Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock

By Michael D. Grubb and Matthew Osborne

Following FCC pressure to end bill shock, cellular carriers now alert customers when they exceed usage allowances. We estimate a model of plan choice, usage, and learning using a 2002–2004 panel of cellular bills. Accounting for firm price adjustment, we predict that implementing alerts in 2002–2004 would have lowered average annual consumer welfare by $33. We show that consumers are inattentive to past usage, meaning that bill-shock alerts are informative. Additionally, our estimates imply that consumers are overconfident, underestimating the variance of future calling. Overconfidence costs consumers $76 annually at 2002–2004 prices. Absent overconfidence, alerts would have little to no effect. (JEL D12, D18, L11, L96, L98)

Cellular phone companies frequently offer consumers contracts with included monthly allowances of voice minutes, text messages, and data usage that are followed by overage charges for higher usage. Consumers are often unaware that they are incurring overage charges during the month, which leads to bill shock at the end of the month. In an October 17, 2011 press release (CTIA 2011a), President Barack Obama declared:

Far too many Americans know what it’s like to open up their cell-phone bill and be shocked by hundreds or even thousands of dollars in unexpected fees and charges. But we can put an end to that with a simple step: an alert warning consumers that they’re about to hit their limit before fees and charges add up.

* Grubb: Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 (e-mail: michael.grubb@bc.edu); Osborne: Institute for Management and Innovation and Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON M5S 3E6, Canada (e-mail: matthew.osborne@rotman.utoronto.ca). A previous version circulated under the title “Cellular Service Demand: Tariff Choice, Usage Uncertainty, Biased Beliefs, and Learning.” We thank Parker Sheppard and Mengjie Ding for research assistance, and Katja Seim, Panle Jia, Eugenio Miravete, Catherine Tucker, Greg Lewis, Chris Knittel, Ron Goettler, Tavneet Suri, and S. Sriram for careful reading and feedback on early drafts. We also thank Ted O’Donoghue and seminar audiences at Duke, Cornell, University of Chicago, University of Rochester, Washington University of Maryland, University of Toronto, the FCC, Dartmouth, NBER, Yale, CMU, Georgetown, Boston University, UBC, UCLA, Berkeley, UCSD, UT Austin, Boston College, Johns Hopkins, and NYU for useful feedback. Finally we thank three anonymous referees for many helpful suggestions. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

† Go to http://dx.doi.org/10.1257/aer.20120283 to visit the article page for additional materials and author disclosure statement(s).

1 Only eight individuals (0.5 percent) in our sample incur a monthly bill in excess of $1,000 and these bills are due to roaming fees. Our study focuses on overage charges which are typically smaller but still often hundreds of dollars. In our sample, 19 percent of individuals incur an overage over $100, 8 percent incur an overage over $200, and 3 percent incur an overage over $300. Consumer surveys suggest that 34 percent of cell-phone users responsible for paying their own bill experience bill shock (GAO 2009) and 17 percent of all cell-phone users experience bill shock (Horrigan and Satterwhite 2010).
President Obama made this statement at the announcement of a new bill-shock agreement between the FCC and cellular firms. As of April 2013, this agreement commits cellular service providers to inform consumers when they approach and exceed their included voice, text, and data allowances (CTIA 2011a). Prior to the agreement, the FCC had proposed a similar regulation, which was strongly supported by consumer groups but opposed by the industry (Altschul, Guttman-McCabe, and Josef 2011; Deloney et al. 2011).2

Will the new bill-shock agreement help or hurt consumers? If firms held their prices fixed after implementing the agreement then it would weakly help consumers. This logic likely lies behind consumer groups’ strong advocacy for bill-shock alerts. However, the bill-shock agreement could hurt consumers once endogenous price changes are taken into account. Moreover, complementary theoretical work by Grubb (forthcoming) shows that the answer is theoretically ambiguous.3

To investigate the effect of bill-shock regulation, we develop and estimate a dynamic model of plan choice and voice usage that makes use of detailed cellular phone data from the years 2002–2004. While our 2002–2004 data are imperfect to directly resolve the policy question today, we use them to predict what effect the policy would have had during 2002–2004. Given our parameter estimates, counterfactual simulations predict the effect of bill-shock regulation had it been implemented in 2002. Holding prices fixed, our simulations predict an annual increase in consumer welfare of $103 per consumer and an accompanying annual reduction in industry profits of $196 per consumer. There is no reason, however, to expect firms to maintain the same prices. We predict that firms would respond to bill-shock regulation by reducing overage rates, reducing included minute allowances, and raising fixed fees. In response, 2 percent of consumers terminate service and more than 25 percent switch to more expensive plans. As a result, firms maintain annual profits close to unregulated levels (rising by just $7 per person). However, annual total welfare falls by $26 per person and annual consumer surplus falls by $33 per person (or 4 percent of the average annual bill).

The first prediction, that bill-shock regulation would benefit consumers but reduce profits if prices were unchanged, is explained by three stylized facts that we document in our data. First, a sharp increase in calling when free off-peak calling begins shows that consumers’ usage choices are price sensitive. Second, an absence of bunching at tariff kink points and other evidence discussed in Section II show that consumers are uncertain about the ex post marginal price when making calling choices. Third, novel evidence from call-level data suggests consumers in our sample are inattentive to their remaining balance of minutes. The latter two facts imply that bill-shock alerts are informative to consumers. The first fact implies that the information in bill-shock alerts is decision relevant for consumers. Together these three facts imply that consumers will respond to bill-shock alerts by reducing calling following an alert. The marginal cost of calls is small, so reduced calling lowers

---

2The wireless industry trade group, CTIA—The Wireless Association, argued that proposed bill-shock regulation “violates carriers’ First Amendment protections….against government compelled speech” (Altschul, Guttman-McCabe, and Josef 2011).

3Bill-shock alerts do not directly affect market power so their effect on profits is unclear. If profits change little then consumers benefit when social surplus increases. Thus whether consumer surplus rises or falls may depend on whether or not consumer choices become more closely aligned with firm costs, something that is unclear a priori.
social welfare. Thus the benefit to consumers of avoiding overage charges is less than the corresponding reduction in firm profits.

The second prediction—that bill-shock regulation would harm consumers rather than firms when endogenous price changes are taken into account—requires more explanation. As bill-shock regulation does not affect firms’ market power directly, we should expect that average monthly fees will rise to offset lost overage revenue so that there is little net effect on profits. This means that consumer surplus falls by approximately the same amount as social surplus. Social surplus is predicted to fall for two reasons. First, while several offsetting effects mean that average calling does not change, the average value of calls made falls. This is because those who remain on the same plans reduce calling in response to lower allowances and bill-shock alerts, while those who switch up to more expensive plans increase calling in response to lower marginal prices. Those cutting back forgo relatively high-value calls, while those calling more add relatively low-value calls, and average value falls. Second, although price changes are almost revenue-neutral, consumers perceive the net changes as a price increase because they undervalue bill-shock alerts (as explained below). Thus the outside option becomes a more competitive alternative and gains share, reducing social surplus.

The reason that consumers underestimated the value of bill-shock alerts stems from an additional stylized fact we document: some consumers in our sample make ex ante mistakes. We find that consumers in our sample systematically choose overly risky plans (those plans that yield high average bills and a chance of a very large bill given underlying uncertainty about usage). While this choice pattern could be due to risk-loving preferences, we assume that consumers are risk neutral, and hence infer that they underestimate the risk they face. In particular, we infer that consumers are overconfident because they underestimate the noise in their own forecasts about their future tastes for calling (by 62 percent at our estimates). Overconfident consumers undervalue the provision of bill-shock alerts because they underestimate the variance of their future consumption, and hence underestimate the likelihood of incurring an overage and benefiting from an alert.

Overconfidence explains an otherwise puzzling question: Why did cellular companies not unilaterally choose to provide bill-shock alerts? Our answer is that overconfident consumers underestimate their value. As a result, our counterfactual simulation predicts that a firm that introduced bill-shock alerts unilaterally would lose $165 per customer annually. A firm could not unilaterally raise monthly fees sufficiently to offset lost overage revenues without losing consumers to competitors. When regulation requires all firms to provide bill-shock alerts, profits are much less affected because average monthly fees can rise at all firms without affecting market shares. (Our prediction that industry profits actually rise slightly may seem surprising in light of the industry’s initial opposition to the regulation but may help explain their eventual voluntary agreement.)

\*According to a pricing manager at a top US cellular phone service provider, “people absolutely think they know how much they will use and it’s pretty surprising how wrong they are.”
Overconfidence proves to be important for another reason. Note that bill-shock alerts are only relevant if firms offer three-part tariffs, which Grubb (2009) shows are tailored to exploit overconfidence. Thus, absent overconfidence, we find that firms offer two-part tariffs rather than three-part tariffs, so that bill-shock regulation has no effect. As our predictions about bill-shock regulation are sensitive to estimated overconfidence, it is natural to ask whether learning could change our predictions by reducing the impact of overconfidence over time. This is a relevant question because the remaining stylized fact that we document is that consumers learn about their own tastes over time and switch plans in response. Importantly, our econometric model takes account of this behavior by assuming that consumers are Bayesian learners. Because we model learning, we account for the impact of learning on the introduction of bill-shock regulation at the beginning of our sample. Moreover, allowing for additional learning prior to conducting our counterfactual simulation of bill-shock regulation leads to slightly smaller but qualitatively similar predicted effects.

To make our prediction about bill-shock regulation’s effect on consumer welfare we make two additional contributions by advancing (i) demand modeling under marginal price uncertainty and (ii) estimation of overconfidence. Our first additional contribution is important because marginal-price uncertainty arises naturally whenever consumers make a series of small purchase choices that are aggregated and billed under a multipart tariff, as in cellular phone service, electricity, and health care. Addressing such marginal-price uncertainty represents a challenge for the literature which has typically side-stepped the issue by assuming that consumers can perfectly predict their future usage (Cardon and Hendel 2001; Reiss and White 2005; Lambrecht, Seim, and Skiera 2007), or that consumers believe they can perfectly predict their usage up to an implementation error which they ignore (Iyengar, Ansari, and Gupta 2007). (As discussed in Section I, Yao et al. 2012 and Jiang 2013 address marginal-price uncertainty using approaches markedly different to our own.)

We provide new evidence on how consumers make consumption choices under marginal-price uncertainty and estimate a tractable model incorporating such realistic behavior. In particular, we find evidence consistent with the student consumers in our sample being inattentive to their remaining minute balance. Given such inattention, we assume that consumers respond optimally to exogenously arising calling opportunities by choosing a calling threshold and making only those calls more valuable than the threshold. This approach has been proposed in earlier work, but has not been implemented in a structural model. Unlike standard models, our approach allows consumers to endogenously adjust their calling behavior in

---

5 A three-part tariff charges a fixed monthly fee for an included allowance of units followed by a constant marginal price for additional units.

6 We follow Erdem and Keane (1996); Ackerberg (2003); Crawford and Shum (2005); and Goettler and Clay (2011).

7 In the context of electricity demand, Borenstein (2009) proposes independently that consumers choose behavioral rules, such as setting the thermostat, similar to our calling threshold. Borenstein (2009) uses the behavioral rule assumption to motivate using expected marginal price rather than realized marginal price in reduced-form estimates of electricity price elasticities. Saez (2002) also suggests a very similar model for labor choice by income tax filers.
response to bill-shock alerts in counterfactual simulations. In contrast, attentive consumers would not find new information in a bill-shock alert.

Turning to our second additional contribution, we relax the standard rational expectations assumption by allowing for consumer overconfidence, or systematic underestimation of variance in tastes, and two potential mean biases, or systematic misestimations of average tastes. We are able to identify overconfidence from these mean biases separately due to the rich choice set of plans in our data that importantly include both three-part tariffs and a two-part tariff. Allowing for mean biases ensures that we do not misattribute mean bias to overconfidence.

To identify bias, we infer the true distribution of consumers’ tastes from their usage choices and infer consumers’ prior beliefs from their initial plan choices and subsequent switching decisions. The joint distribution of beliefs and tastes determines whether beliefs are biased in the sample population. For instance, suppose that we consider the subset of consumers who all share a particular prior belief about their own tastes. Absent an aggregate shock, rational expectations implies that this belief coincides with the distribution of tastes within this subset of the population. We relax this assumption, separately identify both beliefs and the distribution of tastes conditional on beliefs, and then compare the two distributions. We label differences between these distributions as biases.

Our results should be treated with caution when extrapolating to the policy being implemented today, for three reasons. First, we predict what effect the policy would have had during 2002–2004, which was a period when people used cellular phones to talk to each other. The current policy may have large effects on text messaging and data plans or roaming, all of which are absent from our study. However, an advantage of our 2002–2004 data is that the market is more tractable to model. Second, our sample consists entirely of university students who may not be wholly representative. Third, our supply model makes many simplifying assumptions, which are all caveats to the analysis. Nevertheless, at least three of our predictions should be generally robust. First, firms should respond to bill-shock regulation by lowering overage rates to encourage calling even after a bill-shock alert is received. Second, firms should adjust monthly fees and included allowances to increase revenue from monthly fees and offset lost overage fee revenue. Third, accounting for such endogenous price changes should lead to lower predicted firm losses and consumer benefits, relative to a prices-fixed analysis. Our paper is one of the first empirical projects showing that equilibrium effects can undermine the potential benefits of Thaler and Sunstein (2008)-style “nudges” designed to improve individual decision making.

