	1.300	0.339
17	1.264	0.320
18	1.231	0.303
19	1.200	0.288
20	1.171	0.275
25	1.053	0.222
30	0.965	0.187
35	0.896	0.161
40	0.840	0.141
45	0.793	0.126
50	0.753	0.114
55	0.719	0.104
60	0.689	0.095
65	0.662	0.088
70	0.639	0.082

Last column uses $\mu_i = 0, m_{i0} = \sigma_0$.

Table 6: Effect of mistakes, switching costs, and uncertainty

Model Description	Usage: initial, final (Plan share: initial, final) Plan 1	Plan 2	Plan 3	Revenue discounted ($\frac{Rev_{final}}{1-\beta_{firm}}$)	CS realized (expected)
Using estimates	.812, .622	.454, .067	.240, .014	472.9	-45.9
	(.129, .049)	(.321, .321)	(.551, .630)	(448.4)	(118.5)
No mistakes	.993, .615	.723, .084	.056, .008	499.7	-45.4
(i.e., optimal s_{i0})	(.184, .090)	(.222, .222)	(.593, .688)	(476.7)	(139.0)
No switching costs	.945, .930	.404, .408	.016, .011	193.2	-19.5
($\delta = 0$, optimal s_{i0})	(.540, .030)	(.057, .012)	(.403, .958)	(171.0)	(159.6)
No uncertainty	.915, .915	.401, .401	.012, .012	182.7	6.0
($\sigma_0 = 0$, optimal s_{i0})	(.032, .032)	(.012, .012)	(.956, .956)	(182.7)	(6.0)

Values in parentheses correspond to the label in parentheses in the column header.
All revenue and surplus values are in dollars per consumer.
Weekly $\beta_{firm} = .997976$. Hence, one dollar per week has present value of \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
Optimal s_{i0} indicates the initial plan choice is optimal (i.e., no ex-ante mistakes).
All values generated by simulating 100,000 consumers over 100 weeks.

Table 7: Optimal Tariffs: Base Model

Tariff Description	Usage: initial, final (Plan share: initial, final) Plan 1	Plan 2	Plan 3	Revenue discounted ($\frac{Rev_{final}}{1-\beta_{firm}}$)	CS realized (expected)
$F_1 = 4.85$, $p_1 = 0$	.929, .298			998.1	-59.5
(flat fee tariff)	(.417, .417)			(998.1)	(138.3)
$F_3 = 0$, $p_3 = 6.12$			.426, .071	232.4	-9.6
(per-use tariff)			(1.0 , 1.0)	(215.4)	(164.5)
$F_2 = 4.85$, $p_2 = .85$		.927, .272	.	1005.0	-60.8
(1 two-part tariff)		(.400, .400)		(1003.7)	(129.6)
$F_1 = 4.85$, $p_1 = .85$	.930, .273	.801, .129		1009.7	-61.4
$F_2 = 4.84$, $p_2 = 5.11$	(.380, .380)	(.020, .020)		(1008.0)	(127.8)
(2 two-part tariffs)					

Values in parentheses correspond to the label in parentheses in the column header.
All revenue and surplus values are in dollars per consumer.
Weekly $\beta_{firm} = .997976$. Hence, one dollar per week has present value of \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
All values generated by simulating 100,000 consumers over 100 weeks.
For ease of comparison, single tariff "menus" also appear as "Plan 2" or "Plan 3."
An additional ex-post only tariff offered no advantage over 2 two-part tariffs.
Plan 2 in the last row was only chosen by consumers making mistakes.

Table 8: Optimal Tariffs: Base Model with No Mistakes

Tariff Description	Usage: initial, final (Plan share: initial, final) Plan 1	Plan 2	Plan 3	Revenue discounted ($\frac{Rev_{final}}{1-\beta_{firm}}$)	CS realized (expected)
$F_1 = 4.85$, $p_1 = 0$	.932, .297			998.1	-59.6
(flat fee tariff)	(.417, .417)			(998.1)	(138.3)
$F_3 = 0$, $p_3 = 6.51$			.420, .066	230.9	-10.6
(per-use tariff)			(1.0 , 1.0)	(213.1)	(160.3)
$F_2 = 4.85$, $p_2 = 1.15$		.930, .263	.	1005.9	-61.2
(1 two-part tariff)		(.395, .395)		(1004.0)	(126.5)

Values in parentheses correspond to the label in parentheses in the column header.
All revenue and surplus values are in dollars per consumer.
Weekly $\beta_{firm} = .997976$. Hence, one dollar per week has present value of \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
All values generated by simulating 100,000 consumers over 100 weeks.
For ease of comparison, single tariff "menus" also appear as "Plan 2" or "Plan 3."
2 two-part tariffs offered no advantage over 1 two-part tariff.

