Tariff Choice with Consumer Learning: Sorting-Induced Biases and Illusive Surplus*

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

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

August 2, 2007

*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

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 three elements that influence consumer behavior and affect pricing strategies: beliefs, experiential learning, and switching costs. We test the rational expectations assumption against a weaker assumption about beliefs that retains unbiasedness, on average across all consumers, but allows beliefs to be biased conditional on tariff choice. The rational expectations assumption is rejected: consumers on flat fee tariffs tend to have optimistic priors whereas consumers on per-use tariffs tend to have pessimistic priors. If switching costs are sufficiently high, this sorting-induced bias implies that flat fee tariffs can yield high profits for the firm even after optimistic consumers revise their beliefs. This dynamic lock-in effect of flat fees augments their traditional role in static settings of extracting consumer surplus. Biased priors also lead to biased expectations of consumer surplus. Realized surplus is on average negative, despite expectations of a large and positive discounted lifetime surplus. Regarding the use of tariff 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. For example, cellular phone and internet service providers offer an array of volume discounts, and cable and satellite television providers offer various bundles of programming. Consumers with high valuations of the product choose high quantity offers whereas consumers with low valuations select lower quantity offers. In many cases the tariff or subscription choice is complicated by uncertainty regarding the quality or value of the product. Furthermore, as consumers learn their valuations of the product, they may change their selection from the menu. Switching costs, due to either explicit charges or hassle costs, however, can lead consumers to remain on plans that are sub-optimal given their revised beliefs. These dynamic issues complicate the determination of optimal tariff offerings. For example, the fixed fee component of a two-part tariff is no longer merely a device to extract consumer surplus, as in the static setting. With consumer uncertainty and switching costs, the fixed fee is also a device to "lock-in" consumers to paying for the product even after they reduce their valuation of it. Of course, consumers are aware of their uncertainty and the potential costs of switching, and choose accordingly.

In this paper we study consumer behavior and its implications for pricing strategies when consumers are uncertain about their valuations of a nondurable good or service. In particular, we assess the effectiveness of tariff menus as screening mechanisms to price discriminate among heterogeneous consumers. Our empirical focus is the online grocer market in which consumers order groceries online to be delivered to their home.[1] Since this service is new, consumer uncertainty is particularly relevant. Furthermore, this service is an "experience good," in the terminology of Nelson (1970), since the uncertainty is primarily resolved by using the service. As such, forward looking consumers have an informational incentive to experiment with the service, which leads them to trade off current utility for future utility. Though our data describe behavior for an experience good, our results also apply to non-experience goods for which uncertainty is resolved independently of consumption choices.[2]

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. 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 was fixed over the seventy-week period of our data. The model we estimate highlights

[1] This growing industry had revenues of $2 billion in 2006, which is a small fraction of the $500 billion supermarket industry.

[2] Courty and Hao (2000) and Miravete (1996,2002) investigate screening mechanisms for non-experience goods. DellaVigna and Malmendier (2005) investigate tariff choice for health clubs (arguably an experience good) but not in the context of learning.

three elements that influence consumer behavior and affect pricing strategies: beliefs, experiential learning, and switching costs.[3]

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.[4] 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 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 could be interpreted as consumers acting irrationally or not having 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 they have 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 impose a weaker informational assumption than the rational expectations requirement that consumers know the distribution of their match-value. In the Bayesian learning model, each consumer's initial belief is a weighted average of the population mean and her unbiased private signal. The respective weights are the variance of the private signal and the population variance of match-values.[5] Our econometric specification of beliefs deviates from rational

[3]Narayanan, et.al (2007) study tariff choice and consumer learning about telephone usage rates in a model with no switching costs.

[4]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.

[5]The precision of the private signal reflects the degree to which the consumer knows how she differs from the

expectations by allowing the perceived dispersion of match-values to exceed the true dispersion in the population. As such, consumers will place more weight on their private signals than if they knew the true population variance. Hence, consumers receiving signals above their match-value will have optimistic initial beliefs, whereas consumers receiving signals below their match-values will have pessimistic initial beliefs. The key advantage of this assumption for explaining our data is that despite being an unbiased signal for each consumer, the private signal induces a bias in beliefs when conditioning on the consumer's tariff choice. For example, consumers with optimistic priors tend to have high initial beliefs and therefore tend to choose tariffs with high flat fees and low per-use prices. The notion that consumers who choose high fee tariffs are likely to have received optimistic signals is quite intuitive. The fact that this intuition is incompatible with rational expectations, however, has yet to be explored.

The second puzzle—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 \$372 higher on an alternative plan. Given our estimated discount factor, this cost is equivalent to \$4.84 per week.

One interpretation of the third puzzle—that some consumers sign up for plans with fixed fees and never use the service—is that some consumers simply make mistakes when enrolling. An alternative interpretation is that consumers receive information after enrolling which leads them to revise their beliefs.[6] For example, consumers may begin the process of ordering only to conclude that the procedure is too complicated. Hence, our econometric specification of beliefs includes a post-enrollment signal that allows for such behavior.

In short, we offer a model that explains these three puzzles while maintaining rational behavior. We emphasize that rejecting the informational assumption of rational expectations is not a rejection of rationality itself. The consumers in our model behave optimally and update beliefs optimally, given the information they have: they simply do not have as much initial knowledge as conjectured under the rational expectations hypothesis. 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. Furthermore, our model generates the biases needed to explain our data without relying on aggregate biases (i.e., across all consumers). Of course, if a given dataset is consistent with rational expectations, our model of learning and dynamic tariff choice is still applicable since rational expectations is nested within our framework.

Using our estimated model of consumer behavior, we numerically solve for optimal tariffs that

average consumer. For example, if the private signal's variance is high relative to the population variance, then the initial belief will be near the population mean.

[6]Without explicitly illiciting beliefs, one would have difficulty supporting a claim that consumers make ex-ante mistakes.

maximize discounted profits for a variety of scenarios.[7] 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. Since we focus on consumer uncertainty and learning, we assume the firm knows the distribution of consumers' match-values and the other demand parameters.

Given estimates of high switching costs and consumer uncertainty, our model suggests that a firm should offer 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. This result, however, hinges on switching costs always being high. When we simulate the model with switching costs randomly being zero on occasion, the lock-in benefit of fixed fees is sufficiently reduced such that ex-post (i.e., per use) pricing yields higher profits.

Consistent with previous studies we find that the use of tariff menus as a screening device to price discriminate is largely ineffective. In particular, adding a second two-part tariff increases discounted profits marginally and adding a third tariff (per-use only) offers no additional gain. Miravete (2002,2004) also 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.

In section 2 we present our model of consumer learning. In section 3 we discuss the data. 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 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.[8] 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

[7]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)). As mentioned earlier, we focus on tariffs that are fixed over time.

[8]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.

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).[9] To be incentive compatible, the total payment when the 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.[10]

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 I_{it}[11]:

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

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

[9] 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.

[10] 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.

[11] 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.

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:

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

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 after an experience signal $\mu_i + \epsilon_{it}$ is given by

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

where

$$m_{it} = \frac{\sigma_\epsilon^2 m_{i0} + \sigma_0^2(\mu_i + \epsilon_{it})}{\sigma_\epsilon^2 + \sigma_0^2}, \tag{7}$$

and

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

This model of Bayesian learning is tractable because all 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. Since the posterior belief is also a normal distribution, the updating triggered by the next realized signal follows the same process, with this posterior serving as the prior. Note that the posterior variance deterministically shrinks, regardless of the signal's realized value. Also, if σ_ϵ is zero all uncertainty is resolved after the first 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

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

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 $c_{it} = 1$:[12]

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

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

where γ is Euler's constant, and $m_{it+1}(\cdot)$ and $\sigma_{it+1}(\cdot)$ govern the updating of the posterior mean and variance using (7) and (8). The perceived distribution of experience signals Φ is normal with mean m_{it} and variance $\sigma_{it}^2 + \sigma_\epsilon^2$ to account for both the noise in the signal and the uncertainty of current beliefs.

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 analytic solution, we can solve it numerically. Since Φ is $N(m_{it}, \sigma_{it}^2 + \sigma_\epsilon^2)$, the integral over the signal 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. 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.[13]

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.

[13]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 5333 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).[14] 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.[15] 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. Week t=1 for each consumer is the week beginning with her enrollment date. The usage

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

[15] 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 (since initial beliefs and uncertainty do not appear to be related to the week of enrollment). Also note that we only use consumers who enroll at least ten weeks prior to the end of the dataset.

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. Otherwise, c_{it} is set to 0 to indicate that only the traditional grocer was used. Our focus on weekly usage is motivated by the fact that only 2.8% of customers' weeks with orders have more than one order.

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.[16] Somewhat surprisingly, in our data we never observe consumers who switch plans and order after the switch.[17] 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. The model estimates imply that a substantial number of consumers indeed "quit" by switching to the no fee plan.

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 Consumer behavior and tariff choice

The online grocer in our market offered consumers a menu of three tariffs.[18] 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.[19] 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 12.3 percent of consumers signed up for plan 1, compared to 31.7 percent for plan 2 and 56.1 percent for plan 3. The mean usage rate was .60 for plan 1 enrollees, .38 for plan 2 enrollees, and .21 for plan 3 enrollees. Each consumer's usage rate is computed from the weeks spanning enrollment and 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.

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

[17] The online grocer claims that customers even ignored letters explaining that they should switch to a different plan. The model nonetheless allows for switches since the absence of switching is an endogenous outcome.

[18] 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.

[19] 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.

