

Consumers choosing flat-rate contracts tend to have insufficient usage to warrant the cost, particularly for new products. We propose and estimate a Bayesian learning model of tariff and usage choice that explains this flat-rate bias without relying on behavioral misjudgments or tariff-specific preferences. For new products, consumers are uncertain about both their utility relative to the population mean and the mean itself. We show that this latter uncertainty inflates prior variances, which leads consumers to weight their private signals more heavily when updating beliefs. Posteriors are unbiased across products. For a given product, however, the unknown mean yields a “winner’s curse”: Consumers with high posteriors tend to overestimate their utility. These consumers choose fixed-rate tariffs and lower their usage as they correct their beliefs. The flat-rate bias arises when switching costs deter them from changing tariffs. Using the estimated model, the authors find that tariff menus are ineffective screening devices for price discrimination by an online grocer. Predicted revenues increase by 20% when the per-use tariff is dropped, because more consumers choose and stay with the flat rate.

Keywords: consumer learning, tariff choice, dynamic discrete choice

Tariff Choice with Consumer Learning and Switching Costs

Firms often price discriminate among consumers with heterogeneous preferences by offering menus of tariff or bundle options. Consumers are often uncertain about their future demands and base subscription choices on beliefs of future usage. Researchers have begun to explore the accuracy of these beliefs and of optimal contract design given the uncertainty and potential biases. For example, Grubb (2009) shows that three-part tariffs, as used by cellular phone service providers, are nearly optimal when consumers are overconfident (i.e., underestimate the variance of future demand). Miravete (2003), DellaVigna and Malmendier (2006), and Lambrecht and Skiera (2006) show

that consumers often choose and retain the wrong tariff for telephone calling plans, health clubs, and Internet service, respectively.

The stylized facts indicate a flat-rate bias, with consumers tending to choose flat-rate contracts when per-use contracts would be better. DellaVigna and Malmendier (2006) present evidence that biased beliefs about future behavior best explain the flat-rate bias and attribute such biases to irrational or behavioral inclinations. Lambrecht and Skiera (2006), Lambrecht, Seim, and Skiera (2007), and Iyengar, Ansari, and Gupta (2007) explain the flat-rate bias by estimating a high mean utility for flat-rate contracts in models with fixed effects and idiosyncratic preferences for each tariff. This approach, however, is ill-suited for predicting choices and revenues under alternative tariffs because the alternative tariffs’ mean utilities are unknown. Moreover, the idiosyncratic preferences for tariffs imply that subscriptions necessarily increase as tariffs are added, even if they are identical to those already offered.

In this study, we propose a Bayesian learning model of tariff and usage choice that can explain the flat-rate bias within the rational expectations framework and without invoking tariff-specific preferences. The intuition is simple. Consider

*Ronald L. Goettler is Assistant Professor of Marketing, Booth School of Business, University of Chicago (e-mail: ronald.goettler@chicagobooth.edu). Karen Clay is Associate Professor of Economics, Heinz School of Public Policy and Management, Carnegie Mellon University (e-mail: karen.clay@cmu.edu). The authors gratefully acknowledge helpful comments from Dan Akerberg, J.P. Dubé, Shane Greenstein, Bob Miller, Christine Parlour, Uday Rajan, and Holger Sieg. They especially thank Eric Wolff and the online grocer for providing the data. Ronald Goettler appreciates financial support from the True North Communications, Inc. Faculty Research Fund at the University of Chicago Booth School of Business.

a new product for which consumers' "match-values" (i.e., idiosyncratic utilities) are distributed $N(\mu_{\text{product}}, \sigma_{\text{product}}^2)$. Being a new product, μ_{product} is unknown. Suppose consumers know that new products in this category have μ_{product} distributed $N(\mu_{\text{category}}, \sigma_{\text{category}}^2)$. Therefore, the initial common prior is $N(\mu_{\text{category}}, \sigma_{\text{product}}^2 + \sigma_{\text{category}}^2)$ with a variance that sums the idiosyncratic uncertainty $\sigma_{\text{product}}^2$ and the aggregate uncertainty $\sigma_{\text{category}}^2$. Consumers have private information about themselves that is relevant for their match-value, which we model as an unbiased signal. They incorporate this signal in a Bayesian fashion and choose either the flat-rate or per-use tariff according to whether their posterior mean is above or below some critical value. Aggregate uncertainty (i.e., $\sigma_{\text{category}}^2$) inflates the prior variance, causing consumers to weight private signals more heavily. Although posteriors are unbiased across products, the heavy weight on private signals implies that consumers with high posteriors for a given product tend to overestimate their utility. Consumers eventually correct their beliefs and lower their usage, but switching costs deter them from changing tariffs, which yields the flat-rate bias.

This phenomenon is similar to the winner's curse in common-value auctions (and affiliated value auctions), in that consumers with high posteriors have positive biases. The way these posteriors are used, however, differs: When bidding, a person conditions on having higher beliefs than others, whereas when choosing a tariff, he or she cannot condition on others' beliefs or actions. The winner's curse is actually harmless in auctions because bidders use valuations conditional on winning, which are lower than the unconditional valuations. That is, bidders "shave" their bids because if they win the auction, they are likely to have overestimated the item's value. In contrast, consumers choosing tariffs cannot correct their beliefs before making the choice because they have no way of knowing whether their posterior is high. They should not, however, shave their valuation because the selected tariff is "won" regardless of other consumers' actions and their valuations are actually correct on average across products.

We use our model to study the online grocer market, in which consumers order groceries for home delivery from a monopolist offering a menu with flat-rate, per-use, and two-part tariffs. This new service is an "experience good," in that consumers learn their match-values through consumption (Nelson 1970). Therefore, forward-looking consumers have an informational incentive to experiment with the service, leading them to trade off current utility for future utility. Anticipated costs to switching tariffs also lead to dynamic trade-offs.¹ Previous studies of consumers' tariff choices, including DellaVigna and Malmendier (2006), Miravete (1996, 2002), Lambrecht and Skiera (2006), Lambrecht, Seim, and Skiera (2007), Narayanan, Chintagunta, and Miravete (2007), and Iyengar, Ansari, and Gupta (2007), avoid these dynamic trade-offs by assuming that switching costs are negligible and uncertainty is resolved independent of consumption. Neither of these assumptions suits our study. Using household-level data, we estimate the model and perform counterfactual simulations to evaluate alterna-

tive pricing schemes and the effect of uncertainty and switching costs on revenue and consumer surplus.

Our model builds on Eckstein, Horsky, and Raban's (1989) dynamic brand choice model, in which consumers update their beliefs as they experience the brands. Using variants of this model, Erdem and Keane (1996) investigate the role of advertising in consumer learning about laundry detergent, Akerberg (2003) distinguishes between informative and prestige advertising for a new yogurt, and Crawford and Shum (2005) estimate how rapidly consumers learn the effectiveness and side effects of antiulcer drugs. Predating these applications, Miller (1984) estimates a Bayesian learning model matching workers to jobs.

We extend the dynamic brand choice model to include consumers' initial and subsequent tariff choices. Consumers select tariffs to maximize expected discounted utility given their beliefs, which induces a sorting by beliefs: Consumers who expect to purchase often choose tariffs with high flat fees and low per-use prices. Therefore, the observed high usage of such consumers reflects both the self-selection and the low per-use price. By endogenizing consumers' tariff choices, our model disentangles these two drivers of observed usage.

Using our estimated model, we assess the effectiveness of tariff menus as screening mechanisms to price discriminate among heterogeneous consumers who face uncertainty. We numerically solve for optimal tariffs menus, assuming they are fixed over time (as in our data), and find that menus are ineffective screening devices for price discrimination. Adding a second two-part tariff increases profits marginally, and a third tariff adds no gain. Miravete (2002) also finds limited gains from complex tariffs when consumers learn their demand over time, as do Courty and Hao (2000) when *ex ante* consumer heterogeneity is high. The gains are minimal because the incentive compatibility constraints are costly to satisfy: Offering additional tariffs reduces the firm's ability to extract surplus from high-demand consumers.

Moreover, fixed fees play two roles when consumers face high costs for switching tariffs. In addition to extracting surplus from inframarginal units consumed, they may lock consumers into paying the fee even when they learn their match-value is lower than initially expected. Including a per-use option on the menu may therefore lower revenues because fewer consumers will choose the tariffs with fixed fees. Indeed, eliminating the per-use option from the online grocer's current menu increases predicted revenues by more than 20%.

DATA

We use consumer-level data on grocery deliveries to 5368 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 was a monopolist in the home-delivery business that partnered with an existing local grocery chain to supply groceries. Online prices and discounts were the same as 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 by the partner chain, and displays on the delivery trucks. Most consumers enrolled while shopping in the partner-chain's stores. No free trials were offered. Once enrolled,

¹These switching costs are conceptually distinct from the switching costs related to brand loyalty.

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.

The online grocer offered consumers a menu of three two-part tariffs with weekly fixed fees of $F = (\$5.76, \$1.14, 0)$ and per-delivery prices of $p = (0, \$5, \$11.95)$, respectively. Plan 1 is therefore a flat-rate plan, and plan 3 is a uniform pricing plan. F and p are sometimes called *ex ante* and *ex post* prices, respectively, because consumers pay F before knowing their usage and pay p only for units consumed. Although the online grocer quotes fees on a monthly basis, consumers could change plans at any time, with fees prorated.

We observe each consumer's enrollment date and initial tariff choice, the date of each of his or her orders, the subscription plan at the time of each order, and the amount of each grocery order, which averages \$119. We treat the enrollment date as exogenous, reflecting the randomness with which consumers become aware of the service. To ensure sufficient opportunity to observe usage behavior, we only used consumers who had enrolled by week 60.

Grocery shopping typically occurs weekly, so we model weekly usage of the online grocer, with $c_{it} = 1$ if consumer i orders at least once during week t and $c_{it} = 0$ for no orders. We use $t = 1$ to denote the week of enrollment, not the first week of our sample. Only 2.8% of customers' weeks with orders have more than one order.

In addition, $s_{it} \in 1, 2, 3$ denotes consumer i 's subscription plan in week t . Unfortunately, we only observe a consumer's plan upon enrollment and with each order. Surprisingly, we never observe consumers switch (i.e., no one switches and orders again).² Thus, we assume s_{it} is unchanged between orders. The plan beyond the last order, however, is censored. Although we do not observe switches, many consumers likely "quit" by switching to plan 3, which has no fee, and never ordering again. To account for this event, our likelihood function (Equation A1) integrates over all plans to compute the probability of observing no usage during the "trailing weeks" between a customer's last order and the end of our sample period.

In Table 1, we provide relevant summary statistics for each plan. Only 12.3% of consumers signed up for plan 1, compared with 31.9% for plan 2 and 55.8% for plan 3. The mean weekly usage rate was .61 for plan 1 enrollees, .38 for plan 2 enrollees, and .21 for plan 3 enrollees. We compute each consumer's usage rate from the weeks spanning enrollment and the last observed order. This measure is an upper bound because it ignores weeks beyond the last order during which the consumer may have remained on the plan but did not order. In Table 1, we also report the range of usage rates for which each plan minimizes the expected cost per order. Figure 1 plots this expected cost, which equals $p +$

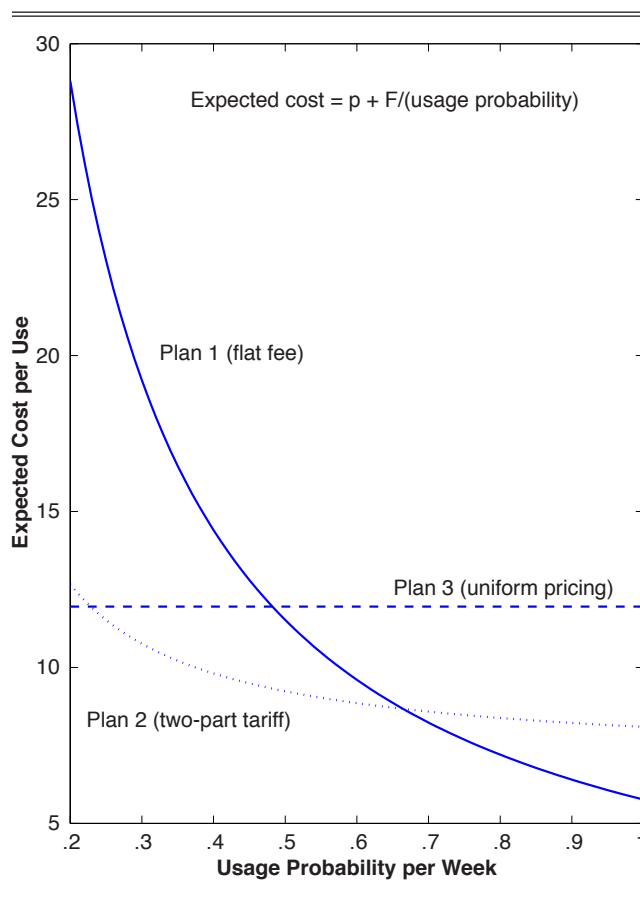
$(F/\text{usage probability})$, for each plan as a function of expected usage. Plan 1 minimizes costs for usage rates above .67, plan 3 minimizes costs for usage rates below .23, and plan 2 minimizes costs elsewhere. As the last row of Table 1 indicates, more than half of the consumers who enrolled in plans with fees minimized *ex post* delivery costs, assuming they quit after the last observed order. The flip side is that nearly half did not minimize costs, which suggests consumers indeed faced substantial uncertainty and switching costs. If consumers were uncertain but faced no switching costs, they would have switched plans if their

Table 1
TARIFF CHOICES

	Plan 1	Plan 2	Plan 3
F, flat fee	\$5.76	\$1.14	\$0
p, per-use price	\$0	\$6.95	\$11.95
Number of households enrolled	658	1712	2998
Initial plan shares	.123	.319	.558
Usage rates for which plan minimizes cost	.67–1	.23–.67	0–.23
Mean usage (between enrollment and last order)	.61	.38	.21
Share of enrollees who never order	.12	.17	.57
Share of ordering enrollees who minimize costs	.57	.53	.32

Notes: Weekly F is the quoted monthly fee (\$24.95, \$5.00, \$0) divided by 4.33.