Section I discusses related literature. Section II describes our data and documents five stylized facts that shape our modeling approach. Sections III and IV describe our model and identification in a simplified setting that does not distinguish between in-network and out-of-network calling and assumes a linear demand curve for calls. Sections V–VIII discuss estimation, present results, and conclude. Additional details

\footnote{Crawford and Shum (2005) and Ching (2010) estimate mean bias while Goettler and Clay (2011) identify underconfidence by a restriction that relates it to mean biases.}

\footnote{Importantly, the distribution of tastes is identified from usage after price sensitivity and beliefs are identified and can be used to map observed usage into underlying tastes. See Section IV.}
about the data, model, and estimation—including a description of how our complete model makes the distinction between in-network and out-of-network calling and implements a piecewise-linear demand curve—are in the online Appendix.

I. Related Literature

Complementary work by Jiang (2013) also evaluates the recent bill-shock agreement via counterfactual simulation, predicting a $1.98 billion annual welfare improvement. In contrast to our own approach, Jiang (2013) imposes rational expectations rather than estimating consumer beliefs and has cross-sectional data so cannot address learning. (A strength of Jiang’s 2013 data is that they are national and cover all firms.)

In our setting, consumers’ usage choices are complicated by the fact that marginal prices increase with usage. The standard approach to this problem assumes that consumers can forecast their usage perfectly, and so respond to the ex post marginal price (Cardon and Hendel 2001; Reiss and White 2005; Lambrecht, Seim, and Skiera 2007). A recent alternative relaxes perfect foresight but assumes consumers attentively track their usage from call to call (Yao et al. 2012). We show in Section IIB that neither perfect foresight nor attentive behavior fit our data. While all models are simplifications, these approaches implicitly assume no scope for bill-shock regulation and so seem inappropriate for our purposes. We assume consumers do not have perfect foresight and are inattentive, making and answering only those calls more valuable than a chosen threshold.

A final alternative assumes that consumers choose a target quantity that is implemented with an exogenous error (Iyengar, Ansari, and Gupta 2007; Jiang 2013). For comparison, our model of consumer choice can be reinterpreted as choice of a target quantity implemented with an endogenous error. A drawback of assuming that implementation errors are exogenous is that the resulting model does not predict how the errors will be affected by bill-shock alerts. Hence, Jiang’s (2013) bill-shock counterfactual is implemented by removing implementation error. In contrast, a strength of our approach is that consumers change calling behavior endogenously in response to information in bill-shock alerts.

In our model, overconfidence corresponds to overestimation of forecast precision. A significant body of experimental evidence shows that individuals are overconfident about the precision of their own predictions when making difficult forecasts (e.g., Lichtenstein, Fischhoff, and Phillips 1982). In other words, individuals tend to set overly narrow confidence intervals relative to their own confidence levels. A typical psychology study might pose the following question to a group of subjects: “What is the shortest distance between England and Australia?” Subjects would then be asked to give a set of confidence intervals centered on the median. A typical finding is that the true answer lies outside a subject’s 98 percent confidence interval about 30 to 40 percent of the time.

A small number of empirical papers relax rational expectations for consumer beliefs and estimate mean biases (Crawford and Shum 2005; Ching 2010; Goettler and Clay 2011). Most similar to our work is Goettler and Clay (2011), which estimates mean biases as well as underconfidence. Goettler and Clay (2011) identify underconfidence by a restriction that links it to one of two estimated mean biases. In
contrast, the rich tariff choice-set in our setting enables us to identify overconfidence separately from mean biases. An alternative approach taken by Hoffman and Burks (2014) is to use survey data on beliefs.

To identify beliefs from plan choices, we assume consumers are risk neutral. In contrast, related work on health insurance markets often does the reverse and imposes rational expectations to identify risk preferences from plan choices (Cardon and Hendel 2001; Einav et al. 2012; Handel 2013). Following a third approach, Ascarza, Lambrecht, and Vilkassim (2012) impose rational expectations and risk neutrality but estimate preferences for cellular phone usage that depend directly on whether contracts are two- or three-part tariffs.

Our results are consistent with a related sequence of papers about Kentucky’s 1986 local telephone tariff experiment (Miravete 2002, 2003, 2005; Narayanan, Chintagunta, and Miravete 2007; Miravete and Palacios-Huerta 2014). First, although the standard model of consumer choice does well at explaining behavior in the Kentucky experiment, our estimate that consumers underestimate their average taste for calling is consistent with evidence in Miravete (2003) which documents that on average all consumers who chose a small metered plan would have saved money on a larger flat-rate plan. Second, as in the Kentucky experiment we find that most consumers (65 percent) initially choose the tariff that turns out to be optimal ex post. Moreover, consumers switch plans and most switches appear to be in the right direction to lower bills (Section IIB). (This result is in contrast to Ater and Landsman’s 2013 finding that checking account customers who have paid overage fees switch toward checking plans that raise, rather than lower, their bills.)

II. Background: Data and Evidence for Stylized Facts

A. Data

Our primary data are a panel of individual monthly billing records for all student enrollees in cellular phone plans offered by a national cellular carrier in conjunction with a major university from February 2002 to June 2005. During this period, cellular phones were a relatively new product in the United States, having 49 percent penetration in 2002 compared to 98 percent in 2010. This dataset includes both monthly bill summaries and detailed call-level information for each subscriber. We also acquired Econ One data on the prices and characteristics of all cellular phone plans offered at the same dates in the vicinity of the university (Econ One Research, Inc. 2002–2004). The price menu offered to students differed from that offered by

---

10 We cannot separately identify consumers’ beliefs about the variance of their future tastes from their risk preferences over the resulting variation in their bills. When we observe consumers choose overly risky plans we infer that they underestimate the risk by underestimating the variance of their future tastes. In other words, we infer that they are overconfident. If we did not assume risk neutrality, however, we could not distinguish this explanation from the alternative that consumers are risk loving. If consumers are risk averse then stronger overconfidence is required to explain risky plan choices and our estimates of overconfidence are lower bounds on bias. See also footnote 14.

11 Interestingly, in Miravete (2003) the bias that can be inferred from elicited expectations differs from that inferred from choices. Consumers were not offered three-part tariffs in the Kentucky experiment so their choices do not shed light on overconfidence.

12 This feature makes our data ideal for studying consumer beliefs about new products. Penetration rates are calculated as estimated total connections (CTIA 2011b) divided by total population (US Census Bureau 2011).
the firm directly to the public: university plans included a two-part tariff, a limited three-month contractual commitment, different monthly promotions of bonus minutes, and a $5 per month surcharge on top of firm charges to cover the university’s administrative costs.

Four possible concerns about the external validity of our sample are: (i) students may not be paying their own bills, (ii) students may be more overconfident or make more mistakes than the general population, (iii) students may be less heterogeneous than the general population, and (iv) students may be more or less price sensitive than the general population. With respect to point (i), we note that students were sent individual bills to their campus residence; the students’ phone bills were not bundled with tuition. It is therefore likely that many students pay their own bills and are residual claimants on an allowance or graduate student stipend. As we show in Figure 2, students respond strongly to marginal prices. In response to point (ii), note that a pricing manager from one of the top US cellular phone service providers made the unsolicited comment that the empirical patterns of usage, overages, and ex post ‘‘mistakes’’ documented in Grubb (2009) using the same data were highly consistent with their own internal analysis of much larger and representative customer samples. In response to points (iii) and (iv), online Appendix G presents counterfactual simulations with additional heterogeneity and different price sensitivities that show our qualitative predictions are robust.

The bulk of our work makes use of the monthly billing data. For most analysis, we restrict attention to the period August 2002 to July 2004 and exclude individuals who began subscribing before August 2002. We focus on customer choice between four popular local plans that account for 89 percent of bills in our data. We group the remaining price plans with the outside option. These restrictions leave 1,357 subscribers used in our reduced-form analysis (from which we often exclude prorated bills). We estimate our structural model using 1,261 subscribers and 15,065 subscriber-month observations, as we exclude an additional 7 percent of individuals due to a variety of data problems (see online Appendix A).

Figure 1 shows the four popular plans, which we label as plans 0 through 3. Plan 0 is a two-part tariff that charges $14.99 per month and $0.11 per minute. Plans 1–3 are three-part tariffs that charge monthly fees ($M_j$) of $34.99, $44.99, and $54.99, respectively, include an allowance ($Q_j$) of 280 to 1,060 free peak-minutes, and charge an overage rate ($p_j$) of $0.35 to $0.45 per additional peak minute. Shares of plans 0–3 are 46, 27, 15, and 2 percent of bills, respectively. Plan prices are shown for Spring 2003 in Figure 1 and are described for all dates in online Appendix A Table 6.

All four plans include surcharges of $0.66 to $0.99 per minute for roaming outside a subscriber’s tri-state area and $0.20 per minute for long distance. Plans 1–3 always offer free off-peak calling but plan 0 does so only prior to fall 2003. Plan 0 includes free in-network calling, while plans 1–3 do not with the exception of plan 2 in 2004. Once a customer chooses a plan, the plan terms remain fixed for that customer,

---

13 As discussed in online Appendix G, we use our data to estimate a version of Jiang’s (2013) model adapted to our data. Comparison of the resulting estimates to those reported in Jiang (2013) that are based on national data suggest that students in our sample may indeed be more homogeneous and less price sensitive than the general population.
regardless of any future promotions or discounts, until they switch plans or terminate service. However, the terms of any given plan, such as the included allowances and overage rates for plans 1–3, vary according to the date a customer chooses the plan.

We say that one plan is larger than another if it coincides with the lower envelope of the tariff menu at a higher interval of usage. Plans are numbered in order of size, smallest to largest. Systematic consumer mistakes in choice of plan size identify mean biases. We say that one plan is riskier than another if it yields a higher expected bill for sufficiently high usage uncertainty. Loosely speaking, this orders plans by their degree of convexity. We also say that one plan is riskier than another if it gives a higher risk of a very large bill. Loosely speaking, this orders plans by their average steepness. Given the plans in our data, both notions of plan risk lead to the same ordering: Plan 0 is the safest plan, plan 1 is the riskiest, and plans 1–3 are numbered in order of decreasing risk. Consumer overconfidence is identified by the systematic choice of overly risky plans.\footnote{With a richer choice set we could separately identify a risk-loving preference (from the choice of overly steep plans) from overconfidence (from the choice of overly convex plans). In our data, however, the steepest plans are the most convex so we identify overconfidence by assuming risk neutrality.}

\section*{B. Evidence for Stylized Facts}

\textit{Three Stylized Facts Relevant to Modeling Usage Choices.}—Three features of the data are important to accurately model usage choices by customers of cellular phone service. First, consumers’ usage choices are price sensitive. Second, consumers’ usage choices are made while consumers are uncertain about the ex post marginal price. Third, consumers are inattentive to the remaining balance of included minutes during the course of a billing cycle. These three stylized facts motivate our assumption that, rather than choosing a precise quantity, consumers choose calling thresholds and proceed to make all calls valued above the threshold.

Consumer price sensitivity is clearly illustrated by a sharp increase in calling volume on weekday evenings exactly when the off-peak period for free night and weekend calling begins (Figure 2). This increase in calling is not simply a 9 PM effect, as the increase occurs only on weekdays, and at 8 PM for plans with early
nights-and-weekends. For plans with free weeknight calling starting at 8 PM, there is still a secondary increase in usage at 9 PM (panel C of Figure 2). Restricting attention to outgoing calls made to landlines (recipients for whom the cost of receiving calls was zero) almost eliminates this secondary peak (panel D of Figure 2). This finding suggests that the secondary peak is primarily due to calls to and from cellular numbers with the common 9 PM off-peak start time rather than a 9 PM effect.