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 at most .60 (since the reported usage rate ignores weeks after the last observed order). Furthermore, some consumers enroll on plans with fixed fees, but then never try the service. 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 10 weeks before the end of our sample period.

Rather than assume these consumers are behaving suboptimally or irrationally, our model rationalizes their behavior as being optimal given the available information. We propose a specification for initial beliefs that relaxes the often used assumption that consumers know the distribution from which their match-values arise. Relaxing this information assumption enables the model to predict the surprisingly low average usage rates for consumers on the high fee plan, but does not, by itself, predict that some consumers on the high fee plan never order. Hence, we also assume consumers receive a signal after enrollment regardless of whether they order. This signal can reflect their experience installing the software or navigating the company's website for ordering. Consumers who receive a low post-enrollment signal may choose to quit before even trying the service. The details of this specification and the initial tariff choice appear in the next section.

Our model and information assumptions are also designed to capture the evolution of usage rates as consumers learn their match-values. 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 74 percent of the plan 1 enrollees used the online grocer during their first week. Their usage declined steadily to around .20 by 60 weeks after enrollment.[20] 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 (assuming beliefs are symmetric). 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-

[20]The usage measures become noisy as "weeks since enrollment" increases since few consumers signed up early enough to provide such data.

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 increase at the end of the sample period.

We will compare Figure 2 to similar graphs based on data simulated by various specifications of the model as an eyeball measure of goodness of fit.

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 specify initial beliefs and the consumer's initial tariff choice. We then 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 simulated maximum likelihood.

4.1 Initial beliefs and tariff choice

The standard bayesian learning model with rational expectations assumes consumers know the normal distribution from which match-values are drawn,

$$\mu_i \sim N(\mu_{pop},\ \sigma^2_{pop}), \tag{11}$$

where the subscript *pop* refers to population. In the absence of additional signals, each consumer's prior for her own match-value would be $N(\mu_{pop}, \sigma^2_{pop})$.

To enable consumers to have different ex-ante beliefs, we assume each receives an unbiased private signal $(\mu_i + \epsilon_{i,pre}) \sim N(\mu_i, \sigma^2_{pre})$, where the subscript *pre* indicates the signal is pre-enrollment. She then updates her prior to

$$\mu_i \sim N(m_{i0},\ \sigma^2_{i0}), \tag{12}$$

where

$$m_{i0} = \frac{\sigma^2_{pre}\mu_{pop} + \sigma^2_{pop}(\mu_i + \epsilon_{i,pre})}{\sigma^2_{pop} + \sigma^2_{pre}}, \tag{13}$$

$$\sigma^2_{i0} = \frac{\sigma^2_{pre}\sigma^2_{pop}}{\sigma^2_{pre} + \sigma^2_{pop}}\ , \tag{14}$$

and the subscript $i0$ refers to the period before i enrolls. Hence, each consumer's belief prior to the initial plan choice is normally distributed with a mean that is a weighted average of the population mean and the unbiased private signal, with weights determined by the variances.

We modify this belief structure by allowing consumers to increase or decrease the relative weight on their private signal. In particular, we suppose consumers perceive the population variance to be $\tilde{\sigma}^2_{pop} \neq \sigma^2_{pop}$. Replacing σ^2_{pop} with $\tilde{\sigma}^2_{pop}$ in equations 13 and 14 yields the pre-enrollment beliefs under this alternative information assumption. We expect markets for new products and services, such as the online grocer, to be characterized by $\tilde{\sigma}^2_{pop} > \sigma^2_{pop}$ since consumers are less likely to have precise information about the population distribution of match-values.

A special case of this structure, $\tilde{\sigma}^2_{pop} = \infty$, corresponds to consumers having no information about the population distribution of match-values. In this case pre-enrollment beliefs are centered around the private signal: $\mu_i \sim N(\mu_i + \epsilon_{i,pre},\ \tilde{\sigma}^2_{pre})$. More generally, the consumer's information about the population distribution is decreasing in $\tilde{\sigma}^2_{pop}$. The sorting-induced biases that enable us to rationalize the data are generated by $\tilde{\sigma}^2_{pop} > \sigma^2_{pop}$, which places less weight on μ_{pop} when updating beliefs with the private signal.[21]

Given pre-enrollment beliefs, the consumer simply chooses the initial tariff s_{i0} that maximizes her expected discounted utility. Following enrollment, however, consumers obtain an unbiased signal of their match-value. The source of this information could be their experience with installing the software for placing orders or perusing the items available to purchase. To account for the anticipated post-enrollment signal, we must integrate over the possible signal values $\mu_{i0} = \mu_i + \epsilon_{i,post}$ and consider the option of switching tariffs (i.e., $s_{i1} \neq s_{i0}$) given the realized signal. That is, the

[21] We could further generalize the information structure to allow an aggregate bias in beliefs by considering a perceived $\tilde{\mu}_{pop} \neq \mu_{pop}$. Since we found this additional generalization to be rejected by the data, we do not include it in the specification. That is, conditional on tariff choice we find significant biases, but no aggregate bias in beliefs when averaging across all consumers (i.e., regardless of tariff choice).

optimal initial tariff solves[22]

$$\max_{s_{i0}\in\{1,2,3\}} \int \max_{s_{i1}} \Big\{ V(m_{i1}(\mu_{i0}), \sigma_{i1}, s_{i1}) + \alpha F_{s_{i1}} + \delta_{i0}\mathcal{I}(s_{i1} \neq s_{i0}) \Big\} G_\delta(d\delta_{i0})\Phi(d\mu_{i0}|m_{i0}, \sigma_{i0}), \quad (15)$$

where $m_{i1}(\mu_{i0}) = \frac{\sigma^2_{post} m_{i0} + \sigma^2_{i0}\mu_{i0}}{\sigma^2_{post}+\sigma^2_{i0}}$ is the posterior mean and $\sigma^2_{i1} = \frac{\sigma^2_{post}\sigma^2_{i0}}{\sigma^2_{post}+\sigma^2_{i0}}$ is the posterior variance. The perceived distribution of signals Φ is normal with mean m_{i0} and variance $\sigma^2_{post} + \sigma^2_{i0}$ to account for both the noise in the signal and the uncertainty of current beliefs. Recall that G_δ allows for switching costs to randomly vary from week to week.

After choosing s_{i0} the consumer realizes the post-enrollment signal $(\mu_i + \epsilon_{i,post})$, updates her beliefs to $\mu_i \sim N(m_{i1},\ \sigma_{i1})$, and then chooses her continuation plan s_{i1} according to the maximization embedded in the integral in equation 15. Her subsequent behavior is governed by the value function in equation 10 with $t = 1$.

As we demonstrate in the next section, this specification of initial beliefs with pre- and post-enrollment signals and incomplete information about the population distribution of match-values enables us to rationalize the seemingly irrational consumer behavior without resorting to "mistakes" on the part of consumers or aggregate biases in beliefs.

4.2 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 now provide more detail regarding the identification of particular parameters.

The match-value distribution parameters μ_{pop} and σ^2_{pop} are identified by consumers' usage rates at the end of their histories—after experience signals have eliminated much of the uncertainty. The degree to which $\tilde{\sigma}^2_{pop}$ exceeds σ^2_{pop} is governed by the degree to which sorting-induced biases are needed to fit the data. The marked drop in usage of consumers on plan 1 (with the high fee), relative to the change in usage of consumers on the other plans suggests this difference will be large. The variance of the pre-enrollment signal σ^2_{pre} is identified by the degree of ex-ante beliefs heterogeneity needed to fit variation in consumers' initial tariff choices and initial usage rates. The variance of the post-enrollment signal σ^2_{post} is identified by the degree to which consumers initial usage rates appear inconsistent with their tariff choice. For example, a (relatively) high σ^2_{post} enables the model to predict that some consumers who enroll on the high fixed fee plan will never use the service. The speed with which behavior adjusts over time 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

[22]The expression being maximized is similar to the last two lines of equation 10, which details the continuation value given an expected experience signal.

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 of nonusage 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.

In Figure 3 we use simulated data (for which we know match-values and beliefs) to depict the identification argument for α.[23] Through tariff choice consumers sort themselves: those with high beliefs choose the fixed fee of plan 1, consumers with moderately high beliefs choose the two-part tariff of plan 2, and consumers with low beliefs choose the per-use tariff of plan 3. Prior to observing the post-enrollment signal this sorting is strict, as depicted along the vertical axis (m_{i0}) of the top panel of Figure 3. Consumers with m_{i0} less than -2.98 initially subscribe to plan 3. Consumers with pre-enrollment beliefs greater than 3.7 initially choose plan 1. The remaining consumers with initial beliefs between -2.98 and 3.7 choose plan 2.

Notice, however, that the distributions of true match values (on the horizontal axis) conditional on each initial plan choice are dispersed over the whole range.[24] 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. The bottom plot of Figure 3 depicts this weakened sorting by plotting beliefs after the post-enrollment signal on the vertical axis. As can be seen in the graph, a given range of beliefs contains consumers on each of the plans. Hence, similar consumers face different prices, which enables the identification of price sensitivity.

4.3 The likelihood function

We estimate four econometric specifications of the model. The base specification assumes no parameter heterogeneity beyond the unobserved match-values and signals that have already been discussed. For comparison, we also estimate this specification with myopic consumers. To highlight the importance of allowing sorting-induced biases (by relaxing the rational expectations assumption), we also estimate the model with $\tilde{\sigma}^2_{pop} = \sigma^2_{pop}$ imposed. Finally, we estimate the forward-looking model with all parameters varying across consumers as random coefficients (except β which is fixed at the baseline model's estimate). This latter estimation uses the importance sampling methodology

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

[24] As expected, the mean match values are increasing in the fee of the plan: -3.9 for plan 3, -2.9 for plan 2, and -2.2 for plan 1. That is, consumers with high m_{i0} tend to have (relatively) high μ_i.