Figure 1
EXPECTED COST PER DELIVERY



²Consumers even ignored letters explaining they should switch plans. The firm also assures us new customer identifications would not have been assigned upon switching. Being skeptical, we searched our data for matches within zip codes between new customers and customers who recently placed their last order. Candidates are regular customers on plans 2 or 3 whose last order is shortly before a new customer enrolls in plan 1 and orders regularly. We found only four possible matches, which had a negligible effect on the estimates.

beliefs crossed the expected-usage thresholds depicted in Figure 1. The absence in our data of switches followed by usage therefore suggests switching costs are high.

Some consumers enroll in plans with fixed fees but never try the service: 12% of those on plan 1 and 17% of those on plan 2 never order. Because we only use consumers who enroll at least ten weeks before the end of our sample period, such outcomes suggest consumers can receive information and revise beliefs without actually placing an order. A “post-enrollment” signal can reflect difficulty installing the software or navigating the company’s website for ordering, thereby leading the consumer to quit without ever ordering. The details of this specification and the initial tariff choice are discussed subsequently.

Our data are best summarized by plotting usage rates over time for enrollees of each plan. In Figure 2, the top solid line is the average c_{it} across consumers who enrolled in plan 1, even if they may have quit (i.e., including weeks beyond their last order). Nearly 74% of plan 1 enrollees used the online grocer during their first week. Their usage declined to around .2 by 60 weeks after enrollment. Average c_{it} becomes noisy as “weeks since enrollment” increases because few consumers signed up early enough to provide such data. If switching costs were low, plan 1 enrollees with low initial beliefs quickly dropping after experientially con-

firmed their beliefs could explain this usage drop. That is, consumers with low initial beliefs can justifiably sign up for plan 1 if switching costs are low because the incentive to experiment implies that initial usage is higher than the expected long-term usage. A high probability of needing to switch does not deter consumers from signing up for high-fee tariffs when switching costs are low. However, high switching costs, as the lack of observed switches suggests, imply that plan 1 enrollees should have had initial beliefs high enough to warrant this plan in the long run. Therefore, the quick drop in usage of plan 1 enrollees implies that they were overly optimistic. Our specification of beliefs nests two alternative explanations for this pattern: aggregate biases and conditional biases. Aggregate biases shift all consumers’ beliefs equally, whereas conditional biases depend on consumers’ signals. We show that conditional bias is more important than aggregate bias for explaining our data.

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. Because this conditioning event selects from consumers who are learning their match-values are (relatively) high, the rates are higher than the unconditional usage rates and increase at the end of the sample period.

MODEL

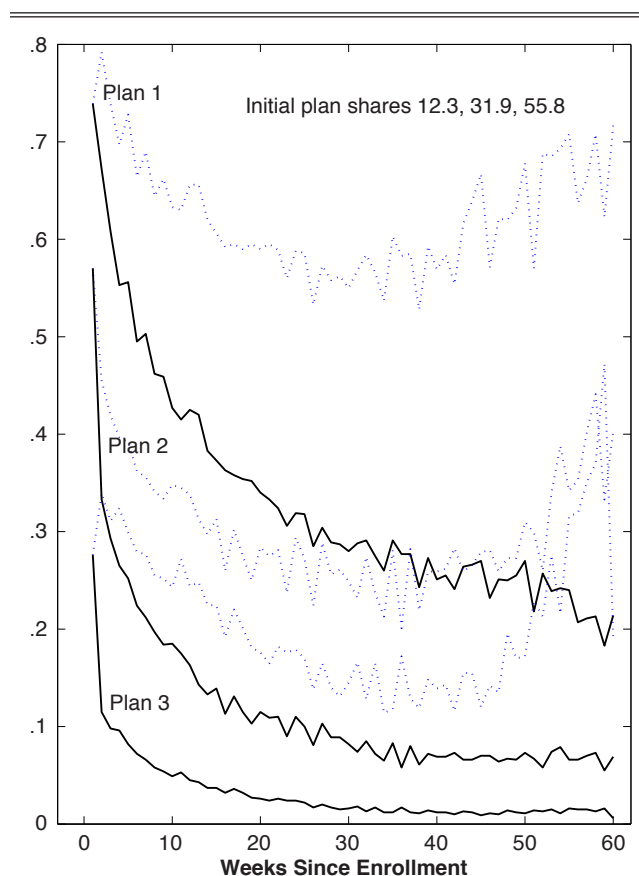
We first propose a Bayesian learning model of consumers’ tariff choices and usage of an experience good. Second, we present comparative statics to illustrate the effects of price sensitivity, switching costs, and uncertainty on tariff choice and usage conditional on tariff choice. We then specify the distribution of match-values and initial beliefs and distinguish between aggregate and conditional biases. We conclude the section with a model of the initial tariff choice, which differs slightly from the subsequent tariff choices.

A Bayesian Learning Model

We model the consumer’s weekly decision of whether to use the online grocer. The consumer’s decision has two dynamic aspects. First, the online grocer’s service is a new experience good about which consumers initially have limited information. As consumers use the service, they learn their utility from it (i.e., their match-values). In particular, consumers may learn whether they can efficiently use the ordering software, whether they like the produce selected, and whether they indeed save much time by using the online grocer. If a consumer’s prior belief suggests his or her match-value is likely low, he or she may still try the online grocer because the lower expected current utility may be offset by the possibility of learning that the service, in fact, provides high utility.

The second dynamic aspect arises from the firm’s use of subscription plans combined with the presence of switching costs. The firm offers a fixed menu of M two-part tariffs denoted by $(F_1, p_1, \dots, F_M, p_M)$, where F denotes the vector of flat fees and p denotes the vector of per-use prices. Each consumer chooses the best subscription plan (i.e., tariff) given his or her beliefs about the value of the service and the costs of switching plans in the event that a switch is warranted. For many products, switching costs are explicit financial charges. Because the online grocer does not charge consumers to change plans, this cost represents the hassle of thinking about which plan is best and calling to request the change.

Figure 2
WEEKLY USAGE RATES BY PLAN



Notes: Solid lines average over consumers initially enrolled in specified plan. Dotted lines condition on ordering in the current or any future week.

Each week, consumers choose plans and usage to maximize expected discount utility from grocery shopping, net of switching costs, conditional on available information I_{it} :

$$(1) \quad \max_{\{s_{it}(I_{it}), c_{it}(I_{it}, s_{it}, u_{it})\}_{t=1}^{\infty}} E \left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} (U_{ic_{it}\tau}(s_{it}, u_{it}) - \alpha F_{s_{it}} - \delta \mathcal{J}(s_{it} \neq s_{\tau-1})) \mid I_{it} \right],$$

where $c_t \in \{0, 1\}$ is the consumer's usage choice in period t ($c_t = 1$ for the online grocer), $s_t \in \{0, \dots, M\}$ is the subscription (i.e., tariff) choice, u_{it} is a vector of i.i.d. shocks to utility from each of the usage choices, β is the weekly discount factor, α is the constant marginal utility of money, δ is the switching cost, and $\mathcal{J}$ is an indicator function. Importantly, the consumer knows u_{it} before the choice of c_{it} but not before the choice of s_{it} .

The utility consumer i obtains from using only the traditional grocery store in period t is simply the idiosyncratic shock: $U_{i0t} = u_{i0t}$. The utility from using the online grocer is

$$(2) \quad U_{i1t} = \mu_i + \varepsilon_{it} - \alpha p_{sit} + u_{i1t},$$

where u_{i1t} is the idiosyncratic shock, p_{sit} is the per-use component of tariff s_{it} , and $\mu_i + \varepsilon_{it}$ is the experience signal, which is noisy because of variation in the provision of the service (e.g., the quality of the fresh produce, the time it took for the delivery to arrive).³ The consumer uses the noisy signal to learn about his or her match-value μ_i .

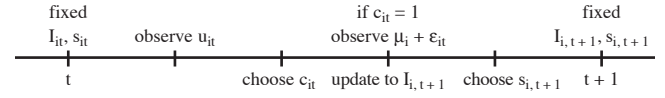
Following Eckstein, Horsky, and Raban (1988), we specify a Bayesian learning process that exploits conjugate distributions, as described in DeGroot (1970). In particular, $\varepsilon_{it} \sim \text{i.i.d. } N(0, \sigma_{\varepsilon}^2)$ with a prior $\mu_i \sim N(m_{it}, \sigma_{it}^2)$ yields a learning process in which the posterior on μ_i after an experience signal $\mu_i + \varepsilon_{it}$ is $\mu_i \sim N(m_{i,t+1}, \sigma_{i,t+1}^2)$, where

$$(3) \quad m_{i,t+1} = \frac{\sigma_{\varepsilon}^2 m_{it} + \sigma_{it}^2 (\mu_i + \varepsilon_{it})}{\sigma_{\varepsilon}^2 + \sigma_{it}^2} \quad \text{and} \quad \sigma_{i,t+1}^2 = \frac{\sigma_{\varepsilon}^2 \sigma_{it}^2}{\sigma_{\varepsilon}^2 + \sigma_{it}^2}.$$

This model of Bayesian learning is tractable because these closed-form expressions for the posterior mean m_{it} and posterior variance σ_{it}^2 capture all the consumer's information regarding μ_i . Because the posterior is normal, the subsequent update follows the same process, with this posterior serving as the prior. Information at t is therefore $I_{it} = (m_{it}, \sigma_{it}^2)$.

Consumers choose their subscription plans prior to observing the idiosyncratic shocks u .⁴ Without loss of generality, 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. Thus, s_{it} is a state variable chosen at the end of period $t-1$. The consumer changes plans if an experience signal shifts his or her

beliefs enough to warrant incurring the switching costs. The sequence is summarized in the following timeline:



Expected discounted utility as a function of the current state, assuming optimal policies are implemented each period, is given by Bellman's equation:

$$(4) \quad V_u(m_{it}, \sigma_{it}, s_{it}, u_{it}) = \max_{c_{it}, s_{i,t+1}} E[U_{ic_{it}} + \beta V_u(m_{i,t+1}, \sigma_{i,t+1}, s_{i,t+1}, u_{i,t+1}) \mid m_{it}, \sigma_{it}, s_{it}, u_{it}, c_{it}],$$

where the expectation is over next period's state and the current period's experience signal $\mu_i + \varepsilon_{it}$. Following Rust (1987), we integrate over the i.i.d. u shocks to remove them from the state space as they only affect current utility. Assuming u are Type I extreme value, this integration has an analytic solution. The stochastic evolution of beliefs is the remaining unknown determinant of the continuation value. Accordingly, let $\mu_{it} = \mu_i + \varepsilon_{it}$ denote the realized experience signal (when $c_{it} = 1$). In the following integrated value function, the first exponent corresponds to $c_{it} = 0$ and the second exponent corresponds to $c_{it} = 1$:

$$(5) \quad V(m_{it}, \sigma_{it}, s_{it}) = \gamma + \ln[\exp(\beta \max_{s_{i,t+1}} \{V(m_{it}, \sigma_{it}, s_{i,t+1}) - \alpha F_{s_{i,t+1}} - \delta \mathcal{J}(s_{i,t+1} \neq s_{it})\}) + \exp(m_{it} - \alpha p_{sit} + \beta \int \max_{s_{i,t+1}} V(m_{i,t+1}(m_{it}, \sigma_{it}, \mu_{it}), \sigma_{i,t+1}(\sigma_{it}, s_{i,t+1}) - \alpha F_{s_{i,t+1}} - \delta \mathcal{J}(s_{i,t+1} \neq s_{it})) \Phi(d\mu_{it} \mid m_{it}, \sigma_{it}))],$$

where γ is Euler's constant and $m_{i,t+1}(\cdot)$ and $\sigma_{i,t+1}(\cdot)$ are the posterior mean and variance from Equation 3. The perceived distribution of signals Φ is normal with mean m_{it} and variance $\sigma_{it}^2 + \sigma_{\varepsilon}^2$ to account for both the signal's noise and the uncertainty of current beliefs.

When $c_{it} = 0$, the beliefs μ_{it} and σ_{it} do not change because no signal is received. When $c_{it} = 1$, the maximization that determines $s_{i,t+1}$ occurs inside the integration over the experience signal because consumers make this choice after observing the signal. Our model of tariff choice differs from Lambrecht and Skiera (2006), Lambrecht, Seim, and Skiera (2007), and Iyengar, Ansari, and Gupta (2007) in that consumers do not have idiosyncratic preferences for tariffs. The optimal tariff in our model maximizes the consumer's expected discounted utility from usage choices. Tariffs themselves do not provide utility in our model; they only affect the cost of usage and have no differential transactions costs.