Two pieces of evidence demonstrate consumer uncertainty about ex post marginal price. First, given clear sensitivity to marginal price, if consumers could anticipate whether they would be under their allowance (zero marginal price ex post) or over their allowance ($0.35 to $0.45 per minute marginal price ex post) we would expect to see substantial bunching of consumers consuming their entire allowance but no more or less. Figure 3 shows there is no bunching, which is consistent with similar findings in the contexts of electricity consumption (Borenstein 2009) and labor supply (Saez 2010). Second, consumers who anticipate being strictly under their allowance should exhibit no price response at the commencement of off-peak hours. However, Figure 4 shows that the sharp increase in calling at 9 PM shown in Figure 2 persists even in months for which the peak allowance is underutilized. This finding is true even for outgoing calls to landlines for which the jump in calling at 9 PM cannot be due to call recipients trying to avoid calling charges. These are natural consequences of usage choices made under uncertainty about ex post marginal price. Hence the standard model (Cardon and Hendel 2001; Reiss and White 2005), which assumes perfect consumer foresight, fits our data poorly.
Now we turn to evidence that consumers are inattentive. Figure 4 shows a sharp increase in weekday outgoing calls to landlines at 9 PM during months for which final usage is 65 percent or less of the included minute allowance. As already noted, the fact that a price response is observed when the ex post marginal price is zero before and after 9 PM is explained by ex ante uncertainty. At the time consumers make their calling choices they place positive probability on an overage and respond to a positive expected marginal price before 9 PM.
Evidence for inattention comes from comparing panels A and B in Figure 4. Panel A shows usage patterns during the first three weeks of the month and panel B shows usage patterns during the last week of the month. If consumers are attentive, some of their ex ante uncertainty about usage should be resolved by the final week of the month and it should be becoming increasingly clear that there will be no overage in the current month. Thus, if consumers were attentive, we would expect the price response to be diminished in panel B relative to panel A. In fact, usage patterns are remarkably similar in the two panels, consistent with consumer inattention.\footnote{15}

Online Appendix A provides additional evidence of inattention. In that analysis we note that an attentive consumer should cut back usage at the end of the month following high usage at the beginning of the month to adjust for the increased chance of paying overage fees (and vice versa). We look for such attentive behavior in a regression framework but find no evidence for it.

In contrast to our findings, Yao et al. (2012) reject our static calling threshold model in favor of attentive dynamic behavior using Chinese cellular phone data.\footnote{16} The discrepancy between Yao et al.’s (2012) finding and our own may be due in part to the fact that, unlike consumers in our data, the Chinese consumers could check their minute balance. Moreover, the financial incentives to pay attention were likely stronger for Chinese consumers than their American counterparts.

\textbf{Two Stylized Facts Relevant to Modeling Plan Choices.—} Two important features of the data are important to model plan choice by cellular customers accurately.\footnote{15} First, while 30 percent of contract choices are suboptimal ex post, consumers learn about their own usage levels over time and switch plans in response. Second (in the absence of aggregate shocks or risk-loving preferences) the pattern of ex post mistakes implies that some consumers make ex ante mistakes and is consistent with overconfidence.

The students in our sample could switch plans at any time and cancel after only three months, without any cost except hassle costs. Among the 1,357 customers in our data, 183 (14 percent) switch plans and 26 (2 percent) switch plans more than once, leading to a total of 221 plan switches. Of these switches, 85 (38 percent) are to plans that have either dropped in price or been newly introduced since the customer chose their existing plan. These switches could be motivated by price decreases rather than learning. However, the remaining 136 (62 percent) switches are to plans that are weakly more expensive than when the customer chose his or her existing plan. These switches must be due to learning or taste changes.

Not only do consumers switch plans, but they switch toward plans which save them money. To substantiate this claim, we make two calculations for each switch from an existing plan $j$ to an alternate plan $j'$, excluding those switches that could be explained by a price cut for plan $j'$. First, we calculate how much the customer would have saved had they signed up for the new plan $j'$ initially, holding their usage from the original plan $j$ fixed. This calculation provides a lower bound on the benefits of

\footnote{15}The finding is perhaps not surprising because service was resold by a university and, as a result, consumers could not contact the carrier to check minute balances.

\footnote{16}Yao et al. (2012) show that a scatterplot of cumulative weekly usage within a billing cycle against its lag is concave. In contrast, the relationship is linear in our data, which is consistent with our constant calling threshold.
switching to plan $j'$ because, by holding usage on the original plan $j$ fixed, it does not account for the additional benefit from optimizing calling choices for plan $j'$. Second, we calculate how much money the customer would have lost had they remained on existing plan $j$ rather than switching to the new plan $j'$, now holding usage from plan $j'$ fixed. This calculation provides an upper bound to the benefits of switching. It calculates the additional costs that would have been incurred on former plan $j$ given usage on the new plan $j'$, without accounting for the fact that some costs would be avoided by adjusting usage. We conclude that expected benefits from switching are between $10.87 and $24.56 per month and 60 to 69 percent of switches save money (see online Appendix A).

In unreported analysis, additional evidence of learning is that: (i) the likelihood of switching declines with tenure, and (ii) the likelihood of switching to a larger plan increases after an overage. Narayanan, Chintagunta, and Miravete (2007) estimate that consumers in the Kentucky experiment learn to switch up from overuse faster than they learn to switch down from underuse. In the context of retail banking, Ater and Landsman’s (2013) results suggest that the asymmetry could be large enough that banking customers’ tendency to choose overly large plans grows over time through switching. For simplicity, we implement symmetric learning in our structural model.

The presence of ex post mistakes alone shows only that consumers face uncertainty ex ante at the time of plan choice. The pattern of ex post mistakes reveals more, however. Assume that (i) consumers are risk neutral with standard preferences, (ii) there are no aggregate shocks correlated across consumers, and (iii) there are no ex ante mistakes. Then the following must hold: Altering plan choices using a rule that depends only on observables at the time of initial plan choice, while keeping observed usage constant, must weakly increase expected bills. Table 1 shows three violations of this prediction, in which average bills are reduced by moving everyone from one plan to another safer plan. Thus, in the absence of an aggregate shock or risk-loving preferences, some consumers make ex ante mistakes.

Turning to column 1 of Table 1 shows that, in the 2002–2003 academic year when plan 0 offered free off-peak calling, signing the 246 students who selected plans 1–3 up for plan 0 would save an average of $8.73 per affected bill. In the following year, the elimination of free off-peak calling on plan 0 made it a poor choice. However, column 2 shows that an alternative was to sign up the 437 students who chose plan 1 onto plan 2, which would have saved an average of at least $2.68 per affected bill.17 These two savings opportunities show consumers choosing plans that are overly risky. We believe risk-loving preferences are unreasonable in this setting and therefore conclude that there are either aggregate shocks or ex ante mistakes. While we cannot distinguish the two possibilities, the observed choice of overly risky plans is consistent with overconfidence.

Whether a consumer is uninformed about an aggregate shock or makes an ex ante mistake, the consequences for optimal firm pricing, consumer welfare, and policy counterfactual simulations are the same. More important is our assumption about predictability. In our counterfactual simulations we assume that firms anticipate

17 Note that the first savings opportunity is robust to dropping the top 30 percent of customers with the highest average savings, while the second savings opportunity is robust to dropping the top 2 percent of customers.
consumers’ choice patterns based on their knowledge of existing subscribers. In other words, if there is an aggregate shock we assume the firm observes it and knows consumers do not. Equivalently, if there are ex ante mistakes we assume they are predicted by the firm. Column 3 of Table 1 suggests that this assumption is reasonable by replicating the finding from column 2 using only data from the prior year. (The exercise is only suggestive due to the price change between the two periods.)

### III. Model

At each date $t$, consumer $i$ first receives a signal $s_{it}$ about her period $t$ taste shock $\theta_{it}$, next chooses a plan $j$ from a firm $f$, and finally chooses peak and off-peak quantities summarized by the vector $q_{it} = (q_{it}^{pk}, q_{it}^{op})$. (The text suppresses the distinction between in-network and out-of-network calling, which are covered in online Appendix C.) Total billable minutes for plan $j$ are

$$q_{ij}^{\text{billable}} = q^{pk}_{it} + O_{Pj} q^{op}_{it},$$

where $O_{Pj}$ is an indicator variable for whether plan $j$ charges for off-peak usage. At the end of period $t$, consumer $i$ is charged

$$P_j(q_{it}) = M_j + p_j \max\{0, q_{ij}^{\text{billable}} - Q_j\},$$

where pricing plan $j$ has monthly fee $M_j$, included allowance $Q_j$, and overage rate $p_j$. (A guide to these and other model parameters is provided in online Appendix B.)

We assume consumers are risk neutral, consumers have quasilinear utility, and peak and off-peak calls are neither substitutes nor complements. Consumer $i$’s money-metric utility in month $t$ from choosing plan $j$ and consuming $q_{it}$ units is

$$u_{ij} = \sum_{k \in \{pk, op\}} V(q_{it}^k, \theta_{it}^k) - P_j(q_{it}) + \eta_{itf},$$

where

$$V(q_{it}^k, \theta_{it}^k) = \frac{1}{\beta} q_{it}^k \left(1 - \frac{1}{2} \left(q_{it}^k/\theta_{it}^k\right)^\beta\right)$$

is the value from category $k \in \{pk, op\}$ calling, which depends on a pair of non-negative taste shocks $\theta_{it} = (\theta_{it}^{pk}, \theta_{it}^{op})$, and $\eta_{itf}$ is a firm-specific i.i.d. standard logit

### Table 1—Savings Opportunities

<table>
<thead>
<tr>
<th>Opportunity</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollment dates</td>
<td>10/02-8/03</td>
<td>9/03 onward</td>
<td>10/02-8/03</td>
</tr>
<tr>
<td>Enrollment change</td>
<td>plan 1-3 → plan 0</td>
<td>plan 1 → plan 2</td>
<td>plan 1 → plan 2</td>
</tr>
<tr>
<td>Affected customers</td>
<td>246 (34 percent)</td>
<td>437 (56 percent)</td>
<td>96 (14 percent)</td>
</tr>
<tr>
<td>Savings per affected bill</td>
<td>$8.73</td>
<td>$2.68</td>
<td>$5.45</td>
</tr>
</tbody>
</table>

Notes: Savings opportunities indicate that consumers choose overly risky plans (overconfidence) predictably. Savings estimates are a lower bound because we cannot always distinguish in- and out-of-network calls.
error that captures firm differentiation via network coverage and other firm characteristics. (The value function used for estimation is slightly richer than that in equation (2), leading to piecewise-linear rather than linear demand, as elaborated in online Appendix C.) The marginal value of a dollar is normalized to one. The coefficient on the logit error is assumed to be one but is calibrated for our bill-shock counterfactual simulations in Section VII. The price sensitivity parameter, $\beta$, determines how sensitive calling choices are to the marginal price of an additional minute of calling time. Our choice of functional form for $V(q_{it}^k, \theta_{it}^k)$ implies that the taste shock $\theta_{it}^k$ enters demand multiplicatively and can be interpreted as the minutes of category-$k$ calling opportunities that arise, as discussed below.

A. Quantity Choices

Recognizing that consumers are uncertain about the ex post marginal price when making usage choices from three-part tariffs is a key feature of our model and where we take a new approach (also suggested independently by Borenstein 2009). We assume that at the start of billing period $t$, consumer $i$ is uncertain about her period $t$ taste shock $\Theta_t$. She first receives a signal $s_{it}$ that is informative about $\Theta_t$, next chooses a plan $j$, and finally chooses a calling threshold vector $v_{itj} = (v_{itj}^{pk}, v_{itj}^{op})$ based on chosen plan terms and her beliefs about the distribution of $\Theta_t$. During the course of the month, the consumer is inattentive and does not track usage but simply makes all category-$k$ calls valued above $v_{itj}^k$. Over the course of the month, for $k \in \{pk, op\}$ this behavior cumulates to the choice

$$q_{it}^k = q(v_{itj}^k, \theta_{it}^k) = \theta_{it}^k \varphi(v_{itj}^k).$$

where $\varphi(v) = 1 - \beta v$ and $\varphi(0) = 1$ (see online Appendix B).

The interpretation is that $\theta_{it}^k$ is the volume of category-$k$ calling opportunities that arise and $\varphi(v)$ is the fraction of those calling opportunities worth more than $v$ per minute. Timing is summarized in Figure 5. Figure 6 shows the peak-calling threshold $v_{itj}^{pk}$ and resulting consumption choice $\theta_{it}^{pk} \varphi(v_{itj}^{pk})$ in relation to a consumer’s realized inverse demand curve for peak-calling minutes, $V_q(q_{itj}^{pk}, \theta_{it}^{pk})$.

Making all peak calls valued above the constant threshold $v_{itj}^{pk}$ is the optimal strategy of an inattentive consumer who does not track usage within the current billing cycle and hence cannot update his beliefs about the likelihood of an overage within the current billing cycle. (It is analogous to an electricity consumer setting a thermostat rather than choosing a quantity of kilowatt hours.)

When marginal price is constant, a consumer’s optimal calling threshold is simply equal to the marginal price. Thus for plan 0, which charges $0.11 per minute for all billable calls, $v_{itj}^{pk} = (0.11, 0.11OP_j)$. Further, $v_{itj}^{op} = 0$ for plans 1–3 because they offer free off-peak calling.