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 some consumers remain on the wrong plan forever.

Since consumers' match-values and beliefs are not observed we integrate over μ_i, $\epsilon_{i,pre}$, $\epsilon_{i,post}$, and ϵ_{it} (the experience signals) to obtain a likelihood function. Let θ denote the vector of parameters to estimate. In the base specification $\theta = (\mu_{pop}, \sigma_{pop}, \tilde{\sigma}_{pop}, \sigma_{pre}, \sigma_{post}, \sigma_\epsilon, \beta, \alpha, \delta)$.

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 $\tau_i - 1$ be the i^{th} consumer's last week with $c_{it} = 1$ (so τ_i is the first week with censored s). Integrating over the unobserved match-value and signals, the likelihood for consumer i is then

$$\begin{aligned} L_i(\theta) \;=\; & \int \Bigg[Pr(s_{i0}|m_{i0}(\epsilon_i,\mu_i),\sigma_{i0};\theta) \prod_{t=1}^{\tau_i-1} Pr(s_{it}|m_{it}(\epsilon_i,\mu_i),\sigma_{it},s_{it-1};\theta) Pr(c_{it}|m_{it}(\epsilon_i,\mu_i),\sigma_{it},s_{it};\theta) \\ & \sum_{s_{i\tau_i}} Pr(s_{i\tau_i}|m_{i\tau_i}(\epsilon_i,\mu_i),\sigma_{i\tau_i},s_{i\tau_i-1};\theta) \prod_{t=\tau_i}^{T_i} Pr(c_{it}|m_{it}(\epsilon_i,\mu_i),\sigma_{it},s_{i\tau};\theta) \Bigg] \Phi(d\epsilon_i, d\mu_i;\theta) \end{aligned} \tag{16}$$

where $m_{it}(\epsilon_i, \mu_i)$ makes explicit the dependence of beliefs on the unobserved match-value μ_i and signals $\epsilon_i = (\epsilon_{i,pre}, \epsilon_{i,post}, \{\epsilon_{it}\}_{t=1}^{\tau_i-1})$, The first line of this likelihood entails the probability of the weeks from enrollment to the last usage, while the second line entails the likelihood for the trailing weeks after the last usage. The summation in the second line integrates over the censored tariff choice $s_{i\tau}$ after the last usage. Since s_{it} only changes after signals are received $s_{it} = s_{i\tau}$ for $t \geq \tau$.

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 utility shock u_{it}, the value of choosing $c_{it} = 0$ is

$$V_{0it} = \beta \max_{s_{it+1}} \Big\{ V(m_{it}, \sigma_{it}, s_{it+1}) + \alpha F_{s_{it+1}} + \delta \mathcal{I}(s_{it+1} \neq s_{it}) \Big\} \tag{17}$$

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

$$\begin{aligned} V_{1it} \;=\; & m_{it} + \alpha p_{s_{it}} + \beta \int \max_{s_{it+1}} \Big\{ V(m_{it+1}(m_{it},\sigma_{it},\mu_{it}), \sigma_{it+1}(\sigma_{it}), s_{it+1}) + \\ & \alpha F_{s_{it+1}} + \delta \mathcal{I}(s_{it+1} \neq s_{it}) \Big\} \Phi(d\mu_{it}|m_{it},\sigma_{it}) \;. \end{aligned} \tag{18}$$

Both these equations are simplifications of the expressions for continuation values in (10) to account for the fixed, rather than random, switching costs δ. We can then write

$$Pr(c_{it}|m_{it},\sigma_{it},s_{it};\theta)=\frac{\exp(V_{c_{it}it})}{\exp(V_{0it})+\exp(V_{1it})}\ . \tag{19}$$

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 plan choice s_{it} is 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. Mathematically,

$$Pr(s_{it}|m_{it},\sigma_{it},s_{it-1};\theta)=\mathcal{I}\{s_{it}=s(m_{it},\sigma_{it},s_{it-1};\theta)\}, \tag{20}$$

where $s(m_{it},\sigma_{it},s_{it-1};\theta)\equiv\text{argmax}_s\{V(m_{it},\sigma_{it},s)+\alpha F_s+\delta\mathcal{I}(s\neq s_{it})\}$ denotes the optimal tariff that solves the maximization embedded in the continuation value integrands of both equations 15 and 10. Though this choice is deterministic given beliefs, from the econometrician's perspective s_{it} is probabilistic since beliefs are unobserved.

Now consider the integration over the censored plan choice for the trailing weeks $t\geq\tau_i$. Beliefs are fixed after the last usage since no more signals are received. Hence, the censored plan is simply $s(m_{i\tau_i},\sigma_{i\tau_i},s_{i\tau_i-1};\theta)$.[25] Integration over the censored plan is therefore automatically handled by the integration over unobserved beliefs.

We use monte carlo simulation with 1000 draws to evaluate $L_i(\theta)$ for each consumer. Our estimator is obtained by maximizing the product of the consumers' simulated likelihoods, using the nested fixed-point algorithm of Rust (1987).

Hajivassiliou and Ruud (1994) show that simulated maximum likelihood yields an inconsistent estimator for a fixed number of draws. To increase the efficiency of our 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.[26] In the absence of this sampling scheme, the contribution to the likelihood of many of the draws of (μ_i,ϵ_i) would be zero due to a zero probability of the tariff choice in equation 20. Sampling from truncated normals and reweighting also yields a smooth simulated likelihood function since the weights are smooth functions of the parameters. This smoothness aids in the numerical optimization and computing of standard errors.

[25] With fixed switching costs, as in the estimated specifications, plan changes in the model only occur immediately after beliefs are updated. 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 $1-Pr(\delta_1)^{t-\tau}$.

[26] Due to switching costs the set of experience signals for which a given plan is optimal may contain two non-contiguous regions, somewhat complicating this approach.

4.4 Exogenous quits and censored plans

Many households in the data use the online grocer regularly over a long period of time, and then suddenly stop. While such behavior could be explained by a slow learning process, it more likely reflects permanent household shocks, such as moving from the area, marriage, divorce, childbirth, retirement, etc. To account for such unobserved shocks, we assume that each household in each period *exogenously quits* with probability γ.[27]

To account for exogenous quits in the likelihood, we multiply, in the first line of equation 16, each $Pr(c_{it}|m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta)$ by $1-\gamma$ and replace $\prod_{t=\tau_i}^{T_i} Pr(c_{it}|m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{i\tau}; \theta)$ in the second line with

$$\sum_{t=\tau_i}^{T_i} \gamma^{I(t<T_i)} \left[(1-\gamma) Pr(c_{it}|m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{i\tau}; \theta)\right]^{t-\tau_i}. \tag{21}$$

Each trailing week's inactivity may be due to either an exogenous quit in the current period or a prior period, or due to choosing the traditional store despite still being a subscriber. Once the consumer exogenously quits, however, the subsequent periods of inactivity no longer represent observations. For example, if $T_i = 52$ and $\tau_i = 51$, the probability of these trailing weeks is the sum of the probability she exogenously quits at $t = 51$, the probability she chooses no usage at $t = 51$ and exogenously quits at $t = 52$, and the probability she chooses no usage at both $t = 51$ and $t = 52$. The expression in 21 sums the probabilities of these different explanations for the trailing weeks of non-usage.

We fix $\gamma = 0.003$ based on the frequency of household relocations in our market, which implies a .145 annualized exogenous quit rate. In principle, we could estimate γ. Given the absence of observed plan switches, estimating both switching costs and γ, however, is impractical since both are identified off the need to explain the trailing weeks with no orders.

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

[27] We model consumers as being unaware of the possibility of these shocks. That is, we do not modify the dynamic choice problem to account for a possible truncation of the infinite horizon. This is consistent with consumer's having a (biased) prior of $\gamma = 0$, or with expecting an online grocer in their new location (in the event of moving).

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,

$$L_i(\rho) = \int L_i(\theta_i) g(\theta_i|\rho) d\theta_i = \int L_i(\theta_i) \frac{g(\theta_i|\rho)}{h(\theta_i)} h(\theta_i) d\theta_i \tag{22}$$

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

$$\tilde{L}_i^{NS}(\rho) = \frac{1}{NS} \sum_{ns=1}^{NS} L_i(\theta_i^{ns}) \frac{g(\theta_i^{ns}|\rho)}{h(\theta_i^{ns})} \tag{23}$$

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 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.[28]

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 3 presents estimates for four specifications: the base model (with forward looking consumers and sorting-induced biases), the base model except with myopic consumers (i.e., $\beta = 0$), the base model except without sorting-induced biases (i.e., $\tilde{\sigma}_{pop} = \sigma_{pop}$), the base model with random coefficients on all parameters (except β which is fixed at .9856). Estimates are precisely estimated and are quite similar across all three models, with a few exceptions. Our discussion of the estimates

[28] For each consumer we use 50 draws of the parameters that affect the consumer's dynamic program (i.e., $\tilde{\sigma}_{pop}, \sigma_{pre}, \sigma_{post}, \sigma_\epsilon, \beta, \alpha, \delta$). For each vector of these dynamic parameters, we solve the dynamic program and evaluate the likelihood for 50 draws of μ_{pop}. This approach enables us to more efficiently integrate over the random μ_{pop} which does not enter the fixed point computation. (Note, μ_{pop} affects consumers' prior means, but does not affect behavior given a certain mean belief.)

and their implications focuses on the base model without random coefficients. Of the four models, this is the most parsimonious one that captures the dynamic trade-offs of interest and the sorting-induced biases.