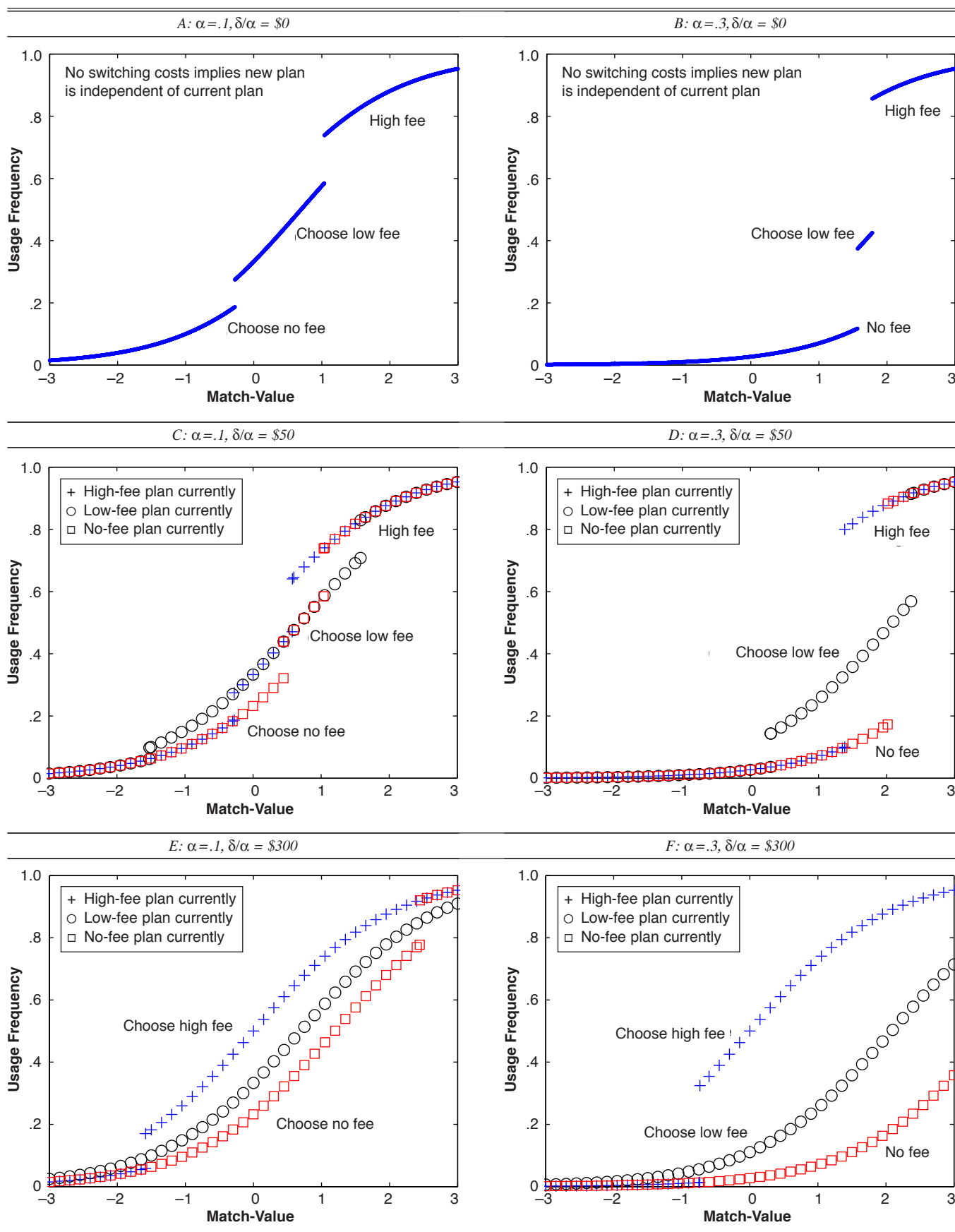
Implications and Comparative Statics

We numerically solve the model and present comparative statics to illustrate the effect of price sensitivity, switching costs, and uncertainty on tariff choice and usage across the three plans, when consumers face the tariff menu observed in our data. In Figure 3, we present consumers' optimal tariffs and usage rates after match-value uncertainty has been

³We do not consider risk aversion in the usage choice, because Coscelli and Shum (2004) show that risk aversion is not identified when priors are estimated separately from the distribution of match-values. We also do not consider risk aversion in the tariff choice because consumers do not face the risk of "overage" charges, as they do with three-part tariffs for cellular phone service.

⁴Clay, Sibley, and Srinagesh (1992) compare *ex post* pricing (i.e., choosing tariffs after usage occurs) with *ex ante* tariff choice. They find that *ex ante* choice yields higher profits for low to moderate levels of uncertainty.

Figure 3
OPTIMAL NEW PLAN AND USAGE GIVEN CURRENT PLAN AND NO REMAINING UNCERTAINTY



fully resolved as functions of the match-value (on the x-axis) and the consumer's current plan (distinguished by plotting symbols). The price coefficient, α , is .1 for plots in the left column and .3 in the right column. Switching costs in dollar terms, δ/α , are zero for the top row, \$50 for the middle row, and \$300 for the bottom row.

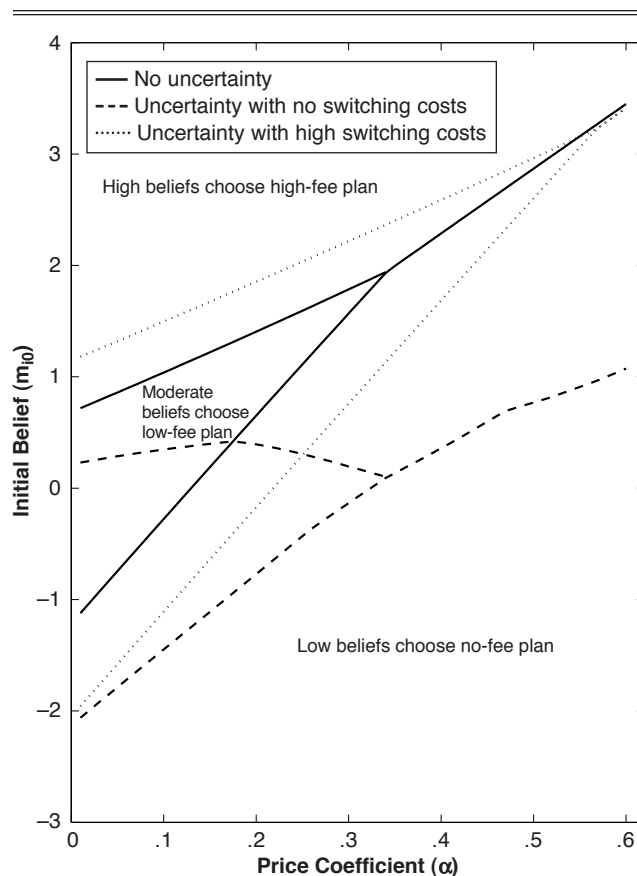
In the top row, we do not distinguish across current plans because without switching costs, plan choice is independent of the current plan. Consumers with low match-values choose the no-fee plan, consumers with high match-values choose the high-fee plan, and moderate match-value consumers choose the low-fee plan. With known match-values, usage frequencies by plan necessarily fall within the ranges, depicted in Figure 1, for which the chosen plan minimizes expected delivery costs, $p + (F/\text{usage probability})$. Usage increases monotonically with match-value, with discrete jumps at the thresholds separating no-fee from low-fee subscribers and low-fee from high-fee subscribers. The usage jumps increase in α because higher α implies a greater sensitivity to the different marginal prices. The thresholds also shift to the right as α increases because higher match-values are needed to justify using the service. Moreover, the threshold separating no-fee from low-fee subscribers shifts more than the threshold separating low-fee from high-fee subscribers because the expected cost per use is lower for consumers with higher match-values. Consequently, the range of match-values choosing the low-fee plan shrinks as α increases. Indeed, no consumers choose the low-fee plan when α exceeds \$.3414.

The lower plots demonstrate the effect of switching costs on tariff choice, given a previous plan. With moderate switching costs of \$50 and low price sensitivity, high-fee subscribers will switch to either the low-fee or no-fee plan if their realized match-value falls sufficiently from the relatively high prior that induced the consumer to initially choose the high-fee plan. Moderate switching costs with high price sensitivity, however, lead the high-fee subscriber to either stay on his or her current plan or to switch to the no-fee plan. The higher α reduces usage on the low-fee plan enough that switching to it is never worth \$50.

In Panels E and F of Figure 3, we show that with high switching costs, consumers never switch to or from the low-fee plan and only switch across the no-fee and high-fee plans in extreme cases. Consequently, consumers with similar match-values face different marginal prices if their initial beliefs varied sufficiently that they first chose different plans.

Whereas Figure 3 focuses on tariff choice and usage given a current subscription plan and no remaining uncertainty, Figures 4 and 5 illustrate the determinants of the initial tariff choice, when consumers' match-values are unknown. In Figure 4, we plot the thresholds that divide match-value beliefs on the y-axis, according to the corresponding optimal initial plan, as functions of price sensitivity on the x-axis. We depict three scenarios: no uncertainty and uncertainty with and without switching costs. The no-uncertainty case, the solid line, is the same scenario as Panels A and B of Figure 3. As such, the threshold line dividing no fee from low fee converges to the line dividing low fee from high fee as price sensitivity increases: No consumers choose low fee for $\alpha > .3414$. As α increases, the market share of no fee monotonically increases. When consumers are uncertain

Figure 4
EFFECT OF UNCERTAINTY, SWITCHING COSTS, AND PRICE SENSITIVITY ON INITIAL TARIFF CHOICE



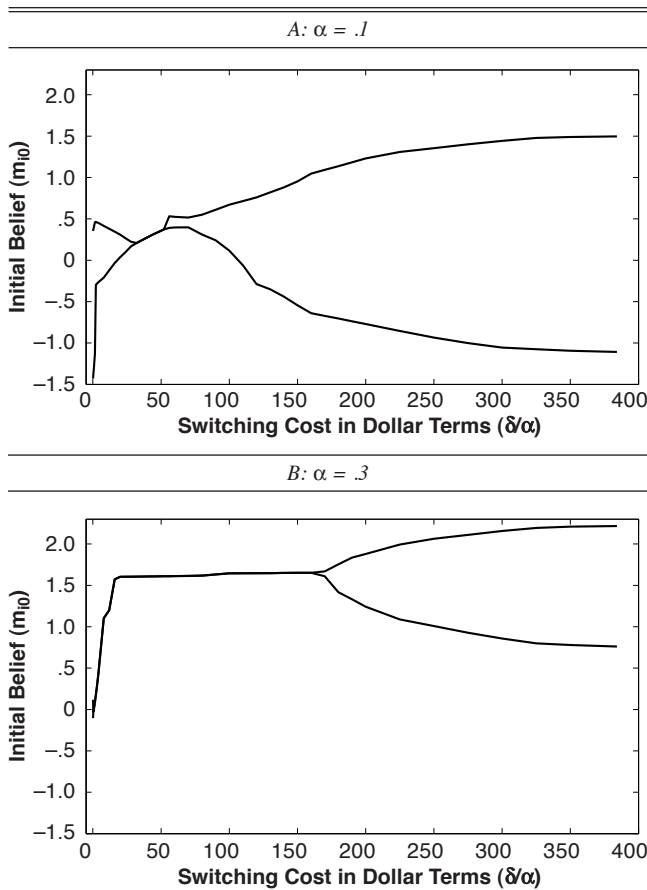
Notes: Upper line of each pair is lowest belief for which high fee is optimal. Lower line of each pair is highest belief for which no fee is optimal. Between each pair are beliefs, if any, for which low fee is optimal.

and face high switching costs (the dotted line), the set of beliefs for which the low-fee plan is optimal expands. The increased attraction of the middle plan is that it enables consumers to at least be near the plan that will be optimal after the uncertainty has been resolved, without incurring the switching costs. When consumers face uncertainty but no switching costs (the dashed line), the threshold lines shift down because consumers know they will increase their usage, relative to the no-uncertainty case, to gain information about their match-value. Higher usage with no cost of being stuck on the wrong plan implies that consumers will shift to the higher-fee plans. Figures 3 and 4 show that α influences both the initial plan shares and the gaps in usage rates across plans. Thus, these moments help identify α when we estimate the model.

The effect of switching costs on initial plan shares and usage conditional on plan (i.e., the allocation of match-values across plans) implies δ is identified by more than just switching behavior. Figure 4 illustrates this effect for two particular values of switching costs. In Figure 5, we put switching costs on the x-axis to provide more details of this effect, with α of .1 in the top plot and .3 in the bottom plot. In both plots, each plan has some subscribers for low

Figure 5

EFFECT OF SWITCHING COSTS ON INITIAL TARIFF CHOICE



Notes: Upper line of each pair is lowest belief for which high fee is optimal. Lower line of each pair is highest belief for which no fee is optimal. Between each pair are beliefs, if any, for which low fee is optimal.

switching costs, although in the bottom plot, the range of beliefs for which low fee is optimal is small and requires very low switching costs. As switching costs increase to moderate levels, the low-fee plan loses its appeal but then becomes attractive again as switching costs increase further. In both plots, switching costs near \$400 are sufficiently high that consumers know they will never switch. Note that switching to the no-fee plan will never occur if δ/α exceeds $F/(1 - \beta)$, the present discounted value of paying F each period.

Initial Beliefs: Aggregate and Conditional Biases

We estimate the distribution of match-values across consumers and consumers' initial beliefs.⁵ For parsimony and to facilitate our discussion of rational expectations, we

⁵Because we do not observe the set of consumers who considered enrolling but chose not to, we estimate the distribution of μ_i conditional on enrolling. Consumers slowly became aware of this service, so we would be uncomfortable using an assumed market size to infer the proportion of consumers who deliberately chose not to sign up. Without the distribution of μ_i for nonenrollees, we cannot predict the full effect of lower prices. However, we can predict the effect of higher prices (because nonenrollees remain nonenrollees), and if lower prices are found to be more profitable for the enrollee population, we can indeed conclude that overall profits would increase because adoption by some nonenrollees would only add to profits.

assume that match-values are distributed $N(\mu_{\text{pop}}, \sigma_{\text{pop}}^2)$, though any distribution may be used. Beliefs are distributed normal, with a common initial prior $\mu_i \sim N(\tilde{\mu}_{\text{pop}}, \tilde{\sigma}_{\text{pop}}^2)$, where $\tilde{\mu}_{\text{pop}}$ and $\tilde{\sigma}_{\text{pop}}^2$ denote consumers' beliefs about the population mean and variance of match-values.