Conditional on choosing one of plans 1–3, which include free off-peak calling and an allowance of peak minutes, consumer $i$ chooses her period $t$ peak-calling threshold

---

18Leider and Şahin’s (2012) calling choice experiment finds that a majority of lab subjects use threshold rules when choosing which calls to make.
\[ v_{itj}^{pk} = p_j \Pr \left( \theta_{it}^{pk} \geq Q_j / \hat{q} \left( v_{itj}^{pk} \right) \mid \mathcal{S}_{it} \right) \frac{E \left[ \theta_{it}^{pk} \mid \theta_{it}^{pk} \geq Q_j / \hat{q} \left( v_{itj}^{pk} \right) \mid \mathcal{S}_{it} \right]}{E \left[ \theta_{it}^{pk} \mid \mathcal{S}_{it} \right]} . \]

The calling threshold \( v_{itj}^{pk} \) is between zero and the overage rate \( p_j \) and is increasing in the consumer’s belief about the mean and variance of calling opportunities, as both increase the anticipated likelihood of paying overage fees. Figure 9 in online Appendix B plots \( v_{itj}^{pk} \) as a function of beliefs.

Note that choosing threshold \( v_{itj}^{pk} \) is equivalent to choosing a target peak-calling quantity \( \hat{q}_{it}^{T} \equiv E \left[ \theta_{it}^{pk} \mid \mathcal{S}_{it} \right] \hat{q} \left( \hat{v}_{itj}^{pk} \right) \), which is implemented with endogenous error \( \left( \theta_{it}^{pk} - E \left[ \theta_{it}^{pk} \mid \mathcal{S}_{it} \right] \right) \hat{q} \left( \hat{v}_{itj}^{pk} \right) \). Importantly, consumers are aware of their inability to hit the target precisely and take this limitation into account when making their threshold/target choice.

B. Plan Choices

We model consumers’ choice between the four most popular pricing plans (plans 0–3), comparable AT&T, Cingular, and Verizon plans (Sprint offered no local plans), and an outside option which incorporates all other plans. We adopt Ching, Erdem, and Keane’s (2009) consideration set model by assuming that consumers
make an active choice with exogenous probability $P_C$ and keep their current plan with probability $(1 - P_C)$.\[^{19}\]

Customer $i$’s perceived expected utility from choosing plan $j$ at date $t$ is

$$U_{itj} = E\left[\sum_{k \in \{pk, op\}} V(q(v^k_{ij}, \theta^k_{it}), \theta^k_{it}) - P_j(q(v^*_{ij}, \theta_{it})) | \Im_{it}\right] + \eta_{itf},$$

and from choosing the outside option is $U_{it0} = O + \eta_{it0}$. Conditional on making an active choice, a consumer’s consideration set includes plans offered by her current provider, the outside option, and (to reduce computational time) plans from a randomly selected alternative firm. Consumers myopically\[^{20}\] choose the plan (or outside option) from their consideration set that maximizes expected utility in the current period.

C. Distribution of Tastes and Signals

We assume that the nonnegative taste shocks which determine usage are latent taste shocks censored at zero:

$$\theta^k_{it} = \begin{cases} 0 & \text{if } \tilde{\theta}^k_{it} < 0 \\ \tilde{\theta}^k_{it} & \text{if } \tilde{\theta}^k_{it} \geq 0 \end{cases}, k \in \{pk, op\}.$$

We assume that the latent shock $\tilde{\theta}^k_{it}$ is normally distributed and that consumers observe its value at the end of the billing period even when censored. This assumption adds additional unobserved heterogeneity to the model but preserves tractable Bayesian updating. Censoring makes zero usage a positive likelihood event, which is important since it occurs for 10 percent of plan 0 observations.

The latent taste shock satisfies

$$\tilde{\theta}_{it} = \mu_i + \epsilon_{it},$$

where $\mu_i = (\mu_{ipk}, \mu_{iop})$ is customer $i$’s true type and $\epsilon_{it} = (\epsilon_{itpk}, \epsilon_{itop})$ is a normally distributed mean-zero shock with variance-covariance matrix

\[^{19}\text{This assumption is equivalent to assuming switching costs are zero with probability } P_C \text{ and infinite otherwise.}\]

\[^{20}\text{We assume learning is independent of plan choice, so there is no value to experimentation with an alternative plan. Nevertheless, myopic plan choice is not optimal for several reasons. First, when a consumer is currently subscribed to a plan that is no longer offered (and is not dominated) there is option value to not switching, since switching plans will eliminate that plan from future choice sets. Second, if } P_C < 1, \text{ a forward-looking consumer would tend to discount her current period logit error } \eta_{itf} \text{ and signal } s_{it}. \text{ Third, if } P_C < 1, \text{ a forward-looking consumer should anticipate that her current plan choice may persist in the future but her future calling threshold choices } v^*_{itpk} \text{ will improve as she learns about her type } \mu_i. \text{ This consideration makes plans 1 and 2 marginally more attractive relative to plans 0 and 3 but the effect is not large. We ignore these issues for tractability. See Goettler and Clay (2011) for an investigation of the effect of switching costs on tariff choice when consumers are forward-looking.}\]
Together with the population parameter, consumers are learning about their peak type and identify beliefs about off-peak calling. For simplicity, we assume that while consumers make plan and calling threshold choices before learning the taste shock \( \theta_i \), however, prior to making these choices consumers observe a standard normal signal \( s_{it} \sim N(0, 1) \) that is jointly normal with the taste innovation \( \varepsilon_{it} \), with correlations \( \rho_{s, pk} = \text{corr}(s_{it}, \varepsilon_{it}^{pk}) \) and \( \rho_{s, op} = \text{corr}(s_{it}, \varepsilon_{it}^{op}) \). For technical convenience, we restrict \( \rho_{s, op} = \rho_s \rho_{s, pk} \). Conditional on the signal \( s_{it} \), the taste innovation \( \varepsilon_{it} \) follows a joint normal distribution with mean and variance given by Bayes rule in online estimations is independently of \( \varepsilon_{it} \).

Consumers’ true types, \( \mu_i \), are normally distributed in the population with mean \( \mu_0 = (\mu_0^{pk}, \mu_0^{op}) \) and variance-covariance matrix

\[
\Sigma_{\mu} = \begin{bmatrix}
(\sigma_{\mu}^{pk})^2 & \rho_{\mu} \sigma_{\mu}^{pk} \sigma_{\mu}^{op} \\
\rho_{\mu} \sigma_{\mu}^{pk} \sigma_{\mu}^{op} & (\sigma_{\mu}^{op})^2
\end{bmatrix}.
\]

Consumers learn about their own peak-type \( \mu_i^{pk} \) over time. At date \( t \), consumer \( i \) believes that \( \mu_i^{pk} \) is normally distributed with mean \( \tilde{\mu}_i^{pk} \) and variance \( \tilde{\sigma}_i^2 \); \( \mu_i^{pk} \mid \tilde{\mu}_i^{pk}, \tilde{\sigma}_i^2 \sim N(\tilde{\mu}_i^{pk}, \tilde{\sigma}_i^2) \). At date 1, prior beliefs are \( \mu_i^{pk} \sim N(\tilde{\mu}_i^{pk}, \tilde{\sigma}_i^2) \). Therefore, each new customer is characterized by the individual specific trivariate \( \{\tilde{\mu}_i^{pk}, \mu_i^{pk}, \mu_i^{op}\} \). Together with the population parameter \( \tilde{\sigma}_i^2 \), this triple specifies each customer’s true type \( \mu_i \) and prior belief.

The population is described by the joint distribution of \( \{\tilde{\mu}_i^{pk}, \mu_i^{pk}, \mu_i^{op}\} \), which we assume is a trivariate normal distribution: As described above, the marginal true type distribution is \( (\mu_i^{pk}, \mu_i^{op}) \sim N(\mu_0, \Sigma_\mu) \). The marginal distribution of initial point estimates is \( \tilde{\mu}_i^{pk} \sim N(\tilde{\mu}_i^{pk}, (\tilde{\sigma}_i^2)^2) \). Finally, correlations between point estimates and true types are \( \rho_{\mu, pk} = \text{corr}(\tilde{\mu}_i^{pk}, \mu_i^{pk}) \) and \( \rho_{\mu, op} = \text{corr}(\tilde{\mu}_i^{pk}, \mu_i^{op}) \).

\[\text{D. Prior Beliefs and Learning}\]

Estimation of consumer beliefs and learning is focused on a single dimension of usage: total peak calling. We make this restriction because plans 1–3 always offer free off-peak calling and hence the choice data are not rich enough to allow us to identify beliefs about off-peak calling. For simplicity, we assume that while consumers are learning about their peak type \( \mu_i^{pk} \) over time, there is no learning about off-peak demand because consumers know their off-peak types \( \mu_i^{op} \).

\[\text{21} \text{ The signal’s correlation with off-peak tastes, } \rho_{s, op}, \text{ is not well identified because it has little effect on plan choice due to unlimited off-peak calling on most plans. We restrict } \rho_{s, op} = \rho_s \rho_{s, pk} \text{ because it implies that } E[\varepsilon_{it} \mid x_i] \text{ is independent of } \varepsilon_{it}^{op}.\]

\[\text{22} \text{ This assumption does not affect our endogenous-price counterfactual simulations as we assume free off-peak calling.}\]
At the end of each billing period, peak usage $q_{it}^{pk}$ is realized and consumers can infer $\theta_{it}^{pk} = q_{it}^{pk} / \hat{q}(v_{it}^{pk})$. When $q_{it}^{pk} = \theta_{it}^{pk} = 0$, we assume that consumers can observe the latent taste shock $\theta_{it}^{pk}$ and consumers update beliefs according to Bayes rule, as given in online Appendix B.3. Over time consumers learn their own types: $\tilde{\mu}_{i,t+1}^{pk}$ converges to $\mu_{i}^{pk}$ and $\tilde{\sigma}_{i,t+1}^{pk}$ converges to zero.

Following a month with surprisingly high usage, a consumer increases his estimate of his type $(\tilde{\mu}_{i,t+1}^{pk} > \tilde{\mu}_{i,t}^{pk})$ and, hence, his future demand. In the standard model, the only behavior change that might result is a switch to a larger plan. In our model, a consumer might also switch to a larger plan but, conditional on not switching, would cut back on usage by choosing a higher calling threshold $(v_{i,t+1}^{pk} > v_{i,t}^{pk})$ and being more selective about calls.

E. Bias

Our model allows for biased consumer beliefs in two ways. First, we relax rational expectations restrictions typically placed on the joint distribution of point estimates $\tilde{\mu}_{i,t}^{pk}$ and true types $\mu_{i}^{pk}$. Second, we allow consumers to misperceive the variance of the signal $s_{it}$ and peak taste innovation $\epsilon_{it}^{pk}$. This flexibility in the model allows for overconfidence and two mean biases described below.

**Overconfidence.**—Overconfidence causes consumers to choose overly risky plans and underestimate the likelihood of paying overage charges. In our model, consumers are overconfident in two respects.

First, consumers are overconfident because the precision of their beliefs about their own types is miscalibrated. Let $\sigma_{i}$ denote the population standard deviation of true peak types conditional on a time $t$ point estimate: $\sigma_{i} = SD[\mu_{i}^{pk} | \tilde{\mu}_{i,t}^{pk}]$. (In period 1, $\sigma_{1} = \sigma_{\mu_{i}^{pk}}^{pk} \sqrt{1 - \rho_{\mu_{i}^{pk}}^{2}}$ and for later periods see online Appendix B.3.) In our notation, $\sigma_{i}$ is a consumer’s uncertainty about her peak type in period $t$. Let $\tilde{\sigma}_{1} = \delta \sigma_{1}$. Rational expectations implies that $\delta = 1$ so that consumer uncertainty is correctly calibrated in period 1: $\tilde{\sigma}_{1} = \sigma_{1}$. We do not impose this assumption. If $\delta < 1$, then consumers exhibit overconfidence: they overestimate the precision of their initial point estimates $\tilde{\mu}_{i,t}^{pk}$.

Second, consumers are overconfident because they underestimate the volatility in their own tastes. In particular, we assume that consumers underestimate the

---

23 In fact, given our assumption that consumers know $\mu_{i}^{op}$, consumers can also infer $\epsilon_{it}^{op}$ from off-peak usage which is informative about $\mu_{i}^{pk}$ because it is correlated with $\epsilon_{it}^{pk}$. We assume consumers only update beliefs using $\theta_{it}^{pk}$ and not $\epsilon_{it}^{op}$ for two reasons. First, consumers are unlikely to pay attention to off-peak usage when they are on a contract with free off-peak calls. Second, we only assume consumers know $\mu_{i}^{op}$ for simplicity as we cannot identify off-peak beliefs. In reality, consumers are unlikely to know $\mu_{i}^{op}$ so cannot actually infer $\epsilon_{it}^{op}$.