Figure 4 depicts simulated moments for the base model 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. In particular, the usage rates of plan 1 enrollees drops steadily from a high around .73 in week 1 to a low around .22 in week 60. Another moment to fit is the percent of consumers who enroll on each plan. In the base model, predicted shares for each plan are .151, .325, and .525, respectively, compared to the actual shares of 0.123, 0.317, and 0.561. The over-prediction of enrolless on plans 1 and 2 reflects the difficulty of matching all the moments in the data with only nine parameters. Typical discrete choice models have a mean utility for each option, which enables the model to perfectly predict market shares. The value of each plan in our discrete choice setting, however, is derived from the value function. Our model has no plan-specific utility parameters to specifically match these moments.

We also note that the model underpredicts the week 1 usage rates of plan 2 and plan 3 enrollees. The very steep declines in usage from week 1 to week 2 by consumers on these plans, relative to the declines in subsequent weeks, suggests that the week 1 usage may be driven by an element not present in our model. For example, some people may be signing up online and using the service at the time they signup. To avoid possible specification error of the week 1 usage, we instead opted to ignore each consumer's week 1 usage when estimating each model.[29] That is, $Pr(c_{it}|m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta)$ in the likelihood function is evaluated as unity if $t = 1$.

Simulated moments for the model imposing $\tilde{\sigma}_{pop} = \sigma_{pop}$ are presented in Figure 5. The graph illustrates clearly that the rational expectations assumption is unable to explain the data. Without sorting-induced biases, the model is unable to predict the significant drop in usage by consumers who enroll on plans with fixed fees (plans 1 and 2). The log-likelihoods presented in Table 3 are sufficiently different that a Chi-square test of $\tilde{\sigma}_{pop} = \sigma_{pop}$ is easily rejected.

The price coefficient is $-.232$ which implies a price elasticity of $-.65$ over the 70-week period. The elasticity for week 1, however, is only $-.22$ compared to $-.79$ for week 70. The expected discounted value of revenues declines by 3.3 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 estimates of $\tilde{\sigma}_{pop}$ (13.966) suggests that consumers were very uncertain about the value of this service for the overall population. Indeed, this amount of uncertainty is about five times the degree of heterogeneity in actual match values ($\sigma_{pop} = 2.712$). This stark difference leads

[29]Including the week 1 usage reduces the fit in later weeks. Given the higher potential for misspecification of week 1, we opt to exclude it.

the consumer to place relatively more weight on her pre-enrollment private signal, which has a a precision of $\sigma_{pre} = 10.136$. The uncertainty she faces prior to enrolling is the posterior standard deviation $\sqrt{10.136^2 * 13.966^2/(10.136^2 + 13.966^2)} = 8.20$. Recall from equation 13 that her mean belief prior to enrolling is a weighted average of $\mu_{pop} = -3.329$ and her private signal, with weights determined by $\tilde{\sigma}_{pop}$ and σ_{pre}. The high estimate of $\tilde{\sigma}_{pop}$ leads the consumer to place more weight on her private signal, which generates the sorting-induced biases: consumers who chose the high fixed fee tariff, on average, received optimistic private signals. Given our estimates, the average biases (mean belief - mean match-value) for plans 1, 2, and 3 are 9.550, 3.050, and -4.830, respectively.

Table 4 details the rates at which these biases are resolved as consumers update their beliefs with the post-enrollment signal and the usage signals.[30] The initial biases appear in the last three columns of the first row. After the post-enrollment signal (which has precision $\sigma_{post} = 4.705$), the biases are dramatically reduced to 2.364, .755, and -1.196, for enrollees on each of the three plans. After the first usage experience (which has precision $\sigma_{post} = 7.641$), these biases are further reduced to 1.839, .587, and -.930. After ten usage signals, the biases are down to .614, .196, and $-.31$. These units of utils may be converted to dollars by dividing by $-\alpha = .232$. Hence, after 10 uses consumers' beliefs are, on average, within one to three dollars of the service's value.

The discount factor is estimated to be .987, which seems low given that each period is only one week.[31] Nonetheless, the hypothesis that consumers are myopic (i.e., $\beta = 0$) is easily rejected given the differences in log-likelihoods in the last row of Table 3. 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 7 plots the model's predicted probability of usage in week 1 as a function of the week 1 belief m_{i1} and the subscribed plan. The solid lines correspond to forward-looking consumers and the dashed lines depict myopic consumers.[32] For a given belief m_{i1} 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 .45.

The effect of per-use prices on usage for given beliefs is also evident in Figure 7. 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 86.373 utils by the price coefficient (.232 dollars per util) implies that the switching

[30]The last three columns of 4 reveal the evolution of posterior mean beliefs for a hypothetical consumer with a match-value of zero (i.e., $\mu_i = 0$) who receives an initial signal optimistic by the amount of the average bias associated with each plan, and receives experience signals equal to the match-value of zero. The posterior mean when the signal equals μ_i is equivalent to the expected posterior mean with random experience signals centered around μ_i.

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

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

cost is equivalent to \$372. That is, consumers would switch plans only if the value of expected discounted utility were at least \$372 higher on an alternative plan. While \$372 may seem high, this cost is equivalent to \$4.84 per week (using the estimated discount factor of .987).

The estimated β is sufficiently high that switching is still predicted to occur: 55.7 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 are predicted to switch since the required weekly difference of \$4.84 is less than the \$1.14 weekly fixed fee component of plan 2.

5.1 Illusive 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}. A consequence of consumers not knowing the population distribution of μ_i is that their realized surplus will differ from their expected surplus. Two forces are at work to yield a realized surplus less than expected when $\tilde{\sigma}_{pop} > \sigma_{pop}$. The first is that sorting-induced biases lead consumers to signup for plans with fixed fees. This is not an irrational action. Such consumers are using all of their information as rationally as possible. Nonetheless, many of these consumers learn they should be on a different plan, but are unwilling to incur the switching costs needed to switch plans. Hence, these consumers end up with less surplus than expected.

One might expect the negative biases of consumers who signed up for the per-use charges of plan 3 to compensate the losses of the plan 1 consumers. However, even these plan 1 consumers have an average realized surplus lower than the expected levels. In this case, however, the discrepancy is due to over estimating the "option value" of the service. The perceived option value is based on the high $\tilde{\sigma}_{pop}$, but the realizations are driven by the much lower population heterogeneity determined by σ_{pop}. That is, the perceived probability of having a sufficiently high μ_i to earn high surplus exceeds the actual probability of this outcome, despite the pessimistic initial mean belief.

The differences between expected and realized surplus due to these two forces are significant. 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 44.4 utils. The expected discounted utility, averaged across all consumers is 196.4 utils. The expected consumer surplus (CS), is therefore $(196.4 - 44.4)/\alpha$, which equals \$655. On a weekly basis, the expected surplus is $\$655(1 - \beta) = \$8.52$.

For comparison, if consumers knew their match-values after the pre-enrollment signal (i.e., $\sigma_{pre} = 0$) the expected (and realized) consumer surplus would average a mere \$20.2 per consumer. Averaging over consumers who in fact use the service at least once in a 100-week simulation, this surplus increases to \$280. Hence, a few consumers would enjoy high surplus from this service.

5.2 Parameter heterogeneity

The estimates reported in Table 3 reveal that the random coefficients model is quite similar to the base 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.

Instead, we regress consumers' posterior means of θ_i (given the model, g, and choices c_{it} and s_{it}) on their demographics. The unadjusted R-squared measures for these eleven regressions ranged from .013 to .041.[33] Hence, consumers differ but their differences are essentially unrelated to observed 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 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

[33] We do not bother reporting the estimates since the few statistically significant coefficients have little economic significance.

(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 switching costs and uncertainty

Table 5 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 50,000 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 5.1, consumers' realized surplus is negative ($-81.0) despite an expected surplus of $655. In the long run (i.e., in week 100) the firm receives $1.02 per consumer, yielding a steady-state discounted revenue of $503.7 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 removes switching costs by setting δ to zero. 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 $\$-19.5$ and revenue falls by more than half to $246.4. The influence of switching costs on initial tariff choice is illustrated by 92.4 percent of consumers initially choosing plan 1 without switching costs, compared to only 14.9 percent with the estimated switching costs.

The final model in Table 5 removes uncertainty about match-values by setting $\sigma_{pre} = 0$. Switching costs are inconsequential since there is no uncertainty. Realized CS is finally positive, although only $20.2 per consumer (same as expected CS) since so few consumers actually use the service. The firm's discounted revenues of $201.4 are lower than the previous model with uncertainty and no switching costs since revenues are received from the experimenting consumers.

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 numerically compute optimal tariffs for various levels of consumer uncertainty and switching costs. In each case we assume the firm knows the distribution of match-values and the other demand parameters.[34] We find that the optimal tariff varies considerably across the different scenarios, which suggests that the effectiveness of different types of tariffs is an empirical question specific to the particular market of interest. For our market flat fees are effective only when switching costs are high and consumers have substantial uncertainty about match-values. When switching costs are low (even if only occasionally) or when consumers have little uncertainty about their match-values, 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.[35] 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

[34]Since this paper focuses on consumer uncertainty and learning, we do not consider a need for the firm to discover the demand curve.