Consumers typically have private information before enrolling, such as their proximity to grocers and computer savvy. We model such information as unbiased private signals $(\mu_i + \varepsilon_{i, \text{pre}}) \sim N(\mu_i, \sigma_{\text{pre}}^2)$, where the subscript pre indicates pre-enrollment. Consumers then update beliefs to $\mu_i \sim N(m_{i0}, \sigma_{i0}^2)$, where

$$(6) \quad m_{i0} = \frac{\sigma_{\text{pre}}^2 \tilde{\mu}_{\text{pop}} + \tilde{\sigma}_{\text{pop}}^2 (\mu_i + \varepsilon_{i, \text{pre}})}{\tilde{\sigma}_{\text{pop}}^2 + \sigma_{\text{pre}}^2}, \quad \sigma_{i0}^2 = \frac{\sigma_{\text{pre}}^2 \tilde{\sigma}_{\text{pop}}^2}{\tilde{\sigma}_{\text{pop}}^2 + \sigma_{\text{pre}}^2}.$$

In empirical learning models, "rational expectations" means consumers know the distribution of match-values. In most applications, such as Miller (1984) and Crawford and Shum (2005), this notion is applied at the product level, implying $\tilde{\mu}_{\text{pop}} = \mu_{\text{pop}}$ and $\tilde{\sigma}_{\text{pop}}^2 = \sigma_{\text{pop}}^2$. For new products, this interpretation is too narrow: Consumers may overestimate the value of a particular new product while still being correct on average across many new products.

In line with Akerberg (2003), we take a broader view of rational expectations and consider the population mean itself distributed normal: $\mu_{\text{pop}} \sim N(\mu_{\text{category}}, \sigma_{\text{category}}^2)$ for some category of products. Rational expectations over the category implies an initial prior with $\tilde{\mu}_{\text{pop}} = \mu_{\text{category}}$ and $\tilde{\sigma}_{\text{pop}}^2 = \sigma_{\text{pop}}^2 + \sigma_{\text{category}}^2$ for each product because the perceived uncertainty $\tilde{\sigma}_{\text{pop}}^2$ accounts for both the dispersion in match-values for the focus product and the uncertainty in its mean match-value.

The posterior mean m_{i0} is unbiased with the usual conditioning on the consumer's information set: $m_{i0} = E(\mu_i | I_{i0})$, where I_{i0} consists of the initial prior and the realized signal. But firms and researchers are usually interested in whether consumers' beliefs are correct regarding a particular product. Therefore, we define product-level bias by also conditioning on μ_{pop} :

$$(7) \quad \text{bias}(m_{i0}; \mu_{\text{pop}}) \equiv m_{i0} - E(\mu_i | I_{i0}, \mu_{\text{pop}}) \\ = m_{i0} - \frac{\sigma_{\text{pre}}^2 \mu_{\text{pop}} + \sigma_{\text{pop}}^2 z(m_{i0})}{\sigma_{\text{pop}}^2 + \sigma_{\text{pre}}^2},$$

where $z(m_{i0})$ is the signal value that leads to m_{i0} when μ_{pop} is unknown (i.e., using Equation 6). The term $[\sigma_{\text{pre}}^2 \mu_{\text{pop}} + \sigma_{\text{pop}}^2 z(m_{i0})]/(\sigma_{\text{pop}}^2 + \sigma_{\text{pre}}^2)$ is the posterior mean if μ_{pop} were known (i.e., $[\tilde{\mu}_{\text{pop}}, \tilde{\sigma}_{\text{pop}}^2] = [\mu_{\text{pop}}, \sigma_{\text{pop}}^2]$), which therefore equals $E(\mu_i | I_{i0}, \mu_{\text{pop}})$.⁶

Product-level biases do not imply irrational or suboptimal behavior. Rational expectations require beliefs to be unbiased given consumers' information sets, but product-level bias

⁶By inverting m_{i0} in Equation 6 for the signal $\mu_i + \varepsilon_{i, \text{pre}}$ and plugging the resulting expression for $z(m_{i0})$ into Equation 7, we can solve for the m_{i0} that yields zero bias:

$$m_{i0}^{\text{NoBias}} = \mu_{\text{pop}} - \frac{(\tilde{\mu}_{\text{pop}} - \mu_{\text{pop}})\sigma_{\text{pop}}^2}{\tilde{\sigma}_{\text{pop}}^2 - \sigma_{\text{pop}}^2}.$$

When $\tilde{\sigma}_{\text{pop}}^2 > \sigma_{\text{pop}}^2$, all consumers with $m_{i0} > m_{i0}^{\text{NoBias}}$ have positive product-level bias. Because μ_{pop} is unknown, m_{i0}^{NoBias} is also unknown: Consumers cannot infer when their m_{i0} is biased.

conditions on the unknown μ_{pop} . Nonetheless, we follow the convention of using “bias” to describe product-level discrepancies between expectations and (average) realizations.

We decompose product-level bias into two components: aggregate bias due to $\tilde{\mu}_{\text{pop}} \neq \mu_{\text{pop}}$ shifting m_{i0} equally for all consumers and conditional bias due to $\tilde{\sigma}_{\text{pop}}^2 \neq \sigma_{\text{pop}}^2$ shifting m_{i0} by an amount that depends on each consumer’s signal. For new products, $\tilde{\sigma}_{\text{pop}}^2 > \sigma_{\text{pop}}^2$, which increases the signal’s weight, causing consumers with high signals for a given product to be overly optimistic, on average, for that product.⁷

We illustrate product-level bias and its implications for tariff choice in Figure 6. For this example, we use $\mu_{\text{category}} = 0$, $\sigma_{\text{category}}^2 = 3$, $\sigma_{\text{pop}}^2 = 1$, $\sigma_{\text{pre}}^2 = 1$, and consider three products with μ_{pop} realizations of -2 , 0 , and 2 . If consumers knew these μ_{pop} realizations, their priors would be $N(-2, 1)$, $N(0, 1)$, and $N(2, 1)$, respectively, and their posteriors would be unbiased at the product level. Because the μ_{pop} are unknown, the prior for each product is $N(0, 1 + 3)$ and the posteriors exhibit product-level biases as plotted in the top graph.

The quality shock μ_{pop} affects all consumers equally, shifting bias ($m_{i0}; \mu_{\text{pop}}$) up and down. We refer to this shift as “aggregate bias.” The posterior’s aggregate bias is less than the difference between $\tilde{\mu}_{\text{pop}}$ and μ_{pop} because the informative signal partially corrects beliefs.

The degree of uncertainty about μ_{pop} (i.e., $\sigma_{\text{category}}^2$) determines the slope of product-level bias. As $\sigma_{\text{category}}^2$ increases, consumers place more weight on their private signals, which increases the relationship between product-level bias and m_{i0} . We refer to this variation in bias across beliefs as “conditional bias,” because the bias is conditional on one’s signal.

Consider the case of no aggregate bias (i.e., $\tilde{\mu}_{\text{pop}} = \mu_{\text{pop}}$). Despite having unbiased priors and signals, consumers’ posteriors exhibit product-level bias: On average, those with signals above μ_{pop} will be overly optimistic, and those with signals below μ_{pop} will be overly pessimistic, as illustrated by the middle line in Panel A of Figure 6. Thus, posteriors can exhibit product-level biases even when both the prior and the signal are unbiased.

Conditional bias is particularly relevant for models of tariff choice because consumers use threshold rules: Consumers with beliefs above some threshold m^* choose the flat-rate

plan, whereas consumers with $m_{i0} < m^*$ choose plans with lower fixed fees and higher per-use charges. Therefore, consumers on the flat-rate plan will have higher bias than consumers on other plans. If the product is weakly worse than average (i.e., $\mu_{\text{pop}} \leq \mu_{\text{category}}$), consumers on the flat-rate plan will necessarily have positive bias. If the product is better than average (i.e., has a negative aggregate bias due to $\mu_{\text{pop}} > \mu_{\text{category}}$), consumers on the flat-rate plan will, on average, have a positive bias only if threshold m^* is sufficiently high.

We illustrate this result in Panel B of Figure 6, which reports for each product the average product-level bias for consumers with m_{i0} exceeding hypothetical m^* thresholds. At the lowest cutoff, few consumers are excluded, so the lowest reported bias is approximately the average across the population (i.e., the aggregate bias). The average product-level bias across all consumers is 0 for the product with no aggregate bias, as expected, and is approximately .5 and $-.5$ for products with quality shocks of -2 and 2 , respectively. The average bias conditional on $m_{i0} > m^*$ increases in m^* and, for the pessimistic example with $\mu_{\text{pop}} = 2$, becomes positive at $m^* = 1.9$.⁸

Finally, Panel C of Figure 6 presents the share of consumers who are optimistic, conditional on their posterior mean exceeding the hypothetical m^* . Again, the lowest cutoff values exclude few consumers, so the lowest shares are approximately population measures. With $\mu_{\text{pop}} = 0$, half the consumers are optimistic; with the negative quality shock, more than half are optimistic; and with the positive quality shock, less than half are optimistic. Because of conditional bias, the share of consumers who are optimistic increases in the threshold m^* .⁹ Even for products with negative aggregate biases (i.e., $\tilde{\mu}_{\text{pop}} < \mu_{\text{pop}}$), a threshold exists above which consumers tend to be overly optimistic. As such, conditional bias can explain a flat-rate bias even when consumers are on average pessimistic.

Thus far, we have presented a rational expectations framework in which conditional bias provides a basis for the oft-cited intuition that people with high beliefs for a product tend to be overly optimistic and often erroneously choose flat-rate tariffs. However, conditional biases can also arise from behavioral conjectures that violate strict rationality. The high weight on private signals that causes conditional bias may, for example, reflect projection bias.¹⁰ In

⁷Expressions for each of these components can be obtained from Equation 7 by noting that m_{i0} in Equation 6 may be written as follows:

$$m_{i0} = \left(1 - \frac{\sigma_{\text{category}}^2}{\sigma_{\text{pop}}^2 + \sigma_{\text{category}}^2 + \sigma_{\text{pre}}^2} \right) \frac{\sigma_{\text{pre}}^2 [\mu_{\text{pop}} + (\tilde{\mu}_{\text{pop}} - \mu_{\text{pop}})] + \sigma_{\text{pop}}^2 (\mu_i + \varepsilon_{i,\text{pre}})}{\sigma_{\text{pop}}^2 + \sigma_{\text{pre}}^2} + \frac{\sigma_{\text{category}}^2 (\mu_i + \varepsilon_{i,\text{pre}})}{\sigma_{\text{pop}}^2 + \sigma_{\text{category}}^2 + \sigma_{\text{pre}}^2},$$

which implies

$$\text{bias}(m_{i0}; \mu_{\text{pop}}) = \frac{\sigma_{\text{pre}}^2 (\tilde{\mu}_{\text{pop}} - \mu_{\text{pop}})}{\sigma_{\text{pop}}^2 + \sigma_{\text{pre}}^2} + \frac{\sigma_{\text{category}}^2}{\sigma_{\text{pop}}^2 + \sigma_{\text{category}}^2 + \sigma_{\text{pre}}^2} \left\{ \mu_i + \varepsilon_{i,\text{pre}} - \frac{\sigma_{\text{pre}}^2 [\mu_{\text{pop}} + (\tilde{\mu}_{\text{pop}} - \mu_{\text{pop}})] + \sigma_{\text{pop}}^2 (\mu_i + \varepsilon_{i,\text{pre}})}{\sigma_{\text{pop}}^2 + \sigma_{\text{pre}}^2} \right\}.$$

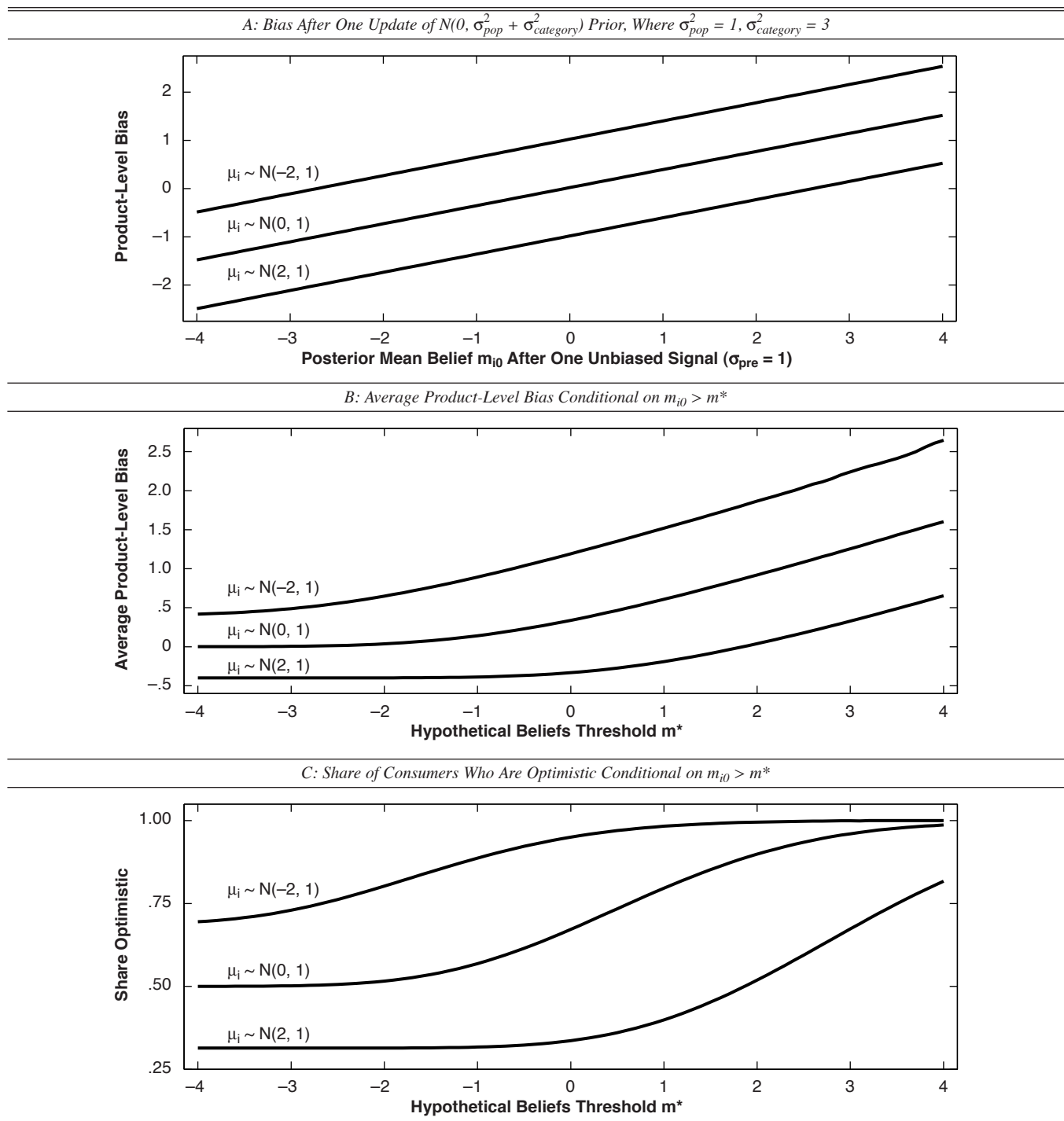
The first term captures aggregate bias; the second term is the conditional bias.