24 In an earlier version of the paper we distinguished this second form of overconfidence as volatility bias and estimated two distinct $\delta$ parameters. While conceptually the two parameters were separately identified by the rate of plan switching, in practice we found that the separation of the two was very sensitive to functional form assumptions. As a result, we now assume that both forms of overconfidence have equal severity and estimate only a single overconfidence parameter $\delta$. We chose this restriction because it implies that overconfidence does not disrupt learning: consumers appropriately weight their priors and new information when following Bayes rule. (See online
unconditional standard deviations of the signal \( s_{it} \) and peak taste innovation \( \varepsilon_{it}^{pk} \) by a factor \( \delta \). As a result, while consumers understand correctly the conditional mean 
\[
E[\varepsilon_{it}^{pk} | s_{it}] = \rho_{s, pk} \sigma_{\varepsilon}^{pk} s_{it},
\]
they underestimate the conditional standard deviation 
\[
\sigma_{\varepsilon}^{pk} = SD[\varepsilon_{it}^{pk} | s_{it}] \]
by a factor \( \delta \). As a result, while consumers understand correctly the conditional mean 
\[
E[\varepsilon_{it}^{pk} | s_{it}] = \rho_{s, pk} \sigma_{\varepsilon}^{pk} s_{it},
\]
they underestimate the conditional standard deviation 
\[
\sigma_{\varepsilon}^{pk} = SD[\varepsilon_{it}^{pk} | s_{it}] \]
by a factor \( \delta \), perceiving it to be 
\[
\tilde{\sigma}_{\varepsilon}^{pk} = \delta \sigma_{\varepsilon}^{pk}.
\]
(Bliefs about \( \varepsilon_{it}^{op} \) conditional on \( s_{it} \) are described in online Appendix B.3.) If \( \delta = 1 \), then consumers’ perceptions match reality. If \( \delta < 1 \), however, then consumers are overconfident and overestimate the precision of their forecasts of peak taste innovations \( \varepsilon_{it}^{pk} \). For tractability, we assume that consumers learn about means but not variances, so this bias is persistent.

Consumers’ plan choices and threshold choices depend on beliefs about the distribution of tastes \( \theta_{it} \). If \( \delta < 1 \) then consumers are overconfident and overestimate the precision of their forecasts about both their types \( \mu_{it}^{pk} \) and their taste innovations \( \varepsilon_{it}^{pk} \). For both reasons, overconfident consumers overestimate the precision of their forecasts about peak tastes \( \theta_{it}^{pk} \). When choosing a plan and a usage threshold at the beginning of period \( t \), consumers believe
\[
\tilde{\theta}_{it}^{pk} | \mathcal{Y}_{it} \sim N(\tilde{\mu}_{it}^{pk}, \tilde{\sigma}^{2}_{\theta_{it}}).
\]
In our notation,
\[
(6)
\tilde{\mu}_{it}^{pk} = \tilde{\mu}_{it}^{pk} + \rho_{s, pk} \sigma_{\varepsilon}^{pk} s_{it}
\]
is a consumer’s point estimate of her peak taste, \( \sigma_{\theta_{it}} \) measures the objective uncertainty surrounding it,
\[
(7) \quad \sigma_{\theta_{it}} \equiv SD[\theta_{it}^{pk} | \tilde{\mu}_{it}^{pk}] = \sqrt{\sigma^{2}_{\theta_{it}} + (1 - \rho^{2}_{s, pk})(\sigma_{\varepsilon}^{pk})^{2}},
\]
but an overconfident consumer’s uncertainty is lower: \( \tilde{\sigma}_{\theta_{it}} = \delta \sigma_{\theta_{it}} \). Off-peak beliefs are given in online Appendix B.3.

**Mean Biases.**—There are two mean biases in our model. Both arise because we relax typical restrictions on the population distribution of true types \( \mu_{it}^{pk} \) and initial point estimates \( \tilde{\mu}_{it}^{pk} \):
\[
\begin{bmatrix}
\mu_{it}^{pk} \\
\tilde{\mu}_{it}^{pk}
\end{bmatrix}
\sim N\left(\begin{bmatrix}
\mu_{0}^{pk} \\
\tilde{\mu}_{0}^{pk}
\end{bmatrix}, \begin{bmatrix}
(\sigma_{\mu}^{pk})^{2} & \rho_{\mu, pk} \sigma_{\varepsilon}^{pk} \tilde{\sigma}_{\mu}^{pk}
\rho_{\mu, pk} \sigma_{\varepsilon}^{pk} \tilde{\sigma}_{\mu}^{pk} & (\tilde{\sigma}_{\mu}^{pk})^{2}
\end{bmatrix}\right).
\]
First, absent an aggregate shock, rational expectations implies that an average individual’s initial point estimate is an unbiased estimate of her true type: \( \mu_{0}^{pk} = \tilde{\mu}_{0}^{pk} \). We relax this assumption and define \( b_{1} = \tilde{\mu}_{0}^{pk} - \mu_{0}^{pk} \). If \( b_{1} \neq 0 \), then there is aggregate mean bias and consumers will choose predictably plans which are too small \( (b_{1} < 0) \) or too large \( (b_{1} > 0) \).
Second, absent an aggregate shock, rational expectations implies that every individual (and not just the average individual) has an unbiased estimate of her true type. Letting

\[ b_2 \equiv 1 - \frac{\text{cov}(\mu_i^k, \hat{\mu}_{i1}^k)}{\text{var}(\hat{\mu}_{i1}^k)} = 1 - \frac{\rho_{\tilde{\mu}, pk} \phi_k}{\delta_{\mu}^k}, \]

it can be shown that

\[ \hat{\mu}_{i1}^k - E[\mu_i^k | \hat{\mu}_{i1}^k] = b_1 + b_2(\tilde{\mu}_{i1}^k - \hat{\mu}_{i1}^k). \]

Thus, initial point estimates \( \hat{\mu}_{i1}^k \) are unbiased if \( b_1 = b_2 = 0 \). We relax the restriction \( b_2 = 0 \) (equivalently \( \text{var}(\hat{\mu}_{i1}^k) = \text{cov}(\mu_i^k, \hat{\mu}_{i1}^k) \)). Instead, we make the weaker restriction that \( b_2 \in [0, 1 - \rho_{\tilde{\mu}, pk}^2] \) (equivalently \( \text{var}(\tilde{\mu}_{i1}^k) \geq \text{cov}(\mu_i^k, \tilde{\mu}_{i1}^k) \) and \( \text{var}(\mu_i^k) \geq \text{cov}(\mu_i^k, \tilde{\mu}_{i1}^k) \)). If \( b_2 > 0 \), then there is conditional mean bias and consumers overreact to their own private information, forming individual point estimates, \( \tilde{\mu}_{i1}^k \), that differ too much from the population average, \( \hat{\mu}_{i1}^k \). Conditional mean bias leads consumers to predictably choose plans that are too extreme. Nevertheless, as can be derived from Goettler and Clay’s (2011) model, the restriction \( b_2 \in [0, 1 - \rho_{\tilde{\mu}, pk}^2] \) implies that both aggregate and conditional mean biases are consistent with rational expectations if there is an unobserved aggregate shock to mean tastes \( (\mu_i^k)^{25} \).

Connection to Goettler and Clay (2011).—The joint distribution of true types \( (\mu_i^k) \) and point estimates \( (\hat{\mu}_{i1}^k) \) is a primitive of our model. Goettler and Clay (2011) show how this joint distribution can be derived from a common prior, an unobserved aggregate mean shock, and private signals. In our notation, this derivation endogenously imposes the restriction \( b_2 \in [0, 1 - \rho_{\tilde{\mu}, pk}^2] \). By imposing this restriction exogenously, we therefore ensure that both aggregate and conditional mean biases are consistent with rational expectations given an unobserved aggregate mean shock.\(^{26}\) Moreover it allows as special cases both rational expectations without aggregate shocks \((b_1 = b_2 = 0)\) and rational expectations with an aggregate shock drawn uniformly from the real line \((\text{Narayanan, Chintagunta, and Miravete’s 2007 assumption} b_1 = 0 \text{ and } b_2 = 1 - \rho_{\tilde{\mu}, pk}^2)\).

\(^{25}\) The upper bound \( b_2 \leq \rho_{\tilde{\mu}, pk}^2 \) does not affect our estimates but the lower bound \( b_2 \geq 0 \) resolves an identification problem. Absent the restriction, \( \delta_{\mu}^k \) and \( \rho_{\tilde{\mu}, pk} \) are not separately identified because there are two local maxima in the likelihood function with similar likelihoods, one with small \( \phi_k \) and large \( \rho_{\tilde{\mu}, pk} \) \((b_2 < 0)\) and another with large \( \phi_k \) and small \( \rho_{\tilde{\mu}, pk} \) \((b_2 > 0)\). We prefer the maxima selected by the \( b_2 \geq 0 \) restriction because its mean biases are consistent with rational expectations given an aggregate shock and the alternative provides poor out-of-sample fits when we use a holdout sample.

\(^{26}\) Note that overconfidence about type \( \mu_i^k \) cannot be rationalized by Goettler and Clay’s (2011) rational expectations model of an unobserved aggregate mean shock. Their model equation (6) also imposes the restriction (which we do not make) that \( \delta^2 = \text{cov}(\mu_i^k, \tilde{\mu}_{i1}^k)/(1 - \rho_{\tilde{\mu}, pk}^2)/\rho_{\tilde{\mu}, pk}^2 \). Given positive conditional mean bias \((b_2 > 0)\), this restriction implies underconfidence about type \( \mu_i^k \) by a factor \( \delta = (1 - b_2)^{-1/2} \geq 1 \).
IV. Identification

Parameters can be categorized into four groups: (i) the price sensitivity parameter $\beta$, (ii) parameters governing beliefs $\left(\tilde{\mu}^{pk}_t, \tilde{\sigma}^{pk}_t, \delta, \mu_\alpha^{pk}, \mu_\sigma^{pk}, \rho_{s, pk}, \rho_{s, op}, \Sigma_\mu, \Sigma_\sigma, \Sigma_v, \Sigma_r, \Sigma_s\right)$, (iii) the true distribution of tastes $\left(\mu_0, \rho_{\mu, pk}, \rho_{\mu, op}, \Sigma_\mu, \Sigma_\sigma\right)$, and (iv) parameters related to switching and quitting $\left(P_C, O\right)$. In Section IVA, we show that the price sensitivity parameter $\beta$ is identified from plan 0 usage independently of other parameters. In Section IVB, we show that beliefs about tastes $\theta_i$ are identified from plan choices conditional on $\beta$. We then argue that the distribution of $\theta_i$ is identified from usage once $\beta$ and beliefs about $\theta_c$ can be used to map observed usage into underlying tastes. The distribution of $\theta_i$ and beliefs about $\theta_c$ jointly determine overconfidence and mean biases. Finally, the rate of switching and the rate of quitting identify, respectively, the plan choice probability $P_C$ and the outside option $O$. Below we expand on the identification of price sensitivity and beliefs.

A. Price Sensitivity Parameter

If consumers’ chosen thresholds $(v^i_{ij})$ were known, the price sensitivity parameter $\beta$ could be inferred from marginal price variation and the induced variation in $q(v^i_{ij})$. Unfortunately, we require $\beta$ to calculate $v^i_{ij}$. We circumvent this problem by relying on a source of marginal price variation for which $v^i_{ij}$ is known. Prior to fall 2003, $v^i_{ij}$ is $0.11$ during peak hours and $0.00$ during off-peak hours for plan 0 subscribers.

We break out the share of calling demand for weekday outgoing calls to landlines immediately before and after 9 PM to help identify the price sensitivity parameter. The shock $r^{opm} = (r^{pk, 9}_{it}, r^{op, 9}_{it}) \in [0, 1]^2$ captures the share of peak and off-peak calling demand that is within 60 minutes of 9 PM on a weekday and is for an outgoing call to a landline. The distribution of $r^{k, 9}_{it}$ for $k \in \{pk, op\}$ is a censored normal, with latent shock $r^{k, 9}_{it} = \alpha^{k, 9}_i + e^{k, 9}_{it}$ and

$$r^{k, 9}_{it} = \begin{cases} 0 & \text{if } \tilde{r}^{k, 9}_{it} \leq 0 \\ \tilde{r}^{k, 9}_{it} & \text{if } 0 < \tilde{r}^{k, 9}_{it} < 1 \\ 1 & \text{if } \tilde{r}^{k, 9}_{it} \geq 1 \end{cases}$$

where $\alpha^{k, 9}_i$ is unobserved heterogeneity and $e^{k, 9}_{it}$ is a mean-zero shock normally distributed with variance $(\sigma^{k, 9}_v)^2$ independent across $i$, $t$, and $k$. We assume that $\alpha^{pk, 9}_i$ is normally distributed in the population with mean $\mu^{pk, 9}_\alpha$ and variance $(\sigma^{pk, 9}_\sigma)^2$.

Our identifying assumption 27 for the price sensitivity parameter is that consumer $i$’s expected outgoing calling demand to landlines on weekdays is the same between 8:00 PM and 9:00 PM as it is between 9:00 PM and 10:00 PM:

$$E[r^{pk, 9}_{it}] E[\theta^{pk}_it] = E[r^{op, 9}_{it}] E[\theta^{op}_it].$$

27 Importantly, our model also assumes that peak and off-peak calls are not substitutes. In reality consumers do delay calls from peak hours to off-peak. Thus the price response measured at 9 PM may overestimate consumers’ overall sensitivity to the price of peak calling (holding off-peak prices fixed). Note, however, that other moments in the data lead to a more conservative estimate of $\beta$ than would be implied from the 9 PM usage jump alone.
In other words, we assume that the increase in observed calling to landlines on weekdays immediately after off-peak begins at 9 PM is a price effect rather than a discontinuous increase in demand at 9 PM.\(^8\) As a result, equation (9) implicitly defines \(\alpha_{i,9}^{op}\) as a function of \(\alpha_{i,9}^{pk}\) and other parameters.