Table 9: Optimal Tariffs: Base Model with Random Switching Costs (Prob($\delta_{it} = 0$) = .1)

Tariff Description	Usage: initial, final (Plan share: initial, final) Plan 1	Plan 2	Plan 3	Revenue discounted ($\frac{Rev_{final}}{1-\beta_{firm}}$)	CS realized (expected)
$F_1 = 3.09$, $p_1 = 0$	.829, .794			155.8	-21.9
(flat fee tariff)	(.654, .082)			(125.0)	(184.9)
$F_3 = 0$, $p_3 = 6.12$			.426, .071	232.4	-9.6
(per-use tariff)			(1.0 , 1.0)	(215.4)	(164.5)
$F_2 = .03$, $p_2 = 6.11$		.558, .139	.	235.4	-10.4
(1 two-part tariff)		(.764, .500)		(217.7)	(163.8)
$F_1 = .58$, $p_1 = 4.54$	.636, .508	.543, .056		240.8	-13.2
$F_2 = .05$, $p_2 = 7.03$	(.318, .110)	(.430, .284)		(219.9)	(169.3)
(2 two-part tariffs)					
$F_1 = .59$, $p_1 = 4.54$	.486, .513	.724, .066	.086, .002	243.0	-13.7
$F_2 = .05$, $p_2 = 7.03$	(.416, .108)	(.260, .236)	(.324, .656)	(221.4)	(167.9)
$F_3 = 0$, $p_3 = 10.78$					

Values in parentheses correspond to the label in parentheses in the column header.
All revenue and surplus values are in dollars per consumer.
Weekly $\beta_{firm} = .997976$. Hence, one dollar per week has present value of \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
All values generated by simulating 100,000 consumers over 100 weeks.
For ease of comparison, single tariff "menus" also appear as "Plan 2" or "Plan 3."

Table 10: Optimal Tariffs: Random Coefficients Model

Tariff Description	Usage: initial, final (Plan share: initial, final) Plan 1	Plan 2	Plan 3	Revenue discounted $(\frac{Rev_{final}}{1-\beta_{firm}})$	CS realized (expected)
$F_1 = 3.28$, $p_1 = 0$	.877, .293			587.2	-44.5
(flat fee tariff)	(.515, .358)			(579.5)	(879.3)
$F_3 = 0$, $p_3 = 6.78$			.401, .056	206.8	-12.0
(per-use tariff)			(1.0 , 1.0)	(189.1)	(784.7)
$F_2 = 3.18$, $p_2 = 1.70$		.868, .237		601.9	-47.0
(1 two-part tariff)		(.485, .333)		(590.3)	(809.5)
$F_1 = 3.66$, $p_1 = 1.21$	.884, .282	.864, .230		602.7	-47.7
$F_2 = 3.14$, $p_2 = 1.74$	(.119, .075)	(.366, .253)		(590.8)	(818.1)

Values in parentheses correspond to the label in parentheses in the column header.
All revenue and surplus values are in dollars per consumer.
Weekly $\beta_{firm} = .997976$. Hence, one dollar per week has present value of \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
Values generated by simulating 5000 consumers over 100 weeks for each of 100 draws of θ.
For ease of comparison, single tariff "menus" also appear as "Plan 2" or "Plan 3."