⁸Averaging $\text{bias}(m_{i0}; \mu_{\text{pop}})$ across all products (i.e., μ_{pop} realizations) yields zero for all m_{i0} when each is weighted by the product’s density of consumers at m_{i0} . The average bias conditional on $m_{i0} > m^*$ is therefore also zero when averaging across products. Although consumers who choose flat-rate tariffs have unbiased beliefs across products, for most products their product-level biases are positive. We obtain the overall bias of zero by balancing positive product-level biases for many products with relatively low weights (i.e., lower density of $m_{i0} > m^*$ due to lower μ_{pop}) with negative product-level biases for fewer products with relatively high weights (i.e., higher density of $m_{i0} > m^*$ due to higher μ_{pop}).

⁹Although the share of consumers who are optimistic about a given product is increasing in their posterior m_{i0} , across all products, exactly half of consumers are optimistic regardless of their m_{i0} , which is consistent with posteriors being unbiased across products.

¹⁰Loewenstein, O’Donoghue, and Rabin (2003) attribute biases about future demand to projection bias, in which consumers overestimate the similarity of their current and future preferences. Consistent with the projection bias hypothesis, Conlin, O’Donoghue, and Vogelsang (2007) report that consumers are more likely to return catalog purchases of winter gear when the temperature is low on the day of ordering. Using experimental data, Lovett, Boulding, and Staelin (2009) also find that consumers overweight private signals.

Figure 6
CONDITIONAL BIASES FOR NEW PRODUCTS



most data, including ours, the behavioral and fully rational scenarios are observationally equivalent. The source of product-level bias does not affect our estimator. Therefore, we separately estimate consumers' initial beliefs ($\tilde{\mu}_{pop}, \tilde{\sigma}_{pop}$) and the distribution of match-values (μ_{pop}, σ_{pop}) and remain agnostic about whether their differences reflect behavioral misjudgments or the rational expectations structure with uncertainty about μ_{pop} .

The special case of $\tilde{\sigma}_{pop}^2 = \infty$ is of particular interest. With flat priors, the posterior after processing one signal is cen-

tered around the private signal: $\mu_i \sim N(\mu_i + \varepsilon_{i, pre} \tilde{\sigma}_{pre}^2)$. This specification maximizes conditional bias and eliminates aggregate bias because $\tilde{\mu}_{pop}$ receives no weight in the posterior (and is therefore not identified). Narayanan, Chintagunta, and Miravete (2007) use this specification to study tariff choice for local telephone service. For models with severe conditional bias, this specification is appealing with two fewer parameters and a similar ability to fit the data, but it is nonetheless a restriction that should be tested.

Consumers' Initial Tariff Choices

Given preenrollment beliefs m_{i0} , the consumer chooses the initial tariff s_{i0} that maximizes expected discounted utility. After enrolling, consumers obtain an unbiased signal of μ_i based, for example, on their experience installing the software or perusing the items available to purchase. To account for this anticipated postenrollment signal, we integrate over the signal value $\mu_{i0} = \mu_i + \varepsilon_{i, \text{post}}$ and consider the option of switching tariffs (i.e., $s_{i1} \neq s_{i0}$) given the realized signal. The optimal initial tariff therefore solves

$$(8) \quad \max_{s_{i0} \in \{1, 2, 3\}} \int \max_{s_{i1}} \{V[m_{i1}(\mu_{i0}), \sigma_{i1}, s_{i1}] - \alpha F_{s_{i1}} - \delta \mathcal{J}(s_{i1} \neq s_{i0})\} \Phi(d\mu_{i0} | m_{i0}, \sigma_{i0}),$$

where $m_{i1}(\mu_{i0}) = (\sigma_{\text{post}}^2 m_{i0} + \sigma_{i0}^2 \mu_{i0}) / (\sigma_{\text{post}}^2 + \sigma_{i0}^2)$ is the posterior mean and $\sigma_{i1}^2 = (\sigma_{\text{post}}^2 \sigma_{i0}^2) / (\sigma_{\text{post}}^2 + \sigma_{i0}^2)$ is the posterior variance. The perceived distribution of signals Φ is normal with mean m_{i0} and variance $\sigma_{\text{post}}^2 + \sigma_{i0}^2$ to account for both the noise in the signal and the uncertainty of current beliefs.

After choosing s_{i0} , a consumer realizes the postenrollment signal ($\mu_i + \varepsilon_{i, \text{post}}$), updates his or her beliefs to $\mu_i \sim N(m_{i1}, \sigma_{i1})$ and then chooses his or her continuation plan s_{i1} according to the maximization embedded in the integral in Equation 8. The value function in Equation 5 governs her subsequent behavior. The postenrollment signal is important for explaining why some consumers enroll on tariffs with high fixed fees and never consume the product. However, this signal plays no role in our explanation of the flat-rate bias.

ESTIMATION

In the Appendix, we derive the likelihood function and discuss our simulation estimator. Because beliefs and signals are unobserved, we use Monte Carlo methods to numerically integrate over the distribution of possible signal realizations. The likelihood involves the joint probabilities of consumers' usage and tariff choices. Because tariff choice is deterministic given beliefs, the probability of a consumer's tariff choice is the mass of beliefs for which the chosen tariff is optimal. To efficiently compute this probability and avoid throwing away signal draws for which the chosen tariff is not optimal, we explicitly solve for the regions of beliefs that map into each tariff being optimal and use a Geweke-Hajivassiliou-Keane-like importance sampler to draw signals only from regions for which posteriors are consistent with the chosen tariffs.

Our likelihood also addresses one other nonstandard feature: Consumers may stop using the online grocer because of an exogenous shock such as moving, marriage, divorce, childbirth, retirement, and so forth. To account for such shocks, we assume each household exogenously quits in each period with probability γ . Allowing for these shocks increases the probability of the string of nonusage weeks at the end of each consumer's usage history.

In the Appendix, we also provide details for using the importance-sampling methodology of Akerberg (2009) to estimate random coefficients.

We now discuss identification. The match-value distribution parameters μ_{pop} and σ_{pop}^2 are identified by consumers' usage rates at the end of their histories, after experience signals have eliminated much of the uncertainty. The extent to which initial beliefs are biased, and whether the bias is

aggregate (i.e., $\tilde{\mu} \neq \mu$) or conditional (i.e., $\tilde{\sigma}_{\text{pop}}^2 \neq \sigma_{\text{pop}}^2$), depends on the degree to which average usage rates adjust over time and how the adjustments vary across plans. 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 a significant conditional bias. We identify the variance of the preenrollment signal σ_{pre}^2 by the degree of heterogeneity in *ex ante* beliefs needed to fit variation in consumers' initial tariff choices and initial usage rates. We identify the variance of the postenrollment signal σ_{post}^2 by the degree to which consumers initial usage rates appear inconsistent with their tariff choice. For example, a (relatively) high σ_{post}^2 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_e , the informativeness of the experience signals.

With no observed plan switches, we might expect infinite switching costs. However, high switching costs reduce the model's ability to explain nonusage between a consumer's last order and the end of our sample. Long trailing periods of nonusage are possibly a result of consumers quitting (i.e., switching to plan 3) and never ordering again. Such events only receive significant weight in the integration over the unobserved plan during trailing weeks if switching costs are not too high. We also identify switching costs by their effect on consumers' initial tariff choices, as depicted by the comparative statics.

Gaps in usage rates across consumers on different plans, as depicted in Figure 3 and discussed in our presentation of comparative statics, identify the price coefficient. The discount factor β is identified by the degree to which early usage exceeds later usage, even for plan 3 enrollees, because the information value of early uses increases in β . However, β would not be identified without parametric assumptions on utility, as Rust (1994) shows.

RESULTS

Table 2 presents estimates for six specifications. The first is the full model with both aggregate and conditional product-level biases and forward-looking consumers. The next four are special cases of the full model: conditional bias only ($\tilde{\sigma}_{\text{pop}} = \infty$), aggregate bias only ($\tilde{\sigma}_{\text{pop}} = \sigma_{\text{pop}}$), no product-level biases ($\tilde{\mu}_{\text{pop}} = \mu_{\text{pop}}, \tilde{\sigma}_{\text{pop}} = \sigma_{\text{pop}}$), and myopic consumers ($\beta = 0$). The additional parameters of the full model, $\tilde{\mu}_{\text{pop}}$, $\tilde{\sigma}_{\text{pop}}$, and β , significantly improve the log-likelihood, as revealed in the bottom row of the table. Thus, chi-square tests reject $\tilde{\mu}_{\text{pop}} = \mu_{\text{pop}}, \tilde{\sigma}_{\text{pop}} = \sigma_{\text{pop}}$, and $\beta = 0$ at the .01 significance level. The final specification adds random coefficients to the full model.

Figure 7 depicts key moments of the data and their simulated values for each of the first five models. Although we do not observe when consumers quit by switching (see the "Data" section), we know when such quits occur in simulated data. Thus, in addition to usage rates and initial plan shares, we report the share of consumers on each plan who switch (and order again) and the share who quit (i.e., switch to plan 3 and never order again).

The full model in Figure 7, Panel B, indeed replicates the data. The model accurately predicts usage by enrollees in each plan and in each week, except week 1. The very steep declines in usage from week 1 to week 2 by consumers on plans 2 and 3, compared with the subsequent declines in the data, suggest

Table 2
PARAMETER ESTIMATES

Parameter		Full Model	Conditional Bias Only	Aggregate Bias Only	$\beta = 0$	No Biases	Random Coefficients	
							Mean θ_i	SD θ_i
μ_{pop}	(mean match quality)	-1.687 (.111)	-4.049 (.112)	-6.871 (.084)	-.604 (.079)	-3.228 (.077)	-3.049 (.084)	2.255 (.089)
$\tilde{\mu}_{\text{pop}}$	(perceived mean match quality)	-6.473 .589	N.A.	-.872 (.070)	-47.357 (7.785)	μ_{pop} (.077)	-4.627 (.316)	2.393 (.278)
σ_{pop}	(SD match quality)	2.200 (.047)	2.839 (.061)	5.849 (.075)	1.524 (.039)	7.209 (.110)		
$\tilde{\sigma}_{\text{pop}}$	(perceived SD match quality)	17.549 (1.135)	∞	σ_{pop}	28.306 2.252	σ_{pop}	14.237 (.257)	3.313 (.159)
σ_{pre}	(SD preenroll signal)	23.565 (5.473)	7.313 (.133)	10.099 (.368)	14.018 (.587)	8.286 (.244)	10.170 (.215)	2.608 (.182)
σ_{post}	(SD postenroll signal)	4.100 (.081)	5.034 (.079)	4.076 (.060)	6.645 (.234)	3.643 (.041)	5.233 (.102)	2.099 (.110)
σ_e	(SD experience signal)	6.200 (.130)	7.621 (.148)	5.718 (.094)	4.164 (.083)	4.616 (.110)	8.272 (.185)	2.791 (.172)
β	(weekly discount factor)	.974 (.001)	.987 .001	.998 (.001)	0	.991 (.001)	.981	
α	(price sensitivity)	.358 (.013)	.176 (.010)	.030 (.003)	.273 (.009)	.099 (.009)	.234 (.004)	.052 (.003)
δ	(switching cost)	74.468 (2.925)	66.658 (6.521)	49.466 (4.170)	∞	49.039 (7.151)	69.504 (1.097)	13.434 (.945)
	log-likelihood	-50,604.7	-50,714.7	-51,500.0	-50,806.7	-52,313.3	-50,502.5	

Notes: We fix β in the random coefficients model to avoid convergence delays when a random β draw is close to 1. We started the random coefficients estimation from two local maxima of the full model. The run based on the local maximum with $\beta = .981$ yielded the highest likelihood. In the random coefficients model, the standard deviation of μ_{pop} serves the role of σ_{pop} . All models use an exogenous weekly quit rate of $\gamma = .003$. $\tilde{\mu}_{\text{pop}}$ is not identified in the conditional-bias-only model because it receives no weight in the posterior. Estimated switching costs when $\beta = 0$ deter all switching, so we fix $\delta = \infty$. Asymptotic standard errors are in parentheses.

that an element not present in our model may drive week 1 usage. In particular, consumers could sign up online as part of placing their first order. To avoid possible specification error of the week 1 usage, we ignore these usage choices (i.e., $\Pr[c_{it} | m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta] = 1$ for $t = 1$ in Equation A1).