Given plan 0 pricing prior to fall 2003, \(\theta_{i,t}^{op} = q_{i,t}^{op}\) and \(\theta_{i,t}^{pk} = q_{i,t}^{pk}/(1 - 0.11\beta)\). Moreover, the pre- and post-9 PM calling shares are always observed because calling thresholds are constant within peak and within off-peak hours: \(r_{it,9}^{op} = q_{it,9}^{op}/q_{it}^{op}\) and \(r_{it,9}^{pk} = q_{it,9}^{pk}/q_{it}^{pk}\). Thus equation (9) can be solved for \(\beta\) as a function of moments of the data:

\[
\beta = \frac{100}{11} \left(1 - \frac{E[q_{it,9}^{op}/q_{it}^{op}] E[q_{it}^{pk}]}{E[q_{it,9}^{op}/q_{it}^{op}] E[q_{it}^{op}]}\right).
\]

**B. Beliefs**

Next, consider identification of consumers’ prior beliefs from plan choices. Choice data are quite informative about beliefs about peak usage, as explained below using Figure 7, but relatively uninformative about beliefs about off-peak usage. Hence we assume consumers know their own off-peak taste distribution (including \(\mu_{i,op}^{\prime}\) and \(\sigma_{i,op}^{\prime}\)).

Prior to fall 2003, when off-peak calling is free on all plans, an individual consumer’s initial plan choice depends only on \(\beta\), which is already identified, and her beliefs about \(\theta_{i,t}^{pk}\), which are described by mean \(\bar{\mu}_{i,1}^{pk}\) and variance \(\bar{\sigma}_{i,1}^{2}\), as defined in equations (6)–(7). Thus initial plan choice shares prior to fall of 2003 are sufficient to identify \(\bar{\sigma}_{i,1}\) and the population distribution of \(\bar{\mu}_{i,1}^{pk}\). Initial choice shares in post-fall 2003 data also aid identification, but require a more complicated argument involving beliefs about off-peak tastes.

Initial plan choices place bounds on each individual’s prior beliefs about the mean \((\bar{\mu}_{i,1}^{pk})\) and variance \((\bar{\sigma}_{i,1}^{2})\) of their first taste shock, \(\bar{\theta}_{i,1}^{pk}\). Using October-November 2002 pricing data and ignoring free in-network calling, the top panel of Figure 7 shows plan-choice as a function of prior beliefs \(\{\bar{\mu}_{i,1}^{pk}, \bar{\sigma}_{i,1}^{2}\}\) given \(\beta = 2\). Consumers joining in October–November 2002 choose the plan corresponding to the shaded region within which their beliefs lie. Figure 7 shows that plan 0 is chosen both by individuals with low expectations of usage (low \(\bar{\mu}_{i,1}^{pk}\)), as it has the lowest fixed fee, and by individuals with high uncertainty about usage (high \(\bar{\sigma}_{i,1}\)), as it never charges more than $0.11 per minute and is therefore a safe option. Plan 1 is only chosen if \(\bar{\sigma}_{i,1}\) is smaller than 107, an upper bound for \(\bar{\sigma}_{i,1}\) implied by plan 1’s observed positive share.

If we were to fix \(\bar{\sigma}_{i,1}\) at any level below 107, individual \(i\)’s plan choice bounds \(\bar{\mu}_{i,1}^{pk}\) to an interval. For example, the bounds are given for \(\bar{\sigma}_{i,1} = 80\) by the vertical

---

\(^8\) We focus on calls to landlines because the other party to the call pays nothing both before and after 9 PM. The assumption would be unreasonable for calls to or from cellular numbers since such calling opportunities increase at 9 PM when the calls become cheaper for the other party and the other party is more likely to call or answer.
lines in Figure 7. Combining plan share data from customers who join in October–November 2002 with these bounds generates a histogram over $\tilde{\mu}_{\theta_1}$ with four bins, one for each of the four pricing plans. Since we assume that $\tilde{\mu}_{\theta_1}$ is normally distributed, this histogram would then identify the distribution. The resulting histogram and fitted normal distribution are both shown in the lower panel of Figure 7 for the case $\tilde{\sigma}_{\theta_1} = 80$ and October–November 2002 new subscriber plan choice shares of 69 percent, 10 percent, 19 percent, and 2 percent for plans 0 to 3, respectively.

The exercise described above identifies consumer beliefs about $\theta_{it}^{pk}$, including uncertainty about initial tastes ($\tilde{\sigma}_{\theta_1}$) and the population mean ($E[\tilde{\mu}_{\theta_0}^{pk}]$) and variance ($\text{var}[\tilde{\mu}_{\theta_0}^{pk}]$) of initial point estimates of $\theta_{it}^{pk}$. In other words, we have identified the distribution of the population over the space of Figure 7 (and equivalently Figure 9 in online Appendix B). As calling thresholds are a function of beliefs, we have thus also identified the distribution of peak calling thresholds $v_{ij}^{pk}$ in the population. Next,
given \( \beta \), the correlation between plan choice and usage identifies the correlation between \( v_{ik}^{pk} \) and usage. This gives us enough information to infer the distribution of tastes \( \theta_{it}^{pk} \) from the distribution of usage \( q_{it}^{pk} \).

Given beliefs about \( \theta_{it}^{pk} \) and its true distribution, we can identify the population mean \( (\hat{\mu}_0^{pk}) \) and variance \( ((\hat{\sigma}_{\mu}^{pk})^2) \) of \( \mu_{i1}^{pk} \) as well as overconfidence \( (\delta) \). Following equations (6)–(7), \( \mu_{i1}^{pk} \) is equal to \( E[\mu_{i1}^{pk}] \), \( (\hat{\sigma}_{\mu}^{pk})^2 \) is equal to \( \text{var}[\mu_{i1}^{pk}] - (\rho_{s,pk} \sigma_{\epsilon}^{pk})^2 \), and \( \delta \) is equal to \( \hat{\sigma}_{\theta_1} / \sigma_{\theta_1} \), where

\[
(11) \quad \sigma_{\theta_1} = \sqrt{((\sigma_{\epsilon}^{pk})^2 - \rho_{s,pk}^2)} \sqrt{1 - \rho_{s,pk}^2} \sqrt{1 - \rho_{s,pk}^2} \sqrt{(\sigma_{\epsilon}^{pk})^2}.
\]

Thus, for identification of \( \tilde{\mu}_0^{pk}, \tilde{\sigma}_{\mu}^{pk}, \) and \( \delta \), it remains to show that \( \sigma_{\epsilon}^{pk}, \sigma_{\mu}^{pk}, \rho_{s,pk}, \) and \( \rho_{s,pk} \) are identified. Now that we have enough information to infer the distribution of \( \Theta_{it} \) from observed usage, the first two parameters are straightforward. The within- and between-individual variances of \( \theta_{it}^{pk} \) correspond to \( (\sigma_{\epsilon}^{pk})^2 \) and \( (\sigma_{\mu}^{pk})^2 \), respectively. Next, \( \rho_{s,pk} \) is identified from the correlation between initial plan choices and long-run usage. We cover \( \rho_{s,pk} \) next.

The signal informativeness, measured by correlation \( \rho_{s,pk} \), is identified by the fraction of switches which are inconsistent with learning based only on past usage. Without a signal in the model \( (\rho_{s,pk} = 0) \), any initial plan choice could be rationalized by an appropriate prior belief \( \tilde{\mu}_{i1}^{pk} \). However, the model requires private signals \( (\rho_{s,pk} > 0) \) to rationalize switches that appear to be in the “wrong” direction given past usage. For example, suppose a customer with high average usage chooses a small plan and subsequently experiences a string of overage charges. A low prior belief \( (\tilde{\mu}_{i1}^{pk} \) small) could rationalize the initial choice of a small plan. However, given the assumption of Bayesian learning, no prior can simultaneously rationalize the initial choice and a subsequent switch to an even smaller plan unless consumer beliefs are informed by more than past usage. Given \( \rho_{s,pk} \), the population mean and variance of initial point estimates \( \tilde{\mu}_{i1}^{pk} \) and overconfidence are all identified.

V. Estimation Procedure

This section summarizes the construction of the likelihood function and the estimation procedure; complete details are in online Appendix D.

An observation in our likelihood is a vector consisting of a consumer’s plan choice and usage in each relevant category (peak, off-peak, 8–9 PM, 9–10 PM, in-network, and out-of-network) for a given month. The likelihood of an individual’s sequence of plan and usage choices has three groups of unobservables that must be integrated out: individual specific unobserved heterogeneity, including \( \mu_{i1}^{pk}, \mu_{i1}^{op} \), and \( \tilde{\mu}_{i1}^{pk} \), private signals \( s_{ij} \), and i.i.d. idiosyncratic taste shocks including \( \epsilon_{it} \) and \( \eta_{it} \).

As the likelihood function does not have a closed-form expression, we turn to Maximum Simulated Likelihood (Gourieroux and Monfort 1993). To form the likelihood, we integrate out some unobservables using Monte Carlo simulation, using 400 shuffled draws from a Sobol quasi-random number generator. In particular, we simulate out individual specific effects, private signals, and taste shocks \( \epsilon_{it}^{j} \) when usage in category \( k \in \{pk, op\} \) is zero. Given a vector of draws, an individual’s
likelihood can be computed in closed-form using the densities of the remaining structural errors, including \( \varepsilon_{it} \) and \( \eta_{it} \). To ensure that our likelihood function is smooth in the parameters, we draw \( s_{it} \) (and \( \varepsilon_{it}^k \) when \( \tilde{\theta}_{it}^k \) is censored) by importance sampling, as explained in online Appendix D.

Computational difficulties in estimation have three sources: First, the model’s high dimensional unobserved heterogeneity requires many evaluations of the likelihood function. Second, the computation of \( v_{itj}^{pk} \) requires a nonlinear equation solver to solve equation (4) numerically. We must do this computation for each simulation draw, at each time period, for every individual, at every choice that is not the outside good or the two-part tariff, plan 0. Third, the importance sampling method we employ to draw \( s_{it} \) requires that the drawn \( s_{it} \) rationalize observed choices. This requirement means that for every choice (other than quitting) we must compute the bounds on the set of \( s_{it} \) values for which the utility of the chosen plan exceeds the utility of other university plans. For example, if consumer \( i \) chooses plan 0, typically the signal \( s_{it} \) will be bounded above.

VI. Results

A. Parameter Estimates

We estimate 28 model parameters in total. The estimates and standard errors for the 21 parameters discussed in the main text are shown in Table 2. Estimates of the seven parameters related to in-network calling are presented in Table 13 of online Appendix C. Turning to Table 2, the calling price coefficient \( \beta \) is 2.7, which indicates that a price increase from $0.00 to $0.11 per minute decreases usage by 30 percent.

The next nine parameters characterize the joint normal distribution of consumers’ initial point estimates, \( \tilde{\mu}_{it}^{pk} \), and true types, \( \mu_{it}^{pk} \) and \( \mu_{it}^{op} \). The average consumer’s point estimate, \( \tilde{\mu}_0^{pk} \), is estimated to be 216 minutes, while the average consumers’ true peak type, \( \mu_0^{pk} \), is 272 minutes. (Accounting for censoring of the latent shock, the average consumer believes the mean of \( \tilde{\theta}_{it}^{pk} \) is 217 minutes and the true mean is 294 minutes.) The average off-peak type, \( \mu_0^{op} \), is larger than the peak value at 419 minutes. (The model predicts an even larger gap between peak and off-peak usage due to consumer price sensitivity.)

The population standard deviations of \( \tilde{\mu}_{it}^{pk} \), \( \mu_{it}^{pk} \), and \( \mu_{it}^{op} \) (\( \tilde{\sigma}_{it}^{pk} \), \( \sigma_{it}^{pk} \), and \( \sigma_{it}^{op} \)) are 210, 148, and 421 minutes, respectively. The estimates of \( \rho_{\tilde{\mu}, pk} \) and \( \rho_{\tilde{\mu}, op} \) indicate that initial beliefs about peak type, \( \tilde{\mu}_{it}^{pk} \), are weakly positively correlated with the true types \( \mu_{it}^{pk} \) and \( \mu_{it}^{op} \). The correlation between the peak and off-peak type \( \mu_{it}^{pk} \) is 0.62.