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

discounted revenues from the delivery tariffs only.

Table 6 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 $F = \$4.98$ generates over four times the discounted revenue of the optimal uniform price of $p = \$8.81$. Surprisingly, the optimal single two-part tariff has the same F as the flat fee tariff, but adds a low per-use price of \$1.31. 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 used by only a few consumers who receive extremely high idiosyncratic utility shocks.

The role of uncertainty in determining whether flat tariffs dominate per-use tariffs is illustrated by the comparative dynamic in Figure 8. For various levels of σ_{pre} ranging from 0 to 9 utils, we compute the optimal flat tariff and the optimal per-use tariff.[36] The per-use tariff is \$7.19 for $\sigma_{pre} = 0$ and increases to \$8.59 when $\sigma_{pre} = 7$.[37] Discounted revenues with per-use tariffs are \$217.9 per consumer with $sigma_{pre} = 0$ and slowly increase to \$253 per consumer when $\sigma_{pre} = 9$. Most of the gain occurs over the range from 0 to 3 utils. The increased revenues with consumer uncertainty are from consumers who have relatively low match-values but use the service until they learn this is indeed the case for them. The flat tariff, on the other hand, rises from \$3.75 at $\sigma_{pre} = 0$ to \$5.03 at $\sigma_{pre} = 2.5$, with revenues increasing dramatically over the whole range of σ_{pre}. These gains are due to the increased degree of sorting-induced biases as consumer uncertainty increases. Regardless of σ_{pre} the optimal fee is low enough to lock-in enrolles. That is, the fees are low enough to ensure that no enrollees quit once they signup. For low levels of uncertainty, the fee is below the highest fee for which even a consumer who simply pays the fee in perpetuity with no additional usage would remain enrolled. For high levels of uncertainty ($\sigma_{pre} \geq 2.5$), the optimal fee is exactly this bound, which equals $(1-\beta)\delta/\alpha = \$4.99$ since $F/(1-\beta)$ is the present value of weekly payments of F. For low levels of uncertainty, charging less than \$4.99 is beneficial since enough additional consumers enroll to offset the lower fee per enrollee. In the graph, ex-ante (fixed fee) pricing dominates ex-post (per-use) pricing for $\sigma_{pre} \geq 1.5$. As we now discuss, this dominance of ex-ante pricing when consumers face uncertainty hinges on switching costs being high in each period.

To assess the importance of switching costs, Table 7 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 \$8.96 generates about 15 percent more revenue than the optimal flat fee of \$6.52. A two-part tariff increases revenue by 1.9 percent,

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

[37] The changes in the tariffs are not as smooth as the changes in the discounted revenues— a consequence of using simulation with a revenue function that is relatively flat (for both flat fees and per-use prices).

compared to the per-use tariff, and adding a second two-part tariff to the menu increases revenue by an additional .9 percent. Adding yet another tariff (a per-use only one) yields an almost imperceptible increase in revenue of less than .1 percent. Hence, we again find that using menus of tariffs is of little benefit to the firm.[38]

An alternative method for assessing the effect of switching costs is presented by the comparative dynamic depicted in Figure 9. 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 \$8.81, which yields discounted revenue of \$252.7.[39] The optimal flat tariff is \$4.75 with no switching costs and declines to a little over \$3 for switching costs ranging from 1.5 to 15 utils.[40] For switching costs of 20 and higher, the optimal flat tariff is again determined by $(1-\beta)\delta/\alpha$, for which enrollees who never use the service continue to pay the fee F. Though the fee drops significantly as δ increases from 15 to 20, the perpetual payment of the lower fee by all enrollees leads to a jump in discounted revenues. Discounted revenues are higher with per-use pricing for switching costs less than or equal to approximately 16 utils (interpolating between the computed values at 15 and 20).

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 8 are generated by 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.[41] The flat tariff generates discounted revenue of \$918.8 compared to only \$243.1 for the per-use tariff. The single two-part tariff generates revenue of \$941.8, which is 2.5 percent higher than the flat tariff. Adding a second two-part tariff increases revenue to only \$945.5. 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.

Comparing Table 6 to the first row of Table 5 indicates that predicted revenues using the parameter estimates are much higher with the optimal flat tariff than with the actual menu offered to consumers (\$1020.6 compared to \$530.5). 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

[38] Optimal tariffs for the model with no match-value uncertainty (i.e., $\sigma_{pre} = 0$) are similar to this model.

[39] 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.

[40] Again, the irregular variation in the optimal flat fee for $\delta \leq 15$ reflects numerical approximation of the optimal tariff being based on simulated consumers. The revenue is relatively flat near these optimal levels.

[41] The optimal flat fees are 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 slightly lower revenue as well.

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 with its optimal tariffs. We therefore put more stock in the optimal tariffs reported in Table 8 for the random coefficients model. In particular, this model suggests a flat tariff of $F =$\$4.19 or an optimal single two-part tariff with $F =$\$3.43 and $p =$\$3.57.

7 Conclusion

Our model of consumer learning and tariff choice is able to rationalize data in which some consumers appear to make tariff choice mistakes, while retaining unbiased beliefs across consumers as a whole and fully rational behavior. This goal is achieved by allowing one parameter in the specification of beliefs to deviate from the value required for rational expectations. Consumers with an initial prior that is less precise than the rational expectations prior (though with the same mean), leads them to place more weight on their private signal prior to enrollment. Though these signals are on average correct, consumers with optimistic signals have posterior means with upward bias. Since such consumers are more likely to have beliefs sufficiently high to warrant choosing the plan with high fixed fees, enrollees of this plan tend to be overly optimistic. That is, the tariff choice leads to sorting-induced biases even though beliefs are unbiased on average and consumers are fully rational in their actions and use of information.

Using the estimated model to evaluate counterfactual pricing schemes, we determine that using menus of tariffs to screen consumers is an ineffective price discrimination tool in this market, regardless of the degree of switching costs and uncertainty. If switching costs are always high, a fixed fee tariff is better than a per-use tariff since many consumers continue to pay the fees despite learning the service is of little value to them. If switching costs are low, even occasionally, per-use pricing is better since the lock-in benefit of the fee is eliminated, or at least reduced.

Interestingly, Peapod, the largest online grocer in the United States, 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. 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.

Regarding future work, one could extend our analysis to consider learning by the firm about demand.

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*, 73, 1137–1174..

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.

Narayanan, S., P.K. Chintagunta, and E. Miravete (2007) "The role of self selection, usage uncertainty and learning in the demand for local telephone service," *Quantitative Marketing and Economics*, 5, 1–34.

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	.123	.67–1	.60	.12
Plan 2	\$1.14	\$6.95	.317	.23–.67	.38	.18
Plan 3	\$0	\$11.95	.561	0–.23	.21	.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: Parameter Estimates

Parameter		Base Model	$\beta = 0$ Model	$\tilde{\sigma}_{pop} = \sigma_{pop}$ Model	Random Coefficients mean θ_i	Random Coefficients std.dev. θ_i
μ_{pop}	(mean	-3.329	-2.644	-3.390	-3.636	2.410
	match quality)	(0.068)	(0.080)	(0.070)	(0.088)	(0.086)
σ_{pop}	(std. dev.	2.712	2.165	7.538	10.520	2.812
	match quality)	(0.047)	(0.045)	(0.099)	(0.183)	(0.143)
$\tilde{\sigma}_{pop}$	(perceived std. dev.	13.966	9.371		13.413	3.052
	match quality)	(0.431)	(0.370)		(0.184)	(0.138)
σ_{pre}	(std. dev.	10.136	6.813	8.650		
	per-enroll signal)	(0.401)	(0.185)	(0.253)		
σ_{post}	(std. dev.	4.705	5.618	3.850	5.227	1.982
	post-enroll signal)	(0.064)	(0.150)	(0.063)	(0.085)	(0.087)
σ_{ϵ}	(std. dev.	7.641	5.153	5.074	8.287	2.538
	experience signal)	(0.138)	(0.117)	(0.119)	(0.149)	(0.134)
β	(weekly	0.987		0.991	0.9856	
	discount factor)	(0.000)		(0.000)		
α	(price coefficient)	-0.232	-0.085	-0.090	-0.189	0.076
		(0.004)	(0.008)	(0.004)	(0.004)	(0.004)
δ	(switching cost)	86.373	0.792	45.401	73.064	13.534
		(1.460)	(0.099)	(1.248)	(0.920)	(0.715)
	Log likelihood	-50561.1	-51078.2	-52169.9	-50393.29	

σ_{pop} is redundant with random coefficients since μ_{pop} already varies across consumers.
All models use an exogenous weekly quit rate of $\gamma = .003$.
The random coefficients model uses a fixed $\beta = .9856$.
Asymptotic standard errors (fixing the monte carlo draws) are in parentheses.