For the models with conditional bias, we might expect usage of consumers on plan 3 to increase as they correct their pessimistic beliefs. For the full model, 98% of these consumers have $m_{i0} < \mu_i$, and the average bias is -8.73 . However, the postenrollment signal raises their beliefs by an average of 8, such that only 58% are pessimistic immediately before their first usage choice. As a result, relative to their extremely low (unobserved) usage rates before the postenrollment signal, these consumers indeed increase their usage. The observed decline reflects the fact that plan 3 enrollees who receive high postenrollment signals develop positive biases. These optimistic consumers tend to use the service and revise their beliefs down, whereas the pessimistic consumers tend to never use the service and therefore never adjust their beliefs.

The full model also fits consumers' plan choices, despite the absence of plan-specific utility terms. Predicted plan shares are .129, .317, and .554, respectively, compared with actual shares of .123, .319, and .558. Moreover, the model predicts that only .25% of consumers switch and order again, whereas more than half of plan 1 enrollees choose to quit. Exogenous quitting due to $\gamma = .003$ adds to these quits, and is the only mechanism for simulated plan 2 enrollees to quit because estimated switching costs are prohibitively high.

The model with only conditional biases is the second-best-fitting model (of the first five) in terms of log-likelihood

and usage patterns, but it overpredicts the plan 3 share by the largest margin across all models. The model with only aggregate biases does not deliver enough of a decline in usage and predicts that nearly a quarter of plan 3 enrollees and 2.2% of plan 1 enrollees will switch plans and order again. Conditional biases are more important than aggregate biases for explaining our data.

The predicted usage rates of the myopic model, in Figure 7, Panel E, decline much faster than in the real data, and the simulated consumers never quit or switch plans. Both of these aspects of the myopic model are unappealing. In Figure 7, Panel F, we observe that the model without any product-level biases fits the data poorly. With neither aggregate nor conditional biases, the decline in usage over time is due only to exogenous quits and the resolution of uncertainty (which reduces the information-acquisition incentive to purchase).

Having established that the full model fits the data well, we now turn to interpreting its parameter estimates. Consumers had imprecise information about the value of this service for the overall population, as implied by $\tilde{\sigma}_{\text{pop}}/\alpha = 17.549/.358 = 49$. Indeed, $\tilde{\sigma}_{\text{pop}}$ is nearly eight times the degree of heterogeneity in actual match-values ($\sigma_{\text{pop}} = 2.2$). For a new service using new technology, we expect uncertainty to be high. Given $\sigma_{\text{pre}} = 23.565$, the weight on the preenrollment signal is .357, and the weight on $\tilde{\mu}_{\text{pop}} = -6.473$ is .643. Because $\tilde{\mu}_{\text{pop}}$ is lower than μ_{pop} (-1.687), the aggregate bias is negative. The average bias (i.e., average $m_{i0} - \mu_i$) across enrollees in each plan is 10.38, 1.75, and -8.73 , respectively.

Figure 8 illustrates the positive relationship between product-level bias and m_{i0} (i.e., mean belief before enrolling) and the resolution of bias as signals are received.

Figure 7
ACTUAL AND PREDICTED USAGE RATES, BY PLAN

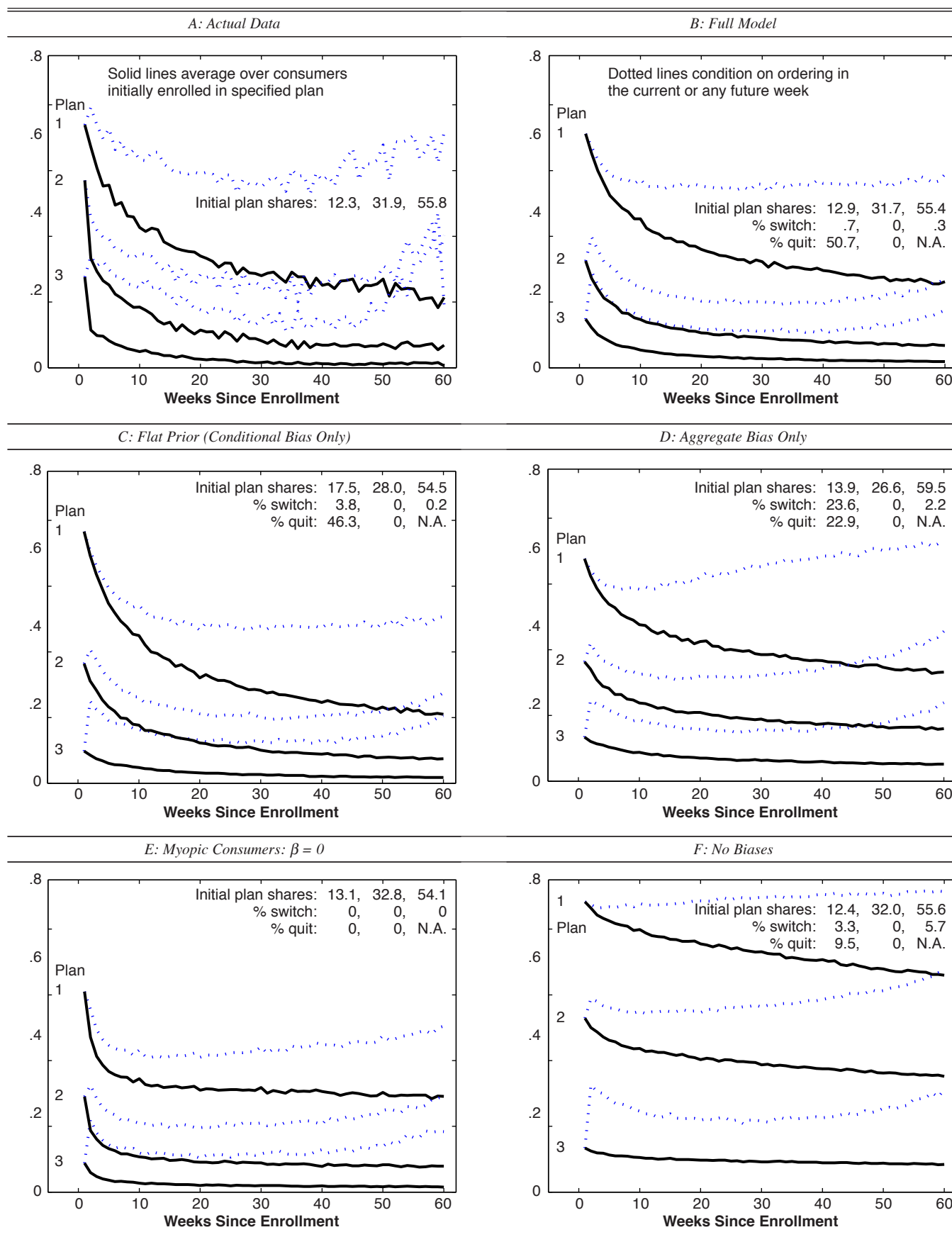
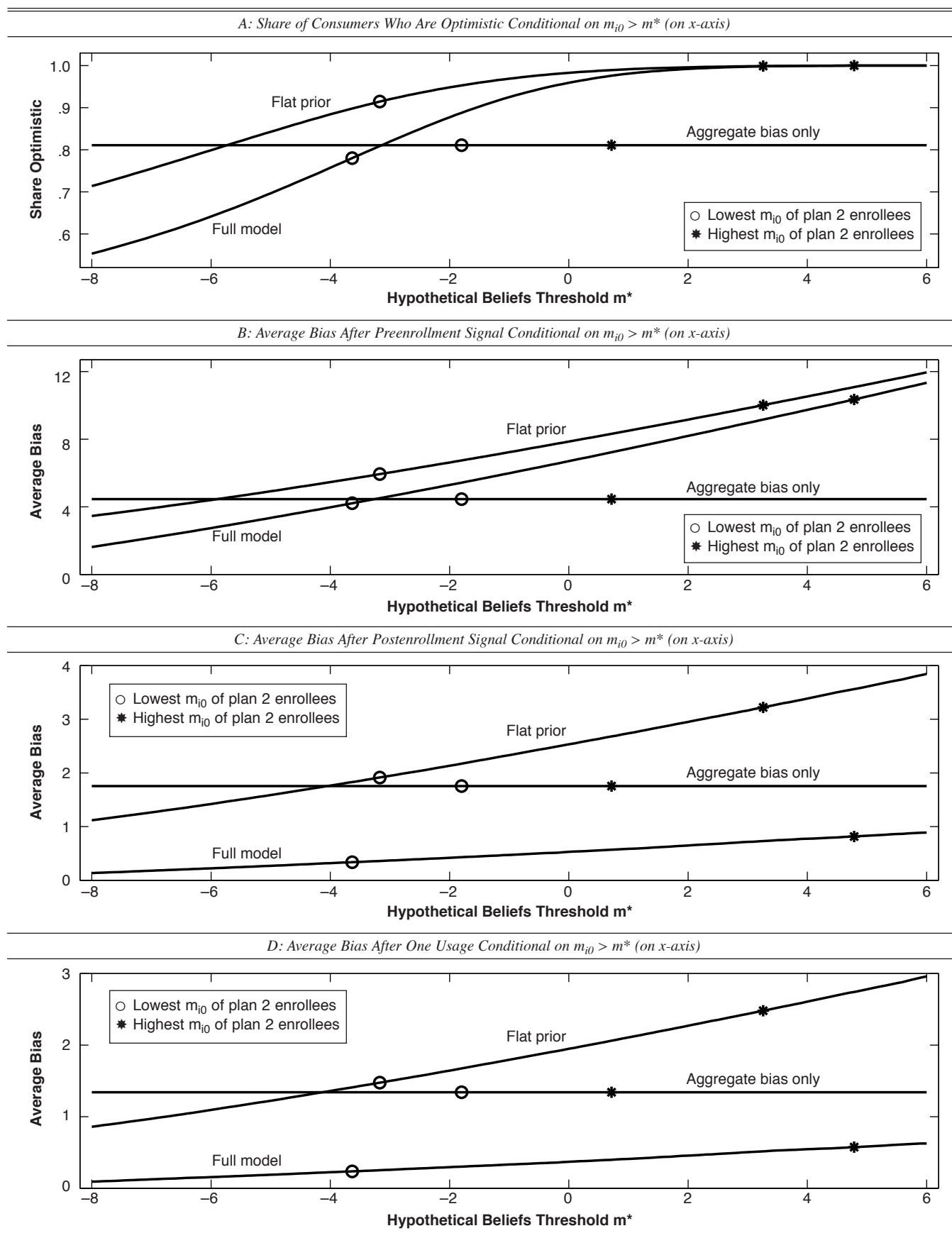


Figure 8
MODEL COMPARISON: PRODUCT-LEVEL BIASES



The top graph plots the share of consumers who are optimistic (i.e., have $m_{i0} > \mu_i$) conditional on their m_{i0} exceeding the hypothetical cutoff m^* that varies along the x-axis. The other three graphs plot the average bias of consumers with m_{i0} exceeding the cutoff m^* at three points in time: after the preenrollment signal, after the postenrollment signal, and after the first experience signal. To provide a sense of the distribution of consumers' m_{i0} , we mark the lowest and highest values of m_{i0} for which plan 2 is chosen (approximately the 55th and 87th percentiles). For comparison, we also show the biases implied by the conditional-bias-only model (i.e., flat prior) and the aggregate-bias-only model. In this latter model, the share of consumers who are optimistic and bias level do not vary across beliefs. For models with conditional biases, the share of consumers who are optimistic and bias levels are increasing in m_{i0} because $\tilde{\sigma}_{\text{pop}}$ exceeding σ_{pop} implies the private signals receive higher weights.

Essentially, all plan 1 enrollees are optimistic in the two models with conditional biases, whereas only 81% are optimistic in the model with aggregate biases only. Fewer plan 2 enrollees are optimistic: 69% for the full model and 86% with the flat prior. Note that the conditional bias in the full model is sufficiently large to overcome the negative aggregate biases ($\tilde{\mu}_{\text{pop}} < \mu_{\text{pop}}$) for both plan 1 and plan 2 enrollees. The lower three panels in Figure 8 reveal the resolution of biases as consumers update beliefs with their preenrollment signal and their first consumption experience.

Although biases conditional on enrollment plan dissipate quickly in the full model, consumers still require many signals to precisely learn their idiosyncratic μ_i . After the post-installation signal, the standard deviation of the posterior is 3.9, or in dollar equivalents (by dividing by $\alpha = .358$), \$11. After three experience signals, this standard deviation is \$7.40 and hits \$4.90 after ten uses. The lack of switching, particularly by plan 3 enrollees who order frequently over long periods, contributes to the estimation of slow learning.