The last row of column 1 and the first three rows of column 2 in Table 2 describe the distribution of the signal \( s_{it} \) and the error term \( \varepsilon_{it} \). The standard deviations of the error terms, \( \sigma_{it}^{pk} \) and \( \sigma_{it}^{op} \), are 248 and 325 minutes for peak and off-peak usage, respectively, and their correlation, \( \rho_{\varepsilon} \), is 0.52. Notice that the variance of the peak usage error is higher than the variance of \( \mu_{it}^{pk} \), indicating that more of the variation

\[29\] We compute standard errors using the subsampling method of Politis and Romano (1994). We first draw 250 datasets, each consisting of a 40 percent subsample of individuals chosen randomly without replacement. We estimate our model on each dataset, and compute the standard error as the standard deviation of the estimates across datasets, multiplied by a scaling factor.
in peak usage can be attributed to monthly volatility than the consumer-level fixed effect. The correlation between the signal and the peak error, $\rho_{s, pk}$, is 0.41, indicating that a substantial portion of the upcoming month’s taste shock is predictable by the consumer.

Continuing down column 2 in Table 2, our estimate of $\delta$ is 0.38, which is consistent with strong consumer overconfidence. The next four parameters of column 2 describe consumers’ tastes for 8:00 PM to 10:00 PM usage. The low value of $\mu_{pk}$ indicates that outgoing 8:00 to 9:00 PM landline usage is small as a fraction of total peak usage, which is consistent with the data. Parameter $\sigma_{pk}$ captures individual specific heterogeneity in the fraction of peak calling falling between 8:00 PM and 9:00 PM, while $\sigma_{e, pk}$ and $\sigma_{e, op}$ capture idiosyncratic volatility.

Shown next is the plan consideration parameter, $P_C$, which we estimate to be 0.06, indicating that consumers seldom look at prices. Below that is the outside good utility, $O$, estimated to be 1.16 (although this estimate is imprecise). Compared to average utilities of about 90, this low estimate of the outside good utility implies that consumers prefer inside goods to the outside good by a large margin and rationalizes the low quit rate observed in the data.

We report the fit of our estimated model to the data in Table 14 and Figure 12 in online Appendix D. Our model does a good job of fitting plan choice shares and the rate of plan switching. The model generally matches observed usage moments but has some difficulty fitting the exact shape of the usage density, which is highest near zero and smoothly drops as usage increases. Our censored normal specification produces a hump near zero that is not replicated in the data.
B. Biases and Learning

Returning to consumer beliefs, we display estimates of bias parameters in Table 3, including those that are functions of our estimated parameters. Our estimate of $\delta$, 0.38, is consistent with strong overconfidence. Within our model, this estimate implies that consumers underestimate the noise in their own forecasts about their future tastes for calling by 62 percent. In particular, the standard deviation of consumers’ initial uncertainty about $\tilde{\theta}_i^{pk}$, $\delta \sigma_{1}$, is 103 minutes, 62 percent less than the correctly calibrated $\sigma_{1} = 269$ minutes. This bias leads consumers to systematically choose overly risky plans such as plan 1. (Note that if consumers are risk averse rather than risk neutral then this estimate is a lower bound on the magnitude of overconfidence.)

Our model attributes overconfidence equally to two factors. First, the standard deviation of consumers’ initial uncertainty about their type, $\delta \sigma_{1}$, is 56 minutes, 62 percent less than the correctly calibrated $\sigma_{1} = 146$ minutes. Second, consumers perceive the standard deviation of monthly volatility in peak usage to be $\delta \sigma_{pk}^{\epsilon}$ = 95 minutes, which is 62 percent less than the correctly calibrated $\sigma_{pk}^{\epsilon} = 248$ minutes.

Next, $b_1 = -55$, consistent with negative aggregate mean bias. Within our model, this estimate implies that the average consumer underestimates her peak type $\mu^{pk}_{i}$ by 55 minutes and will systematically choose too small a plan. Finally, the positive estimate of $b_2$ is consistent with strong conditional mean bias. Within our model, this estimate implies that consumers choose plans which are too extreme and will moderate their plan choices over time.

Our model accounts for learning, which means that mean biases dissipate over time. Our simulation predicts that after one year, aggregate mean bias diminishes 40 percent ($b_1$ increases from $-55$ to $-33$) and conditional mean bias diminishes 33 percent ($b_2$ decreases from 0.86 to 0.58).

C. Fixed-Price Counterfactual: Impact of Biased Beliefs

Before proceeding to simulate endogenous price changes in Section VII, we briefly simulate the change in annual firm profits, consumer welfare, and total welfare that results from debiasing consumers while holding observed prices fixed. We construct these counterfactual simulations at our data in the sense that we hold fixed the number of consumers, and when consumers enter and exit the dataset. Annual surplus changes are measured in dollars per student averaged over the two-year period.

31In the context of grocery home delivery service, Goettler and Clay (2011) also find $b_1 < 0$ and $b_2 > 0$.
period that they are observed. We assume marginal cost is $0.02 per minute based on a calibration described in Section VII. The first three columns of Table 4 show the welfare effects when students face university prices, while the last three columns show the welfare effects when consumers face publicly available prices.

The first row of Table 4 shows profits, consumer surplus, and social welfare at our estimated parameters. The second two rows of Table 4 show the impact of removing biases: row 2 shows the impact of removing overconfidence and row 3 shows the impact of removing all biases. In all cases, debiasing consumers raises consumer surplus and lowers firm profits because debiased consumers make better choices and pay fewer overage charges. Given public prices, consumer surplus increases by $76 when overconfidence is removed and by $91 when all biases are removed. Firm profits fall by a similar magnitude and hence changes in total welfare are relatively small.

### VII. Endogenous-Price Counterfactual: Bill-Shock Regulation

#### A. Endogenous Price Calibration

To calculate endogenous equilibrium prices we introduce a simple supply model:32

We assume that there are three symmetric firms, equilibrium is symmetric static Nash in prices, overage rates are at most $0.50 (see Section VIIC), and each firm offers a menu of three plans. We assume that firms pay a monthly fixed cost $FC$ of serving a customer account, a per-minute marginal cost $c$ to carry peak calls, and no cost for off-peak calls when capacity constraints are slack.

To simulate equilibrium prices it is important to accurately capture the degree of firm market power and firm costs in the industry. In our model, logit errors give consumers idiosyncratic firm preferences (which might result from differences in network coverage or phone availability) that create market power. The degree of market power is governed by the logit error variance, which is normalized to one in our estimation. Now, however, we adjust the model by revising equations (1) and (5) to weight the logit error ($\eta_{itf}$) in consumers’ utility functions by the factor $1/\lambda$.

---

32 Our endogenous price counterfactual simulations relate to the literature on optimal contracting with nonstandard consumers, for which Spiegler (2011) provides a useful guide. Of particular relevance are DellaVigna and Malmendier (2004); Eliaz and Spiegler (2006, 2008); and Grubb (2009, forthcoming). Online Appendix E explains how we compute equilibrium prices.
We calibrate $\lambda$ and firm marginal costs $c$ using supply-side price data: We select the values of $\lambda$ and $c$ that best rationalize observed prices of the major firms conditional on our demand estimates. Our algorithm, which is described in online Appendix F, calibrates $\lambda$ to be 0.03 and peak marginal cost $c$ to be $0.02$ per minute.\(^{33}\) Finally, we assume that monthly fixed costs are $FC = $15 per customer based on industry financial reports.\(^{34}\)

In an unreported specification, we normalized the outside good utility to zero and estimated $\lambda$. The estimate of $\lambda$ is 0.06 compared to our calibrated value of 0.03. We use the value calibrated from observed prices because the estimated value is not well identified by our demand data and implies too little market power to explain observed prices. (Given that only the quitting rate is available to identify the outside good utility and $\lambda$, outside price variation is too limited to identify separately the two parameters.)

Before proceeding, we make two additional comments on our calibration approach. First, one potential problem is that our demand estimates were made conditional on $\lambda = 1$, but different values of $\lambda$ might produce different demand estimates. Fortunately, our demand estimates are relatively insensitive to $\lambda$, which we discuss in online Appendix F. Second, in principle we could have estimated $\lambda$ and the other parameters jointly by using constrained maximum likelihood and constraining $\lambda$ to minimize the calibration objective function. We avoided this approach for two reasons. First, it is computationally infeasible to nest calibration within each likelihood evaluation as each calibration takes a day. Second, we prefer to impose our supply-side structural assumptions (that competition is symmetric static Nash in prices and that our student population is representative) only when they are necessary in the endogenous-price counterfactual simulations.

### B. Modeling Bill-Shock Alerts

In our bill-shock regulation counterfactual, consumers are informed when their usage reaches $Q_j$, their allotment of free minutes. (Alerts are not applicable to two-part tariffs with constant marginal prices.) In response to this new policy, a consumer’s usage rule changes: A consumer will accept all calls valued above $v_{ij}^{pk}$ until she exhausts her included minutes. After that point, she accepts only calls valued above $p_j$. Because the consumer adjusts her calling threshold upon making $Q_j$ calls, the optimal initial threshold $v_{ij}^{pk}$ is lower than that characterized by equation (4). The reduction in $v_{ij}^{pk}$ is never more than about $0.025$. The threshold falls slightly because the consumer can afford to make more calls prior to an alert, using up more included minutes in a low-demand month, safe in the knowledge that a bill-shock alert will protect her from a large bill in a high-demand month. Online Appendix B.2 describes expected utility and characterizes $v_{ij}^{pk}$ under bill-shock regulation.

\(^{33}\) A marginal cost of $0.02$ per minute is reasonable, being positive but small. Hausman (2000) estimates marginal cost to be $0.05$ per minute when taking into account some costs of increasing network capacity.

\(^{34}\) T-Mobile and AT&T financial reports disclose that in 2003 their costs of acquiring a customer, including handset subsidies, were $329$ and $377$, respectively (T-Mobile 2004; AT&T Wireless Services, Inc. 2003). Averaged over a 24-month contract, these correspond to fixed costs of $14$ to $16$ per month.
Counterfactual Simulation Results

Bill-Shock Regulation. — Table 5 shows the results of our endogenous-price counterfactual simulations. Column 1 shows predicted plan prices and welfare outcomes under our estimated demand parameters. (These are the prices which were calibrated to match publicly available calling plans, which are reported in Table 7 in online Appendix A for October 2003.) The model predicts that firms offer a menu of three-part tariffs with monthly fees of $42.88, $48.64, and $58.12, corresponding allowances of 216, 383, and 623 peak minutes, and the maximum overage rate of $0.50 per minute.

Table 5 shows the results of our endogenous-price counterfactual simulations. Column 1 shows predicted plan prices and welfare outcomes under our estimated demand parameters. (These are the prices which were calibrated to match publicly available calling plans, which are reported in Table 7 in online Appendix A for October 2003.) The model predicts that firms offer a menu of three-part tariffs with monthly fees of $42.88, $48.64, and $58.12, corresponding allowances of 216, 383, and 623 peak minutes, and the maximum overage rate of $0.50 per minute.

Column 2 of Table 5 holds constant the predicted prices from column 1 but imposes bill-shock regulation. Holding prices constant, bill-shock alerts help all customers reduce overages and also gives 4 percent of customers the comfort to choose a smaller plan, as they know that the alerts will protect them from overages. In fact, with an overage rate of $0.50 per minute and price sensitivity parameter $\beta = 2.7$, consumers stop almost all calling after receiving an alert and pay negligible overage fees. Avoided overage charges and smaller plan choices correspond both to reduced bills and to reduced calling. Average annual welfare falls by $93 per student because marginal costs are only $0.02 and, hence, the reduced calling hurts consumers more than it lowers firm costs. Reduced bills drive annual firm profits down by $196 per student but reduced calling means average annual consumer surplus rises by only $103.

C. Counterfactual Simulation Results

Note: All welfare and profit numbers are expressed in dollars per customer per year. Because the counterfactuals in columns 4 and 5 produced two-part tariffs, bill-shock regulation has no additional effect. We simulate 10,000 consumers for 12 months.

| Biases: Bill-shock regulation | Estimates Estimates Estimates $\delta = 1$ No biases |
|-----------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| (1) | (2) | (3) | (4) | (5) |
| Plan 1 $M$ | 42.88 | 42.88 | 39.28 | 42.32 | 52.59 |
| $Q$ | 216 | 216 | 0 | 0 | 0 |
| $p$ | 0.50 | 0.50 | 0.17 | 0.13 | 0.07 |
| Share | 39 | 43 | 26 | 42 | 37 |
| Plan 2 $M$ | 48.64 | 48.64 | 50.66 | 70.63 | 69.41 |
| $Q$ | 383 | 383 | 80 | $\infty$ | $\infty$ |
| $p$ | 0.50 | 0.50 | 0.12 | N/A | N/A |
| Share | 38 | 36 | 23 | 46 | 52 |
| Plan 3 $M$ | 58.12 | 58.12 | 68.23 | |
| $Q$ | 623 | 623 | 540 | |
| $p$ | 0.50 | 0.50 | 0.12 | |
| Share | 14 | 11 | 40 | |
| Outside good share | 10 | 10 | 12 | 11 | 11 |
| Usage | 240 | 199 | 239 | 262 | 288 |
| Overage revenue | 223 | 2 | 152 | 136 | 75 |
| Annual profit | 501 | 305 | 509 | 512 | 512 |
| Annual consumer welfare | 903 | 1,006 | 870 | 884 | 907 |
| Annual total welfare | 1,404 | 1,311 | 1,379 | 1,396 | 1,419 |
| $\Delta$ annual profit | $-196$ | 7 | 11 | 11 | |
| $\Delta$ annual consumer welfare | $-103$ | $-33$ | $-19$ | 4 | |
| $\Delta$ annual total welfare | $-93$ | $-26$ | $-8$ | 15 | |
Column 3 of Table 5 imposes bill-shock regulation but allows firms to adjust prices. Firms adjust prices in several ways to mitigate lost overage revenues. First, firms adjust prices by lowering overage fees to $0.17 (plan 1) or $0.12 (plans 2–3), which encourages consumers to make some calls even after receiving a bill-shock alert. Absent bill-shock alerts, inattentive consumers use calling thresholds between $0 and $0.15 per minute (see online Appendix B.1). At the estimated price sensitivity parameter of $\beta = 2.7$, this implies inattentive consumers make 60 to 100 percent of calls whether or not they have exhausted their allowance of peak minutes. With bill-shock alerts, however, an overage rate of $0.50$ essentially stops consumers making all calls after receiving an alert and generates negligible overage fees. By reducing overage rates to $0.17$ or $0.12$, firms ensure that consumers continue to make 54 or 68 percent of calls after receiving an alert. Thus lowering overage rates prevents the complete collapse of overage revenue.