Figure 1: **Expected Cost per Delivery**

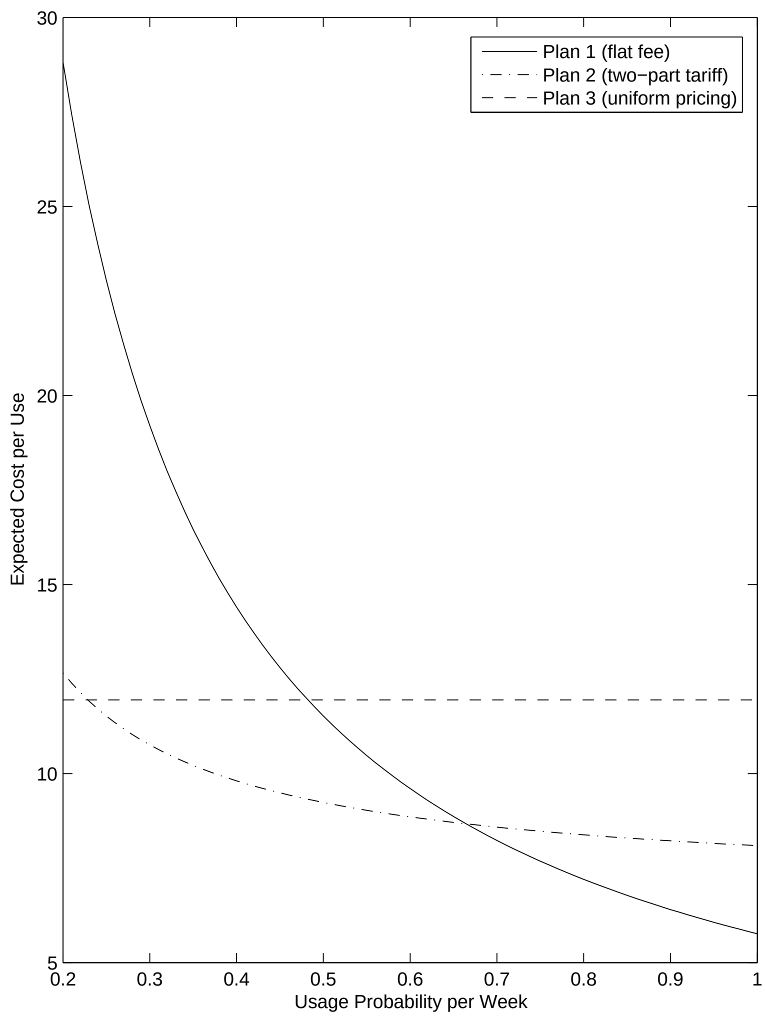


Figure 2: **Usage Rates**

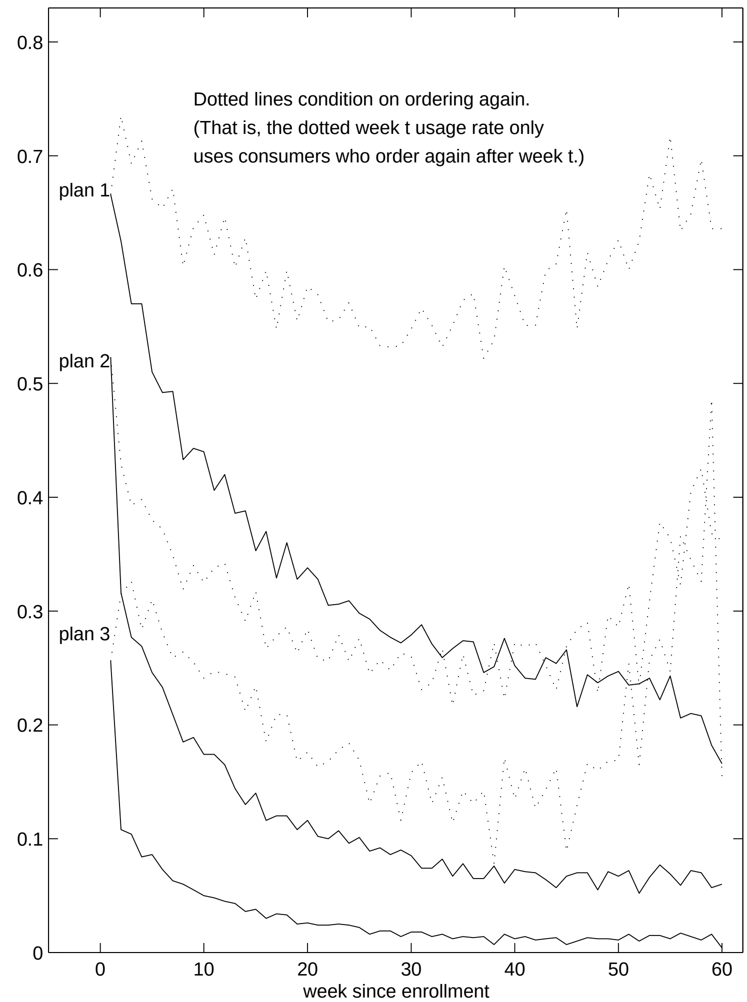