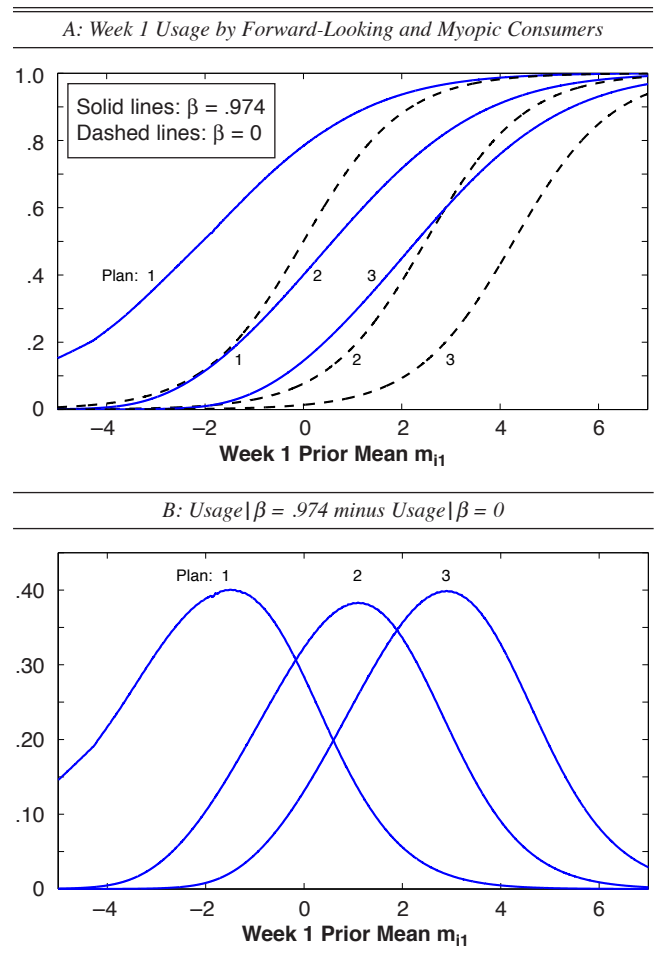
The full model's discount factor is .974. Figure 9 depicts the shift in usage due to information acquisition by forward-looking consumers. In Panel A, the week 1 usage probabilities for forward-looking consumers exceed the usage probabilities for myopic (though otherwise identical) consumers for all values of week 1 beliefs and subscribed plans. The differences in these usage rates reaches .40, as plotted in Panel B.

The implied price sensitivity of consumers is also evident in Figure 9: Usage probabilities, holding beliefs fixed, are substantially higher for consumers on plan 1 than for consumers on the plans with per-use charges.

Given the lack of plan switching, our estimate of δ is high. Dividing the estimate of 74.468 utils by .358 dollars per util implies a switching cost equivalent to \$208. Iyengar, Ansari, and Gupta (2007) estimate even higher "hassle costs" for switching wireless service plans, and Handel (2010) estimates switching costs exceeding \$2,500. As Figure 7, Panel B indicates, we predict that more than half of plan 1 enrollees switch to plan 3 despite the high switching costs.

The last two columns of Table 2 report the means and standard deviations of the random coefficients. The means fall between the estimates of the full model and the model with only conditional bias, except σ_{post} and σ_ϵ are slightly

Figure 9
INFORMATION ACQUISITION



higher. The standard deviations of α , δ , and the σ parameters are 20%–40% of their respective means.¹¹

COUNTERFACTUAL SIMULATIONS

We conduct two sets of counterfactuals. First, we decompose the effects of switching costs and uncertainty on consumer behavior, consumer surplus, and the firm's revenues. Then, we compute optimal menus of two-part tariffs to investigate the effectiveness of menus as screening devices for experience goods. In each counterfactual, we simulate 50,000 consumers over 100 weeks and report discounted revenue and surplus values in dollars per consumer, assuming the firm's annual discount factor is .9.

Decomposing the Effects of Switching Costs and Uncertainty on Outcomes

Table 3 summarizes consumer behavior and surplus and firm's discounted revenues, for various specifications of the

¹¹We might expect priors and match-values to depend on consumers' enrollment dates. We therefore allowed "week of enrollment" to linearly shift μ_{pop} , $\tilde{\mu}_{\text{pop}}$, and their respective standard deviations in the random coefficient model (which accepts conditioning variables without adding dynamic programs to solve). The log-likelihood improves from -50,502.5 to -50,473.3, but the other estimates are virtually unchanged. Given that the random coefficients specification is primarily a robustness check for our finding that menus of tariffs fail to increase revenues, we do not report the estimates.

Table 3
EFFECT OF SWITCHING COSTS, UNCERTAINTY, AND BIASES ON CHOICES, REVENUE, AND SURPLUS

Model Description	Usage: Initial, Final (Plan Share: Initial, Final)			Revenue Discounted	CS
	Plan 1	Plan 2	Plan 3	Rev _{final} (1 - β_{firm})	Realized (Expected)
Full model	.712, .579 (.129, .058)	.324, .071 (.317, .317)	.148, .015 (.554, .625)	497.2 (476.5)	-39.2 (349.0)
No switching costs (full model except $\delta = 0$)	.949, .941 (1.00, .037)	N.A., N.A. (.000, .000)	.011, .010 (.000, .940)	185.4 (166.1)	-10.2 (395.7)
No uncertainty (full model except $\sigma_{\text{pre}} = 0$)	.935, .935 (.044, .044)	N.A., N.A. (.000, .000)	.009, .009 (.956, .956)	175.8 (175.8)	5.6 (5.6)
No-biases model	.883, .913 (.124, .138)	.529, .419 (.320, .320)	.134, .049 (.556, .542)	1196.6 (1190.3)	1415.9 (1415.9)
Aggregate-bias-only model	.677, .738 (.139, .085)	.363, .182 (.266, .266)	.134, .039 (.595, .649)	721.2 (704.1)	6916.6 (25780.1)
Conditional-bias-only model	.766, .492 (.175, .082)	.363, .079 (.280, .280)	.098, .016 (.545, .638)	552.2 (530.0)	-85.4 (856.0)

Notes: Values in parentheses correspond to the label in parentheses in the column header. All revenue and surplus values are in dollars per consumer. Weekly $\beta_{\text{firm}} = .997976$. Thus, one dollar per week has present value of nearly \$500. $\text{Rev}_{\text{final}}/(1 - \beta_{\text{firm}})$ measures the firm's steady-state discounted revenue per consumer. All values were generated by simulating 50,000 consumers over 100 weeks with $\gamma = 0$.

model, when consumers face the actual tariff menu. The first two rows use the full model as estimated. In the long run (i.e., week 100), the firm receives \$.96 per consumer, yielding a steady-state discounted revenue of \$476.5 per consumer. The steady-state revenue is generally lower than the discounted revenue along the transition path because the firm earns revenues from consumers' experiential consumption.

A consequence of consumers not knowing the population distribution of μ_i is that their average realized consumer surplus (CS) differs from their expected surplus. In the base case, consumers lost \$39.2 on average when they expected surplus of \$349.¹² Given $\hat{\mu}_{\text{pop}} < \mu_{\text{pop}}$, one might expect realized CS to exceed expected CS. However, with uncertainty, expected CS is driven primarily by the possibility that μ_i is high, and the actual option value is lower than the perceived option value because σ_{pop} is lower than $\hat{\sigma}_{\text{pop}}$.

The second model in Table 3 removes switching costs by setting δ to zero. Both expected and realized CS increase, though realized CS is still negative because consumers invest too much in experiential consumption as a result of their overestimation of the service's option value. Firms discounted revenues fall dramatically, from \$497.2 per consumer with δ to \$185.4 without δ , as consumers who enroll in plans with fixed fees, given their conditional biases, switch to the fee-less plan 3 after learning their μ_i is lower than expected.

The third model in Table 3 removes uncertainty about match-values by setting $\sigma_{\text{pre}} = 0$, so consumers fully learn μ_i from the preenrollment signal. Discounted revenues of \$175.8 are lower than the previous model with uncertainty and $\delta = 0$ because revenues are no longer received from experimenting consumers. Realized CS equals expected CS and is finally positive, albeit at only \$5.60 per consumer because few consumers use the service. Plan 1 enrollees have CS of a mere \$105 on average, or \$2.68 per week.

¹²Choosing the outside good every week yields expected discounted utility of 22.6 (Euler's constant divided by $1 - \beta$). With the online grocer available, expected discounted utility, averaged across consumers, is 147.4 utils. Expected CS is therefore $(147.4 - 22.6)/\alpha = \349 , or $\$349(1 - \beta) = \9.07 on a weekly basis.

In summary, Table 3 shows that the combination of uncertainty and switching costs largely drive consumers' behavior and surplus and the firm's revenues. The combination of biased beliefs (due to either aggregate or conditional biases) and high switching costs creates a windfall for firms offering tariffs with fixed fees and a costly learning experience for consumers. If consumers' hassle costs of switching are small compared with their cost savings from switching tariffs, firms should consider contracts with termination fees to increase the proportion of consumers who remain on high-fee plans after correcting their beliefs. However, the short-term benefits of using termination fees must be weighed against the potential long-term cost of exploited consumers switching to other firms when the contract ends.

Table 3 also reports simulated outcomes for the models with no bias, only aggregate bias, and only conditional bias. The specification of initial beliefs, which determines product-level biases, has a significant impact on predicted revenues and consumer surplus.

Price Discrimination

We evaluate the use of flat-rate tariffs, two-part tariffs, and menus of two-part tariffs to price discriminate when a firm sells an experience good.¹³ With consumer uncertainty and switching costs, fixed fees play two roles: They extract surplus in the traditional (static) sense from inframarginal units, and they enable firms to capitalize on consumers' biases. The per-use prices determine the degree to which consumers learn: On the one hand, high per-use prices discourage experimentation, thereby reducing the firm's long-term profits because some consumers who should use the product regularly will never discover this. On the other

¹³Firms frequently use two-part tariffs to extract surplus from people who consume multiple units of a given good. In our model, consumers either use the service once or not at all, each week. Nonetheless, because the fee component of the two-part tariff is paid before the consumer's observing the 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.

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 assuming the firm knows the match-value distribution and all demand parameters. Because we focus on consumer uncertainty and learning, we do not consider a need for the firm to discover the demand curve. We assume the firm maximizes expected discounted revenues, which is consistent with profit maximization when net marginal costs are zero, as suggested by industry analysts.¹⁴

Table 4 presents the optimal prices and fees for various tariffs when consumers behave according to the estimated full model. Given the high switching costs, the optimal flat-fee tariff of $F = \$5.32$ generates more than three times the discounted revenue of the optimal uniform price of $p =$

$\$5.39$. In this model, all consumers have the same switching costs, so the optimal fixed fee is just low enough that no one quits (i.e., $F = [1 - \beta]\delta/\alpha$). The optimal single two-part tariff, with $F = \$1.98$ and $p = \$4.09$, yields 20% more revenue than the optimal flat-rate tariff. Adding a second two-part tariff, however, fails to increase revenues.

We also compute optimal tariffs for the random coefficients model to assess whether richer consumer heterogeneity leads to revenue gains with tariff menus. Table 5 presents the results. Although the exact fees and prices differ from those derived from the full model, the broad implications are the same: A fixed fee, either by itself or as part of a two-part tariff, enables the firm to earn approximately four times more than with uniform pricing, and menus offer little to no gain over the optimal two-part tariff. The revenue gain from the second two-part tariff is less than .1%.

In our application, menus of two-part tariffs are unable to segment consumers without letting high-usage consumers retain too much surplus: The incentive compatibility constraints are too costly to satisfy. Courty and Hao (2000) also find that segmenting consumers fails to increase profits when *ex ante* consumer heterogeneity is high.

CONCLUSION

We show that aggregate uncertainty about a new product's quality yields posteriors that exhibit a phenomenon

Table 4
OPTIMAL TARIFFS: FULL MODEL

Tariff Description	Usage: Initial, Final (Plan Share: Initial, Final)			Revenue Discounted	CS
	Plan 1	Plan 2	Plan 3	$\frac{\text{Rev}_{\text{final}}}{(1 - \beta_{\text{firm}})}$	Realized (Expected)
$F_1 = 5.32, p_1 = 0$ (flat-fee tariff)	.662, .332 (.257, .327)			858.0 (858.0)	-49.8 (345.3)
$F_3 = 0, p_3 = 5.39$ (per-use tariff)			.331, .091 (1.0, 1.0)	255.4 (241.5)	-2.0 (404.2)
$F_2 = 1.98, p_2 = 4.09$ (one two-part tariff)		.398, .126 (.814, .826)		1029.8 (1020.2)	-59.6 (352.5)

Notes: Values in parentheses correspond to the label in parentheses in the column header. All revenue and surplus values are in dollars per consumer. Weekly $\beta_{\text{firm}} = .997976$. Thus, one dollar per week has present value of nearly \$500. $\text{Rev}_{\text{final}}/(1 - \beta_{\text{firm}})$ measures the firm's steady-state value. All values were generated by simulating 50,000 consumers over 100 weeks with $\gamma = 0$ for 100. For ease of comparison with the actual menu, single tariffs appear under "Plan 1" if a flat rate, under "Plan 2" if a two-part tariff, and under "Plan 3" if a uniform price. Menus of two-part tariffs offered no advantage over the single two-part tariff. In each simulation, consumers can also choose to not enroll in any plan.

Table 5
OPTIMAL TARIFFS: RANDOM COEFFICIENTS MODEL

Tariff Description	Usage: Initial, Final (Plan Share: Initial, Final)			Revenue Discounted	CS
	Plan 1	Plan 2	Plan 3	$\frac{\text{Rev}_{\text{final}}}{(1 - \beta_{\text{firm}})}$	Realized (Expected)
$F_1 = 4.25, p_1 = 0$ (flat-fee tariff)	.622, .208 (.403, .420)			883.6 (882.2)	-88.4 (662.26)
$F_3 = 0, p_3 = 8.35$ (per-use tariff)			.261, .048 (1.0, 1.0)	220.2 (199.2)	-20.0 (645.0)
$F_2 = 4.25, p_2 = 1.35$ (one two-part tariff)		.604, .183 (.385, .399)		890.8 (886.4)	-89.6 (636.7)
$F_1 = 4.25, p_1 = 1.12$	.621, .194	.447, .117		891.4	-89.5
$F_2 = 4.08, p_2 = 1.55$	(.352, .366)	(.037, .037)		(887.4)	(640.9)

Notes: Values in parentheses correspond to the label in parentheses in the column header. All revenue and surplus values are in dollars per consumer. Weekly $\beta_{\text{firm}} = .997976$. Thus, one dollar per week has present value of nearly \$500. $\text{Rev}_{\text{final}}/(1 - \beta_{\text{firm}})$ measures the firm's steady-state value. Values generated by simulating 5000 consumers over 100 weeks with $\gamma = 0$ for 100 draws of θ . For ease of comparison with the actual menu, single tariffs appear under "Plan 1" if a flat rate, under "Plan 2" if a two-part tariff, and under "Plan 3" if a uniform price. An additional *ex post* only tariff offered no advantage over two two-part tariffs. In each simulation, consumers can also choose to not enroll in any plan.

similar to the winner's curse: Consumers with high posteriors tend to overestimate the product's utility. This conditional bias (i.e., bias conditional on the consumer's posterior) can explain consumers' tariff and usage choices for new products without invoking violations of rationality or explicit preferences for particular types of tariffs.