Next, firms adjust prices by lowering allowances of peak minutes, which helps offset reduced overage revenues in two ways. First, overages are triggered at lower usage levels and second, it causes consumers to choose larger plans (the share of plan 3 increases from 14 percent to 40 percent), and hence raises average monthly fees. Finally, firms increase the plan 3 monthly fee by $10 and make smaller changes to monthly fees of plans 1 and 2. The net effect is that annual overage fees per customer fall to $152$ (rather than to $2$ without price adjustment) and annual profits actually rise by $7$ per person.\(^\text{35}\) (This result seems surprising in light of the industry’s initial opposition to the regulation but may help explain their eventual voluntary agreement. Note that an additional counterfactual simulation predicts that a firm which introduced bill-shock alerts unilaterally would lose $165$ per customer annually, explaining why firms did not offer bill-shock alerts prior to regulation.)

In response to bill-shock alerts and price changes, 2 percent of consumers (largely plan 1 customers) switch to the outside option. This response is not because plan 1 becomes a worse deal. In fact, the monthly fee on plan 1 is reduced sufficiently that the average plan 1 consumer is better off by $13$ per year by staying on plan 1. Unfortunately, bias causes consumers to undervalue both bill-shock alerts and reduced overage rates. Overconfidence causes consumers to underestimate the chance of high usage and hence underestimate the likelihood of paying a reduced overage rate or receiving a bill-shock alert. As a result, the average plan 1 customer perceives himself to be worse off by $137$ per year by staying on plan 1. Thus the outside option becomes a more competitive alternative and gains share. Gains in the outside good share correspond to lost profits, lost social surplus, and lost consumer surplus.

Additional losses to social surplus occur because, while those who maintain service make almost the same number of calls on average (239 rather than 240 minutes monthly) they make on average lower value calls than before. This is because those who remain on the same plans reduce calling in response to lower allowances and bill-shock alerts while those who switch up to plan 3 increase calling in response to unambiguously lower marginal prices. Those cutting back forgo relatively

\(^{35}\)The fact that regulation limiting revenues from one set of fees may lead firms to raise other fees to compensate has been dubbed the “waterbed effect.” Genakos and Valletti (2011) document the waterbed effect following Europe’s introduction of caps on mobile call termination charges.
high-value calls while those calling more add relatively low-value calls and average value falls.

Combining the shift to the outside good with the shift to lower value calls, bill-shock regulation lowers annual social surplus by $26 per person. The fact that firm profits increase overall reflects the fact that profits per customer increase for those that maintain service, in large part from increasing the average monthly fee paid. As a result, average annual consumer surplus falls by $33 per person (or 4 percent of the average annual bill).

The $33 reduction in annual consumer surplus from bill-shock regulation reported in column 3 of Table 5 is an average effect. The left panel of Figure 8 shows a histogram of consumer utility changes due to the regulation. The distribution is right-skewed meaning that there is a tail of individuals who benefit substantially. Moreover, as we assume consumers are risk neutral, the predicted change in consumer surplus depends on changes in average bills but not on changes in bill dispersion. The right panel of Figure 8 shows that bill-shock regulation leads to an increase in small overages below $50 but a decrease in large overages above $50. Regulators may feel that reducing the incidence of large overages justifies reduced market coverage, less efficient calling, and higher firm margins.

Debiasing Consumers.—Turning to columns 4 and 5 of Table 5, we investigate the consequences of debiasing consumers. Column 4 shows the effect of eliminating overconfidence. In column 4, the firm finds it optimal to offer only two two-part tariffs: a $0.13 per minute plan and an unlimited talk plan for monthly fees of $42.32 and $70.63, respectively. Column 5 shows the effect of eliminating all biases (except myopic plan choice). In this case, the menu looks similar to column 4 but plan 1 is more similar to the unlimited talk plan, having a lower $0.07 rate per minute and a higher monthly fee of $52.59.

Comparing pricing in columns 4 and 5 to columns 1–3, notice that three-part tariffs disappear when overconfidence is eliminated—they are only offered to exploit overconfidence. Pricing in column 4 can be understood as a response to conditional mean bias. Estimated conditional mean bias implies that those who choose plan 1
underestimate their usage (so that an overage rate above marginal cost is optimal) but those who choose plan 2 overestimate their usage (so that charging an overage rate below marginal cost is optimal) (Grubb 2009).

Pricing in column 5 is similar to that in column 4 because debiased consumers still make plan choices myopically. In our model, myopia has a similar effect to conditional mean bias. This similarity arises because myopic consumers who receive a large signal $s_{ij}$ will choose a large plan tailored to high expected usage in the first month. Over time, however, inertia ($P_C = 0.06$) will keep them on the same large plan while their future signals and usage revert downward toward the mean. Similarly, those who receive small signals will choose plans appropriate for their first month’s low usage but too small for their long-run usage. As a result, pricing in column 5 shows qualitatively the same distortions away from marginal cost as column 4 but the distortions are smaller.

Somewhat surprisingly, while eliminating all biases raises total welfare, eliminating overconfidence alone lowers total welfare. It is straightforward to see why total welfare is higher in column 5 than column 4: marginal prices are closer to marginal cost in column 5 and so calling choices are more efficient. It is less clear why total welfare is lower in column 4 than column 1. The reason is that the marginal prices in column 4 ($0.13$ and $0.00$) are on average further from marginal cost ($0.02$) than are the calling thresholds ($v^o_{it}$) induced by the price menu in column 1. (Figure 9 in online Appendix B.1 depicts the calling thresholds induced by observed prices.)

Absent overconfidence, firms offer two-part tariffs rather than three-part tariffs and, as a result, bill-shock regulation has no effect. Moreover, the fact that bill-shock regulation is less important for unbiased consumers does not depend entirely on the elimination of three-part tariffs. In a final counterfactual, we simulate the effect of bill-shock regulation while holding observed public prices constant. In this simulation, bill-shock regulation benefits consumers with estimated biases by $25$ per year but benefits debiased consumers by only $1$ per year. Debiased consumers are affected less by bill-shock alerts because (even holding prices constant) they make better plan choices that lead to lower incidence of overages.

Robustness.—It is natural to ask at least four questions about the robustness of our results. First, our student sample may be less heterogeneous and either more or less price sensitive than the general population. Are our results robust to such variation? Second, as learning diminishes bias and debiasing makes bill-shock regulation irrelevant, does additional learning lead to smaller effects? Third, a weakness of our simulations is that the upper bound on overage rates of $0.50$ is binding at our estimates. Does the choice of upper bound affect our results? Fourth, absent an upper bound on overage rates, predicted overage rates would be far too high. What problem with our model can be diagnosed from this overprediction problem?

Online Appendix G answers the first three questions. We show that the predicted effect of bill-shock regulation on consumer welfare is qualitatively robust to (i) variation in price sensitivity and heterogeneity, (ii) three years of additional learning, and (iii) variation in the imposed overage rate cap of $0.35$ to $0.75$. In each case, bill-shock regulation is predicted to reduce average annual consumer welfare a small amount (varying between $1$ and $44$). As expected, additional learning reduces the effect of bill-shock regulation. Moreover, three other predictions are also robust: In
each case, bill-shock regulation is predicted to (i) lower overage rates, (ii) increase monthly fee revenue, and (iii) generate smaller costs to firms and smaller benefits to consumers when endogenous price changes are accounted for than when they are not.

Turning to the fourth question, comparison of columns 1, 3, and 4 in Table 5 shows that high overage rates result from the combination of overconfidence and inattention. The fact that we overpredict overage rates absent our restriction suggests one of two things. (i) We may be missing a countervailing force limiting overage rates such as limited liability, risk aversion, or regulatory threat. (ii) Overconfidence and inattention may be less pronounced in the general population than in our student sample. We omit the forces in (i) to avoid additional complexity. (In unreported analysis we found that adding limited liability at $100 per month to the supply model endogenously restricts overage rates to similar levels charged for roaming.) Like additional learning, less overconfidence or inattention would likely diminish the effects of bill-shock regulation.

While we cannot test the robustness of our predictions for every model variation, the clear logic underlying several of our predictions gives us confidence that they are more general than we have demonstrated: Holding prices fixed, a large effect of bill-shock regulation would be to transfer surplus from firms to consumers by reducing overage payments. Firms should respond, however, by lowering overage rates to encourage calling even after a bill-shock alert is received. Moreover, as bill-shock regulation does not directly affect firms’ market power, in equilibrium firms should make similar profits before and after regulation. This means that prices should adjust so that monthly fee revenues increase to partially offset reductions in overage payments. This could be via direct increases in monthly fees or other price changes (such as reduced minute allowances) that encourage consumers to choose more expensive plans. Therefore endogenous price changes should limit firm losses and consumer gains from bill-shock regulation relative to a prices-fixed world.

VIII. Conclusion

We specify and estimate a model of consumer cellular phone plan and usage choices. We identify the distribution of consumer tastes from observed usage and consumers’ beliefs about their future usage from observed plan choices. Comparing the two we find that students’ usage and plan choices are consistent with biased usage forecasts. In particular, students systematically choose overly risky, overly small, and overly extreme plans. These choices are consistent with underestimating noise in forecasts (overconfidence), overly low forecasts (negative aggregate mean bias), and overly extreme forecasts (conditional mean bias). Note that while we maintain the bias interpretation throughout the paper, there are other interpretations for some biases. For instance, as shown by Goettler and Clay (2011), the mean biases in our model are consistent with rational expectations and an unobserved aggregate shock to average tastes.

We conduct counterfactual simulations in which we (i) eliminate biases and (ii) quantify the welfare impact of implementing bill-shock regulation in 2002–2004. Our model predicts that eliminating biases increases consumer welfare, by $91 annually per consumer holding observed public prices fixed, but only by $4 annually per consumer accounting for firms’ endogenous pricing response. If public
prices do not respond to bill-shock regulation, then the average consumer will benefit by $25 annually. This finding is reversed when firms optimally respond to bill-shock regulation. Although consumers avoid overage fees, firms raise average monthly fees and average consumer surplus falls by $33 annually. This loss is an average and many consumers benefit from the policy. Moreover the incidence of overages over $50 falls substantially. Finally, we find that bill-shock regulation would have little to no effect if consumers were unbiased.

An important caveat to our results is that consumers may be unaware of their own inattention. Grubb (forthcoming) shows that unawareness of inattention creates overconfidence endogenously. If such naïveté is the source of consumer overconfidence then overconfidence itself should be reduced by bill-shock regulation. In this case our counterfactual predictions about the effect of debiasing consumers may be a better guide to the effect of bill-shock regulation. Our predictions should be applied to the bill-shock agreement implemented in 2013 with caution for a number of additional reasons. For instance, consumers today are more experienced and may be less biased. Moreover, text messaging and data plans were beyond the scope of our study. Finally, we reiterate that our sample consists of students who may not be wholly representative. Nevertheless, there are some robust predictions that can be applied with more confidence: Firms should lower overage rates but increase revenue from monthly fees. These endogenous price changes should make firm profits robust to bill-shock regulation, thereby undermining the benefits to consumers.

Our evaluation of bill-shock regulation could be insightful in other contexts. For instance, in 2009 US checking overdraft fees totalled more than $38 billion and have been the subject of new Federal Reserve Board regulation (Federal Reserve Board 2009; Martin 2010). Convincing evidence of consumer inattention (Stango and Zinman 2009, 2014) suggests that this fee revenue would be dramatically curtailed if the Fed imposed its own bill-shock regulation by requiring debit card processing terminals to ask users “$35 overdraft fee applies, continue Yes/No?” before charging fees. Our counterfactual shows that in the cellular context consumers are nevertheless made worse off after accounting for all price changes.

REFERENCES