Figure 3: **Simulated Usage Rates**

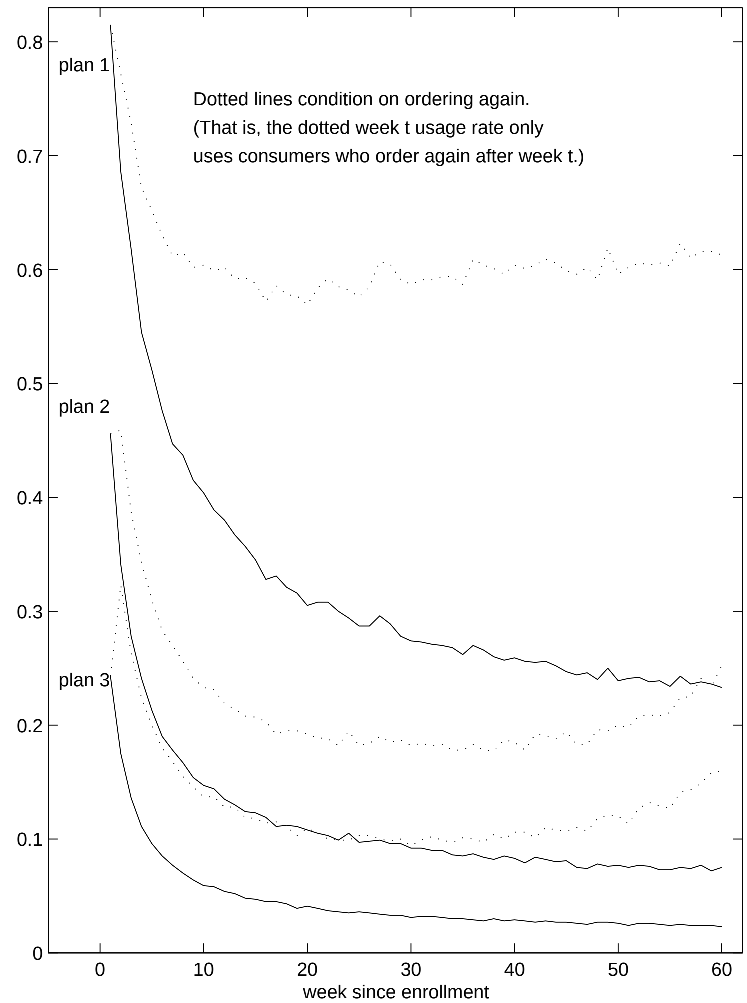
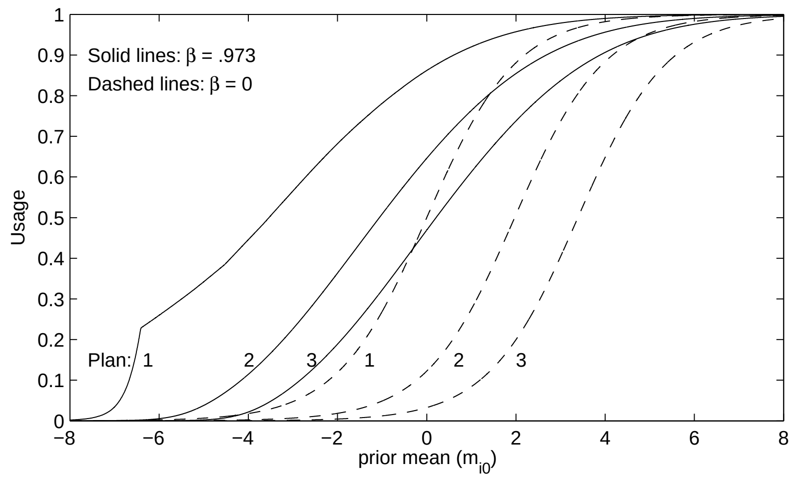