Either conditional biases or positive aggregate biases (which affect all consumers equally) can explain why consumers on flat-rate plans often fail to use the product enough to justify the fixed fee. Our estimation results indicate that for the online grocer, the aggregate bias was in fact negative. DellaVigna and Malmendier's (2006) finding that biased beliefs best explain consumers' contract choices for health clubs suggests our results may be relevant for other products as well. Even when consumers are pessimistic in general, those with high posteriors can have positive biases.

However, our data cannot distinguish between the rational model of the flat-rate bias and alternative behavioral explanations. For example, with switching costs, consumers may commit to future use of a health club by enrolling in a flat-rate tariff with no per-use charge. We leave such distinctions to future work. Although we have focused on tariff choice, our insights about beliefs and choice are potentially relevant for all new products. In particular, the rational model of conditional bias implies that most new products will fail to fulfill buyers' expectations, even when expectations are correct on average across products.

APPENDIX

In this Appendix, we derive the likelihood function, taking into account the possibility of exogenous quits, and discuss our use of importance-sampling to estimate random coefficients.

The Likelihood Function

Because consumers' match-values and beliefs are not observed, we integrate over μ_i , $\epsilon_{i, \text{pre}}$, $\epsilon_{i, \text{post}}$, and ϵ_{it} . Let θ denote the vector of parameters to estimate, which will be a subset of $(\mu_{\text{pop}}, \tilde{\mu}_{\text{pop}}, \sigma_{\text{pop}}, \tilde{\sigma}_{\text{pop}}, \sigma_{\text{pre}}, \sigma_{\text{post}}, \sigma_{\epsilon}, \beta, \alpha, \delta)$, depending on the particular specification.

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 sample. Because 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

$$(A1) \quad L_i(\theta) = \int \left\{ \Pr[s_{i0} | m_{i0}(\epsilon_i, \mu_i), \sigma_{i0}; \theta] \prod_{t=1}^{\tau_i-1} \Pr[s_{it} | m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{i, t-1}; \theta] \Pr[c_{it} | m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta] \right\}$$

$$\sum_{s_{i\tau_i}} \Pr[s_{i\tau_i} | m_{i\tau_i}(\epsilon_i, \mu_i), \sigma_{i\tau_i}, s_{i\tau_i-1}; \theta] \prod_{t=\tau_i}^{T_i} \Pr[c_{it} | m_{it}(\epsilon_i, \mu_i), \sigma_{it}, s_{i\tau}; \theta] \Phi(d\epsilon_i, d\mu_i; \theta),$$

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

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

$$(A2) \quad V_{0it} = \beta \max_{s_{i, t+1}} \{V(m_{it}, \sigma_{it}, s_{i, t+1}) - \alpha F_{s_{i, t+1}} - \delta \mathcal{J}(s_{i, t+1} \neq s_{it})\}$$

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

$$(A3) \quad V_{1it} = m_{it} - \alpha p_{s_{it}} + \beta \int \max_{s_{i, t+1}} \{V[m_{i, t+1}(m_{it}, \sigma_{it}, \mu_{it}), s_{i, t+1}(\sigma_{it}), s_{i, t+1}] - \alpha F_{s_{i, t+1}} - \delta \mathcal{J}(s_{i, t+1} \neq s_{it})\} \Phi(d\mu_{it} | m_{it}, \sigma_{it}).$$

We use the standard normalizations to fix the scale and level of utility and obtain

$$(A4) \quad \Pr(c_{it} | m_{it}, \sigma_{it}, s_{it}; \theta) = \frac{\exp(V_{c_{it}})}{\exp(V_{0it}) + \exp(V_{1it})}.$$

The plan choice s_{it} is deterministic given beliefs m_{it} and σ_{it} . That is, $\Pr(s_{it} | m_{it}, \sigma_{it}, s_{i, t-1}; \theta)$ equals one if s_{it} is optimal given $(m_{it}, \sigma_{it}, s_{i, t-1})$ and equals zero otherwise. Mathematically,

$$(A5) \quad \Pr(s_{it} | m_{it}, \sigma_{it}, s_{i, t-1}; \theta) = \mathcal{J}\{s_{it} = s(m_{it}, \sigma_{it}, s_{i, t-1}; \theta)\},$$

where $s(m_{it}, \sigma_{it}, s_{i, t-1}; \theta) \equiv \arg\max_s \{V(m_{it}, \sigma_{it}, s) - \alpha F_s - \delta \mathcal{J}(s \neq s_{it})\}$ denotes the optimal tariff that solves the maximization embedded in the continuation value integrands of both Equations 8 and 5. Although this choice is deterministic given beliefs, from the econometrician's perspective, s_{it} is probabilistic because 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 because no more signals are received. Thus, the censored plan is $s(m_{i\tau_i}, \sigma_{i\tau_i}, s_{i\tau_i-1}; \theta)$. Integration over the censored plan is therefore automatically handled by the integration over unobserved beliefs.

We use Monte Carlo simulation with 2000 draws to evaluate $L_i(\theta)$ for each consumer and obtain our estimator by maximizing the product of the consumers' simulated likelihoods, using the nested fixed-point algorithm of Rust (1987). We solve the dynamic model numerically using Gauss-Hermite quadrature (Judd 1998) to integrate over μ_{it} . We discretize m_{it} and use linear interpolation to evaluate V at points off the grid. The posterior variance is a determinis-

tic function of the number of experience signals processed, which we allow to vary from 0 to 99.

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 distribution for which the observed plan choice is indeed optimal and reweight the likelihood by the probability mass of this truncated normal. The sampling scheme is a nonstandard application of the Geweke-Hajivassiliou-Keane importance sampler. As a result of switching costs, the set of experience signals for which a given plan is optimal may contain two noncontiguous regions. 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 because of a zero probability of the tariff choice in Equation A5, which is deterministic given beliefs. Sampling from truncated normals and reweighting also yields a smooth simulated likelihood function because the weights are smooth functions of the parameters, which aids in the numerical optimization and computing of standard errors.

Exogenous Quits

Many households use the online grocer regularly over a long period and suddenly stop. Although such behavior can indicate slow learning, it may also reflect permanent household shocks, such as moving, marriage, divorce, childbirth, retirement, and so on. To account for such shocks, we assume each household exogenously quits in each period with probability γ . We assume consumers are unaware of these possible shocks. To account for exogenous quits in the likelihood, multiply each $\Pr[c_{it} | m_{it}(\varepsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta]$ by $1 - \gamma$ in the third line of Equation A1, and replace $\prod_{t=\tau_i}^{T_i} \Pr[c_{it} | m_{it}(\varepsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta]$ in the last line with

$$(A6) \quad \sum_{t=\tau_i}^{T_i} \gamma^{I(t < T_i)} \{ (1 - \gamma) \Pr[c_{it} | m_{it}(\varepsilon_i, \mu_i), \sigma_{it}, s_{it}; \theta] \}^{1 - \tau_i}.$$

Each trailing week's inactivity may be caused by either an exogenous quit in the current period or a prior period or choosing the traditional store despite still being a subscriber. When 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 he or she exogenously quits at $t = 51$, the probability he or she chooses no usage at $t = 51$ and exogenously quits at $t = 52$, and the probability he or she chooses no usage at both $t = 51$ and $t = 52$. Equation A6 sums the probabilities of these different explanations for the trailing weeks of nonusage.

We fix $\gamma = .003$ based on the frequency of household relocations in our market, which implies a .145 annualized exogenous quit rate. We experimented with estimating γ , but its estimate of .019 (which implies an annualized rate of .63) seems implausibly high. The higher γ increases the switching costs estimate because (censored) endogenous quits are less important for explaining the trailing weeks of nonusage when γ is high. If the data contained observed switches, we would be more inclined to estimate γ .

Parameter Heterogeneity

We use the importance sampling methodology of Akerberg (2009) to vary θ across consumers. Integrating over random coefficients typically involves averaging $L_i(\theta_i)$ over many draws of θ_i for each consumer, which leads to an infeasible number of fixed points to compute because 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. Akerberg suggests computing and retaining $L_i(\theta_i)$ 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,

$$(A7) \quad L_i(\rho) = \int L_i(\theta_i) g(\theta_i | \rho) d\theta_i = \int L_i(\theta_i) \frac{g(\theta_i | \rho)}{h(\theta_i)} h(\theta_i) d\theta_i.$$

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

$$(A8) \quad \hat{L}_i^{NS}(\rho) = \frac{1}{NS} \sum_{ns=1}^{NS} L_i(\theta_i^{ns}) \frac{g(\theta_i^{ns} | \rho)}{h(\theta_i^{ns})}.$$

We center the initial h on estimates without random coefficients and use a high variance so the reweighting (with h in the denominator) does not explode. We then iterate by setting h to the previous iteration's g until the estimated ρ implies g is similar to h . We impose restrictions such as $\sigma_{pop} > 0$ by using truncated normals for g and h . For each consumer, we use 50 draws of the parameters that affect the consumer's dynamic program (i.e., $\tilde{\sigma}_{pop}$, σ_{pre} , σ_{post} , σ_ε , β , α , δ). For each vector of these dynamic parameters, we solve the dynamic program and evaluate the likelihood for 100 draws of $(\mu_i, \tilde{\mu}_{pop}, \varepsilon_i)$.

REFERENCES

- Akerberg, Daniel (2003), "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44 (3), 1007–1040.
- (2009), "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," *Quantitative Marketing and Economics*, 7 (4), 343–76.
- Clay, Karen, David Sibley, and Padmanabhan Srinagesh (1992), "Ex Post vs. Ex Ante Pricing: Optional Calling Plans and Tapered Tariffs," *Journal of Regulatory Economics*, 4 (2), 115–38.
- Conlin, Michael, Ted O'Donoghue, and Timothy Vogelsang (2007), "Projection Bias in Catalog Orders," *American Economic Review*, 97 (4), 1217–49.
- Coscelli, Andrea and Matthew Shum (2004), "An Empirical Model of Learning and Patient Spillovers in New Drug Entry," *Journal of Econometrics*, 122 (2), 213–46.
- Courty, Pascal and Li Hao (2000), "Sequential Screening," *Review of Economic Studies*, 67 (4), 697–717.
- Crawford, Gregory and Matthew Shum (2005), "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73 (4), 1137–74.
- DeGroot, Morris (1970), *Optimal Statistical Decisions*. New York: McGraw-Hill.

- DellaVigna, Stefano and Ulrike Malmendier (2006), "Paying Not to Go to the Gym," *American Economic Review*, 96 (3), 694–719.
- Eckstein, Zvi, Dan Horsky, and Yoel Raban (1988), "An Empirical Dynamic Model of Optimal Brand Choice," working paper, Eitan Berglas School of Economics, Tel-Aviv University.
- Erdem, Tulin and Michael Keane (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1–20.
- Grubb, Michael (2009), "Selling to Overconfident Consumers," *American Economic Review*, 99 (5), 1770–1807.
- Hajivassiliou, Vassilis and Paul A. Ruud (1994), "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics*, Vol. 4, R.F. Engel and D.L. McFadden, eds. Amsterdam: Elsevier Science, 2384–2441.
- Handel, Ben (2010), "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," working paper, Department of Economics, Northwestern University.
- Iyengar, Raghuram, Asim Ansari, and Sunil Gupta (2007), "A Model of Consumer Learning for Service Quality and Usage," *Journal of Marketing Research*, 44 (November), 529–44.
- Judd, Kenneth L. (1998), *Numerical Methods in Economics*. Cambridge, MA: MIT Press.
- Lambrecht, Anja, Katja Seim, and Bernd Skiera (2007), "Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs," *Marketing Science*, 26 (5), 698–710.
- and Bernd Skiera (2006), "Paying Too Much and Being Happy About It: Existence, Causes, and Consequences of Tariff-Choice Biases," *Journal of Marketing Research*, 43 (May), 212–23.
- Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2003), "Projection Bias in Predicting Future Utility," *The Quarterly Journal of Economics*, 118 (4), 1209–1248.
- Lovett, Mitchell, William Boulding, and Richard Staelin (2009), "Consumer Learning Models for Perceived and Actual Product Instability," working paper, Simon Graduate School of Business, University of Rochester.
- Miller, Robert A. (1984), "Job Matching and Occupational Choice," *Journal of Political Economy*, 92 (6), 1086–1120.
- Miravete, Eugenio (1996), "Screening Consumers Through Alternative Pricing Mechanisms," *Journal of Regulatory Economics*, 9 (2), 111–32.
- (2002), "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans," *The Review of Economic Studies*, 69 (4), 943–71.
- (2003), "Choosing the Wrong Calling Plan? Ignorance and Learning," *American Economic Review*, 93 (1), 297–310.
- Narayanan, Sridhar, Pradeep Chintagunta, and Eugenio Miravete (2007), "The Role of Self Selection, Usage Uncertainty, and Learning in the Demand for Local Telephone Service," *Quantitative Marketing and Economics*, 5 (1), 1–34.
- Nelson, Philip (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (2), 311–29.
- Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033.
- (1994), "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, Vol. 4, R.F. Engle and D.L. McFadden, eds. Amsterdam: North-Holland, 3081–3143.
- Wellman, D. (1999), "Are We On?" *Supermarket Business*, 54 (12), 35.