Table 4: Posterior beliefs and the resolution of sorting-induced biases

cumulative usage	Posterior std. dev.	Posterior mean $-\mu_i$ Plan 1	Plan 2	Plan 3
pre	8.204	9.550	3.050	-4.830
post	4.082	2.364	0.755	-1.196
1	3.600	1.839	0.587	-0.930
2	3.257	1.505	0.481	-0.761
3	2.996	1.274	0.407	-0.644
4	2.789	1.104	0.353	-0.558
5	2.620	0.974	0.311	-0.493
6	2.478	0.872	0.278	-0.441
7	2.358	0.789	0.252	-0.399
8	2.253	0.720	0.230	-0.364
9	2.161	0.663	0.212	-0.335
10	2.079	0.614	0.196	-0.310
11	2.006	0.571	0.182	-0.289
12	1.941	0.534	0.171	-0.270
13	1.881	0.502	0.160	-0.254
14	1.826	0.473	0.151	-0.239
15	1.776	0.448	0.143	-0.226
16	1.730	0.425	0.136	-0.215
17	1.687	0.404	0.129	-0.204
18	1.648	0.385	0.123	-0.195
19	1.611	0.368	0.118	-0.186
20	1.576	0.352	0.113	-0.178
25	1.431	0.291	0.093	-0.147
30	1.320	0.247	0.079	-0.125
35	1.231	0.215	0.069	-0.109
40	1.158	0.190	0.061	-0.096
45	1.097	0.171	0.055	-0.086
50	1.045	0.155	0.049	-0.078
55	0.999	0.142	0.045	-0.072
60	0.959	0.130	0.042	-0.066
65	0.923	0.121	0.039	-0.061
70	0.891	0.113	0.036	-0.057

"pre" is prior to the initial enrollment.
"post" is after enrollment, before usage.
Sorting-induced biases are initially 9.550, 2.050, and -4.830, respectively, across the three plans. Declining biases are depicted for a consumer receiving signals of exactly μ_i.

Figure 1: **Expected Cost per Delivery**

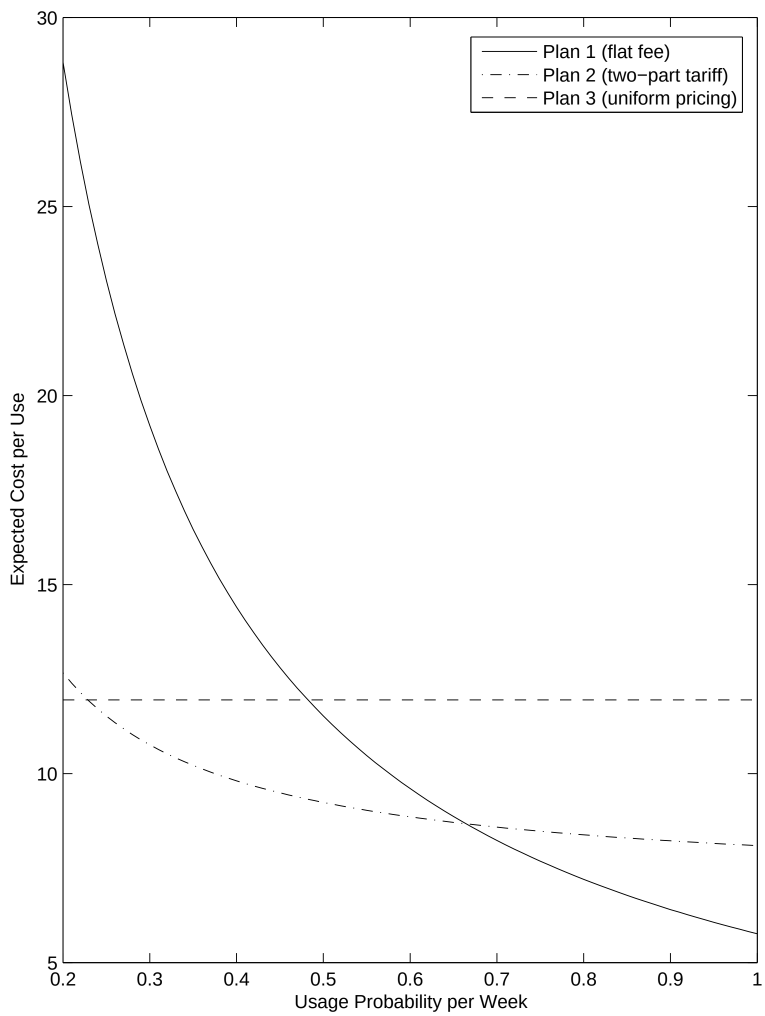


Figure 2: **Usage Rates**

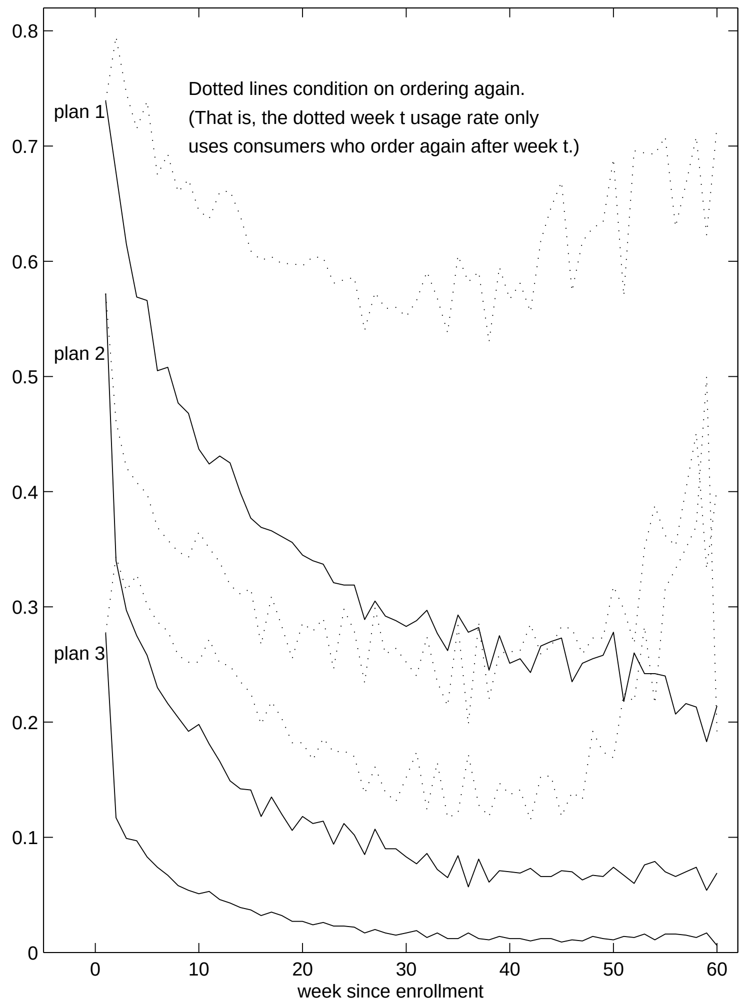
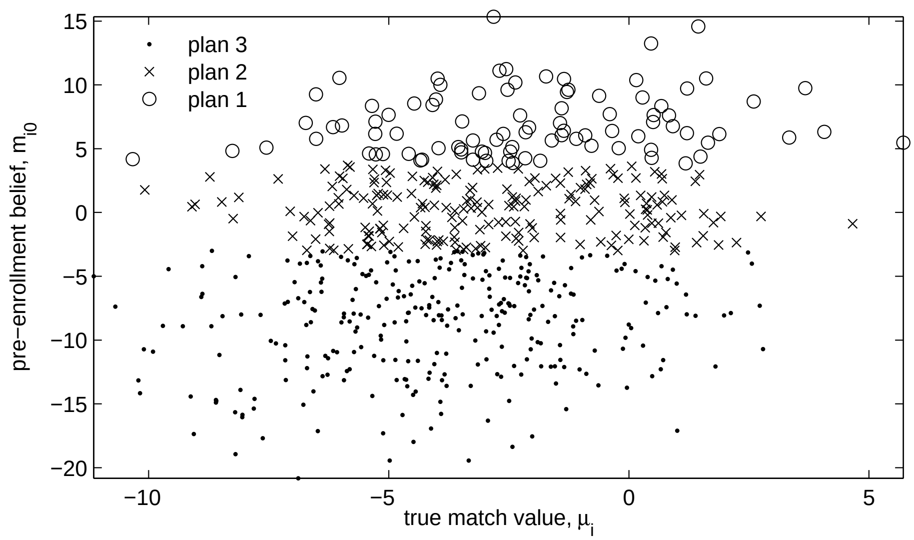