Figure 4: **Information Acquisition**



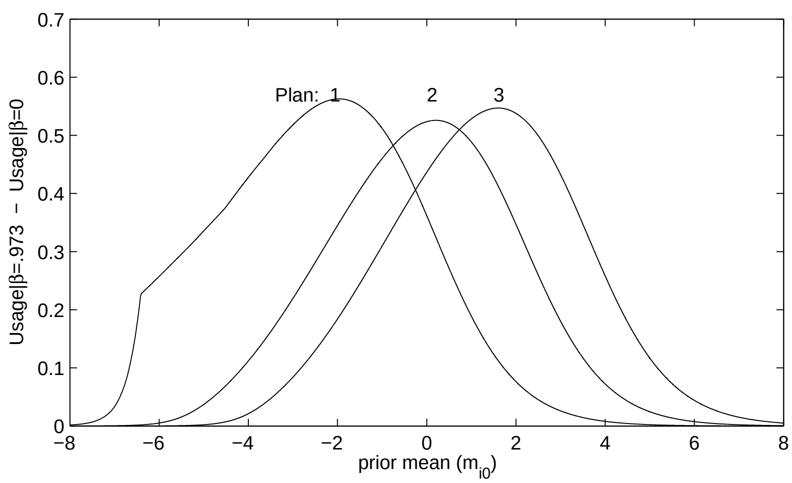


Figure 5: **Tariffs and Revenues as Functions of Initial Uncertainty (σ_0)**

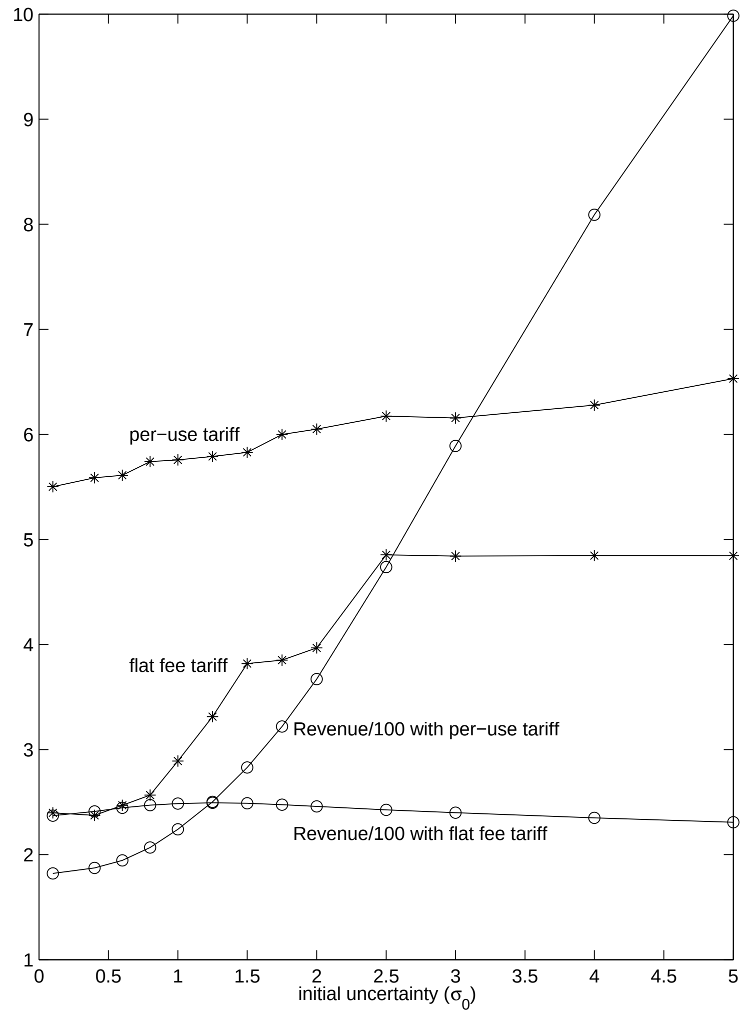


Figure 6: **Tariffs and Revenues as Functions of Switching Costs (δ)**