Figure 3: **Sorting of Beliefs and Match Values across Plans**



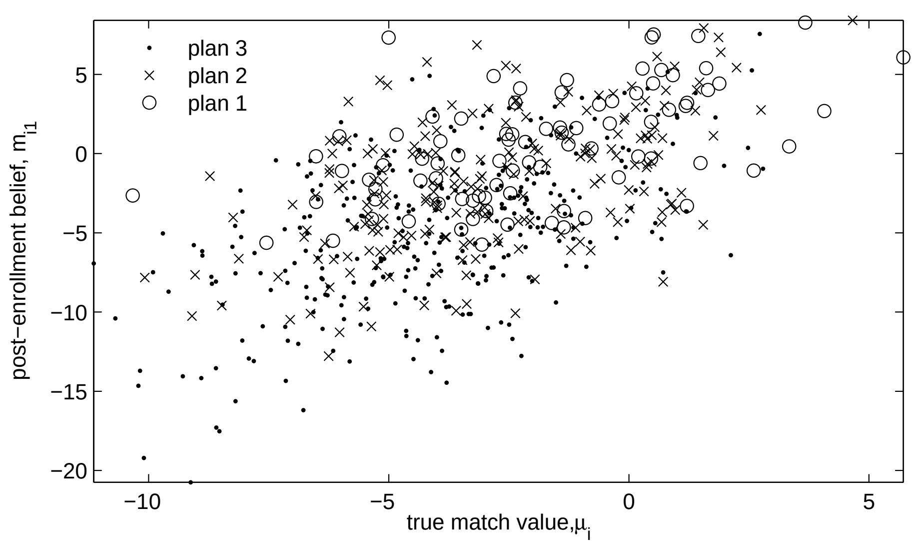


Figure 4: **Simulated Usage Rates, base model**

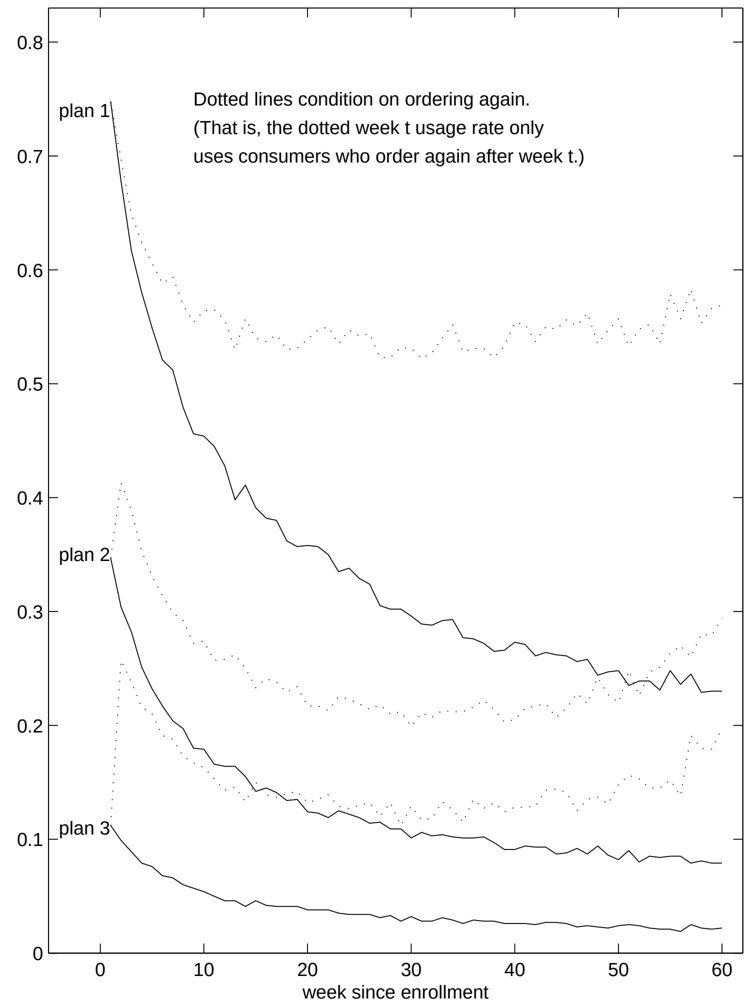


Figure 5: **Simulated Usage Rates, model with** $\tilde{\sigma}_{pop} = \sigma_{pop}$

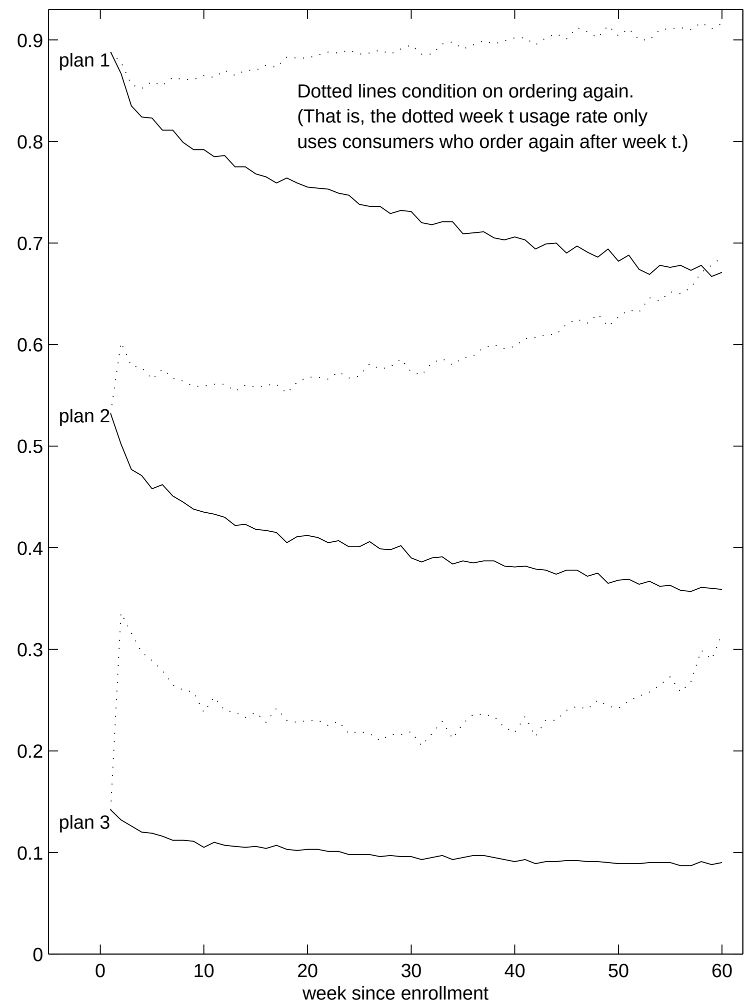


Figure 6: **Simulated Usage Rates, model with $\beta = 0$**

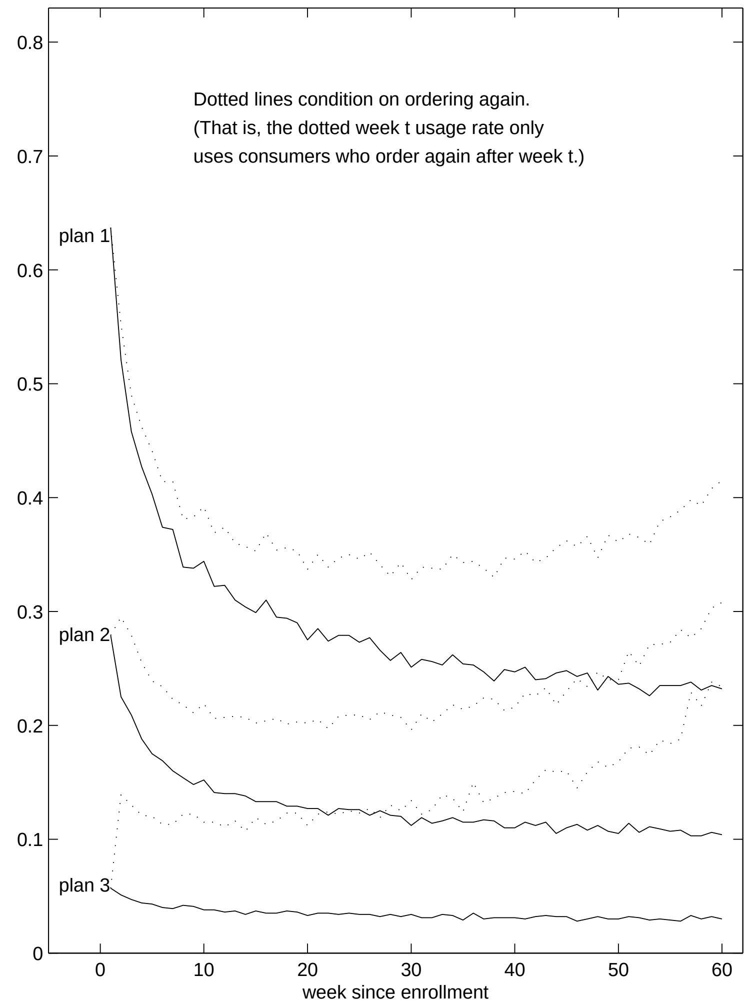


Figure 7: **Information Acquisition**

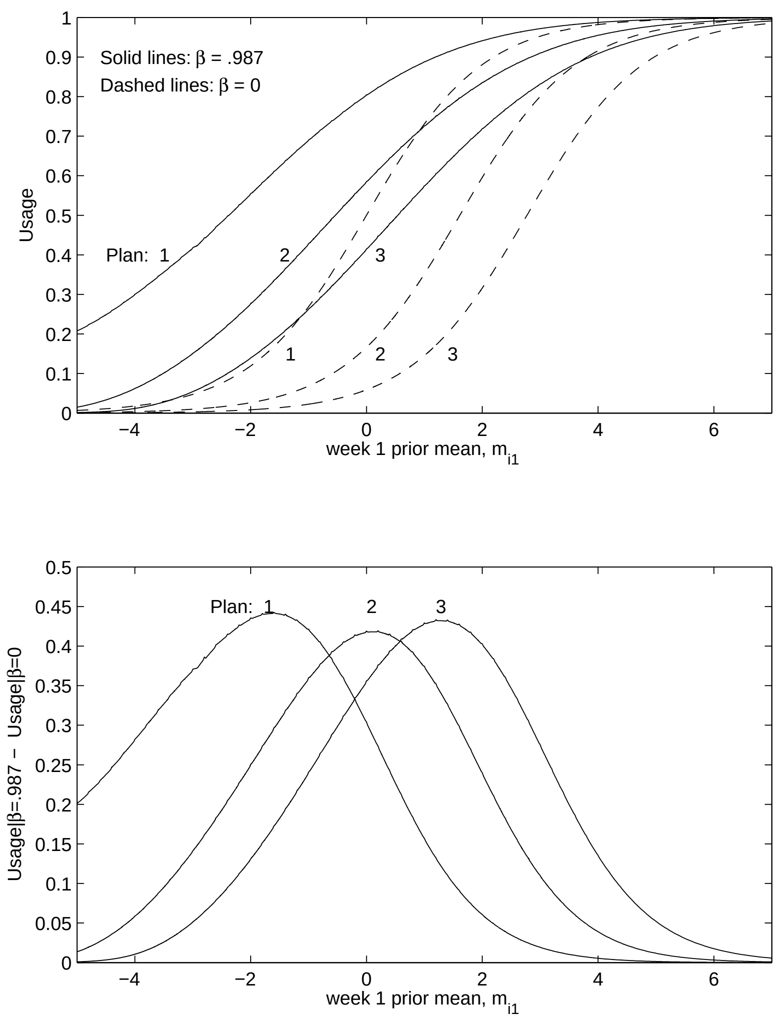


Figure 8: **Tariffs and Revenues as Functions of Initial Uncertainty** (σ_{pre})

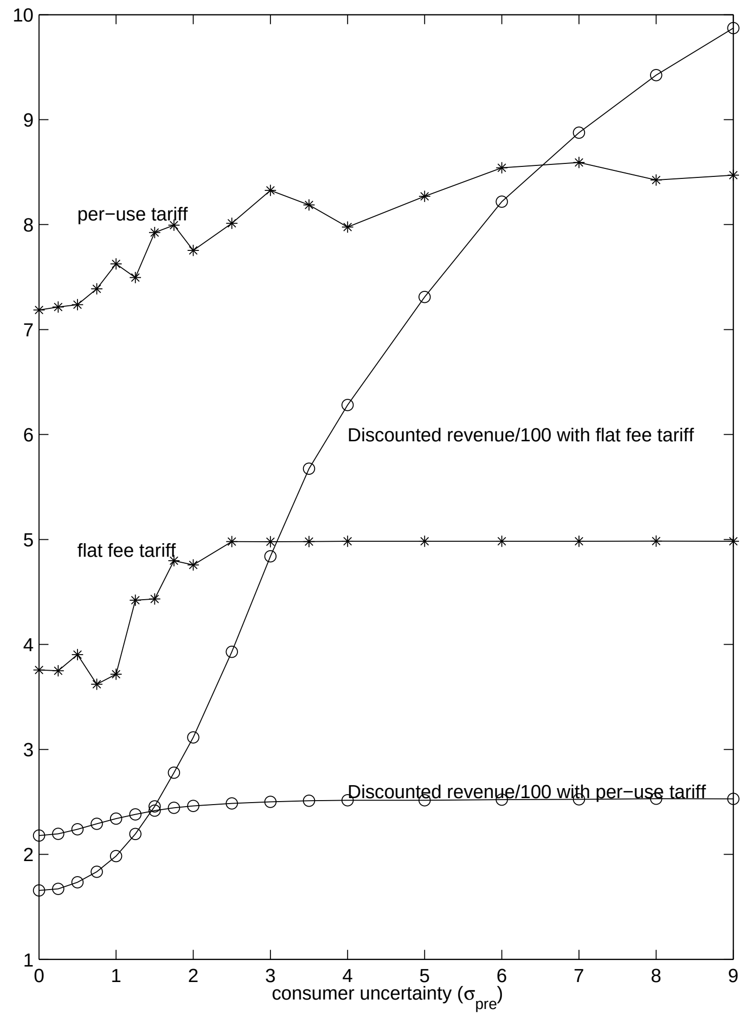


Figure 9: **Tariffs and Revenues as Functions of Switching Costs (δ)**

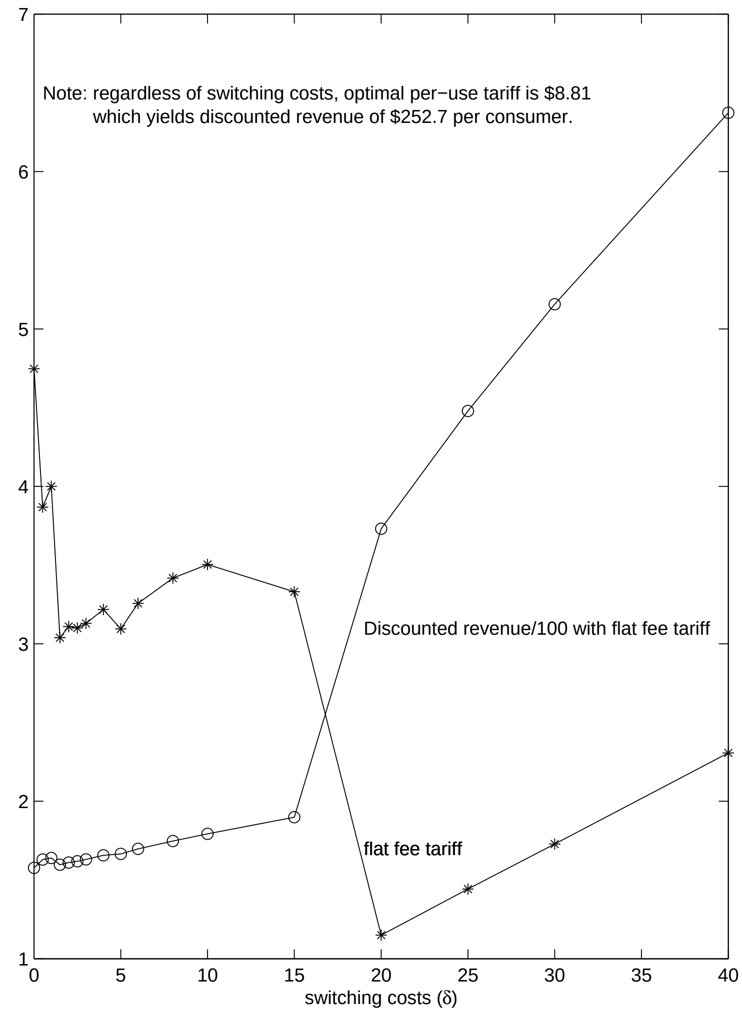


Table 5: Effect of switching costs and uncertainty

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)
Base model estimates	.745, .569	.350, .079	.119, .016	530.5	-81.0
	(.149, .060)	(.329, .329)	(.522, .612)	(503.7)	(655.1)
No switching costs	.924, .909	.394, .400	.018, .013	246.4	-18.5
($\delta = 0$)	(.999, .037)	(.000, .023)	(.000, .940)	(222.7)	(722.5)
No uncertainty	.915, .915	.401, .401	.012, .012	201.4	20.2
($\sigma_{pre} = 0$)	(.036, .036)	(.022, .022)	(.943, .943)	(201.4)	(20.2)

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 nearly \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
All values generated by simulating 50,000 consumers over 100 weeks.

Table 6: 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)
$F_1 = 4.98$, $p_1 = 0$	.630, .229			1020.6	-122.0
(flat fee tariff)	(.361, .414)			(1020.9)	(640.6)
$F_3 = 0$, $p_3 = 8.81$			.259, .053	252.7	-12.9
(per-use tariff)			(1.0 , 1.0)	(230.3)	(642.0)
$F_2 = 4.98$, $p_2 = 1.31$		.612, .202		1026.2	-123.9
(1 two-part tariff)		(.342, .391)		(1023.4)	(602.1)
$F_1 = 4.98$, $p_1 = .78$	.616, .216	.002, .000		1027.0	-125.7
$F_2 = 0.00$, $p_2 = 134.03$	(.352, .404)	(.648, .596)		(1024.0)	(612.7)
(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 nearly \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
All values generated by simulating 50,000 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.

Table 7: 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)
$F_1 = 6.52$, $p_1 = 0$	.537, .886			220.4	-69.4
(flat fee tariff)	(.889, .044)			(140.5)	(696.0)
$F_3 = 0$, $p_3 = 8.96$			.258, .052	253.3	-13.1
(per-use tariff)			(1.0 , 1.0)	(230.5)	(638.0)
$F_2 = .05$, $p_2 = 6.44$		.275, .144		258.2	-14.3
(1 two-part tariff)		(.764, .500)		(234.7)	(598.7)
$F_1 = .21$, $p_1 = 8.08$	.298, .333	.023, .019		260.5	-16.6
$F_2 = .01$, $p_2 = 10.37$	(.966, .137	(.034, .365)		(235.1)	(673.1)
(2 two-part tariffs)					
$F_1 = .21$, $p_1 = 8.04$	.298, .330	.019, .019	.000, .000	260.7	-16.6
$F_2 = .01$, $p_2 = 10.45$	(.970, .140)	(.030, .360)	(.000, .500)	(235.3)	(673.8)
$F_3 = 0$, $p_3 = 51.01$					

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 nearly \$500.
$\frac{Rev_{final}}{1-\beta_{firm}}$ measures the firm's steady-state value.
All values generated by simulating 50,000 consumers over 100 weeks.
For ease of comparison, single tariff "menus" also appear as "Plan 2" or "Plan 3."

Table 8: 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)
$F_1 = 4.19$, $p_1 = 0$	.580, .177			918.8	-130.4
(flat fee tariff)	(.367, .442)			(914.8)	(996.6)
$F_3 = 0$, $p_3 = 11.62$			.235, .038	243.1	-28.6
(per-use tariff)			(1.0 , 1.0)	(216.4)	(883.7)
$F_2 = 3.43$, $p_2 = 3.57$		.507, .121		941.8	-124.0
(1 two-part tariff)		(.359, .488)		(932.8)	(928.5)
$F_1 = 3.58$, $p_1 = 3.30$	.558, .137	.342, .080		945.7	-125.6
$F_2 = 3.32$, $p_2 = 3.99$	(.355, .375)	(.000, .105)		(936.7)	(933.2)

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 nearly \$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."