An Empirical Model of the Effects of "Bill Shock" Regulation in Mobile Telecommunication Markets

Jiawei Chen and Lai Jiang and Saad Andalib Syed Shah*

November 2021

Abstract

We develop an empirical model of consumer usage and price uncertainty under a three-part tariff plan to study the effects of an enacted "bill shock" agreement in mobile telecommunication markets, which requires mobile network operators to inform consumers when they use up the monthly allowance of their mobile phone plan. Using a rich billing dataset, we estimate an industry model of calling, subscription, and pricing. Our counterfactual simulations, which incorporate operators' price responses to compute the new equilibrium, show that the regulation would lead to lower fixed fees, lower allowances, and lower overage prices chosen by the operators. The regulation and the pricing changes in response to it would benefit both the consumers and the operators, and overall the total surplus would increase by $307 million per month. An increase in mobile penetration explains the joint increase in firm profits and consumer surplus.

Keywords: three-part tariff, usage uncertainty, price uncertainty, bill shock, regulation.

*Chen: University of California, Irvine; jiaweic@uci.edu. Jiang: University of California, Irvine; ljiang14@uci.edu. Syed Shah: University of California, Irvine; ssyedsha@uci.edu. We thank John Asker, Luis Cabral, Allan Collard-Wexler, Amihai Glazer, Robin Lee, Gary Lilien, Matthew Osborne, Charles Weinberg, Lawrence White, Chunhua Wu, Ting Zhu and conference and seminar participants at International Industrial Organization Conference, Marketing Science Conference, University of British Columbia, University of Victoria, University of California, Riverside, and University of California, Irvine for their helpful comments and suggestions. We are grateful to the editor and two anonymous reviewers for their constructive suggestions.

1 Introduction

A sudden and unexpectedly high bill for a subscription that shocks the consumer upon receiving it is known as a "bill shock". In mobile telecommunication markets, a bill shock is a sudden and unexpected increase in a mobile phone user's monthly bill that is not caused by a change in service plans. While the issue of bill shock is most prominent in mobile telecommunication markets, consumers have also faced bill shocks in other scenarios, including credit card bills, rental bills, energy bills, and medical bills, causing regulatory concerns (see, for example, Freking and Alonso-Zaldivar (2019) and Hausman (2019)). In many cases, firms design the contractual terms of their products and services in a way to exploit consumers' imprecise usage forecasts and extract surplus at consumers' expense (Grubb, 2015b).

In April 2013, an agreement between the Federal Communications Commission (FCC) and mobile network operators committed operators to alert consumers when they approach and exceed the voice, text, and data allowances included in their mobile phone plans. This agreement was reached as the response to a proposed bill shock regulation, which would require mobile network operators to inform consumers when they use up the monthly allowance of their mobile phone price plan (U.S. mobile network operators charge consumers a three-part tariff: a fixed monthly fee, a monthly allowance of free calling minutes, and an overage price per minute). The objective of the bill shock regulation was to reduce consumers' uncertainty regarding the marginal price they were paying for the next unit of consumption so that they would not be shocked by the bill they would receive at the end of the billing cycle. Under the three-part tariff pricing structure, the source of consumer marginal price uncertainty comes from consumers' usage uncertainty: they cannot keep perfect track of their usage, and so they don't know for sure whether their actual usage is below or above the monthly allowance (the marginal price changes drastically at the point of monthly allowance).

To better understand the effects of the bill shock regulation, in this paper we develop an empirical model of consumer usage and price uncertainty under the three-part tariff plan. We use this model to predict how mobile phone companies would adjust their pricing decisions if the bill shock regulation is implemented to eliminate uncertainty in consumer usage and price.

We present an empirical industry model in which consumers have price uncertainty when they make their calling decision on their mobile phones. This price uncertainty occurs because consumers are unsure of their exact usage relative to the number of free minutes (allowance) included in the plan. We model consumer price uncertainty by including a perception error (actual usage/perceived usage) in consumers' consumption decisions. We assume that the perception error has a mean of 1 and follows a log-normal distribution (this assumption is supported by results from a field study, reported below). With the perception error, consumers cannot keep track of their exact usage; instead, they recall their past usage with error. The presence of the perception error can be interpreted as limited consumer attention to keeping track of the exact usage.

The industry model has three stages. First, mobile network operators decide on the pricing structures of mobile phone plans; second, consumers decide whether to use mobile phones and, if so, which plan to subscribe to; and third, consumers make consumption decisions conditional on their chosen plan. The model is estimated using a rich billing dataset. We jointly estimate the consumers' preference for usage and subscription to mobile phone services. We then back out the mobile network operators' marginal cost using the demand estimates and the optimal pricing condition. Given these estimates, we simulate the price and quantity changes in the counterfactual scenario in which the regulation is implemented.

A crucial step in this estimation is identifying consumer price uncertainty. Our identification strategy is based on the lack of bunching at the point where the marginal price changes discontinuously: under the assumption that the distribution of consumer preference for calling is smooth, if consumers were aware of their exact usage, a mass point of consumers would use exactly their monthly allowance of free minutes; such bunching does not appear in the data, and this is informative about the degree of consumer price uncertainty.

In the counterfactual analysis, we study the effects of the bill shock regulation which requires cell phone companies to alert consumers when their usage hit the allowance. We first allow consumers to readjust their subscription and consumption decisions assuming no price adjustments from the firms. We then allow mobile network operators to readjust their prices in response to the bill shock regulation and, after finding the new price equilibrium, measure how consumer surplus and firm profits would be affected.

Assuming no price adjustments, we find that the elimination of usage and price uncertainty leads to a higher number of total calling minutes and a lower number of total overage minutes, with a large increase in the number of subscribing households using exactly their allowances. Consumers benefit from more calling minutes and lower overage payments, and monthly total consumer surplus increases by $53 million. Monthly total operators profit, on the other hand, decreases by $117 million mainly due to lower overage payments and higher costs associated with the larger number of calling minutes. Overall, monthly total surplus decreases by $64 million.

When price adjustments are incorporated into the analysis, we find that all major operators decrease the allowances, the fixed fees, and the overage prices in response to the regulation. Following the elimination of consumers' usage and price uncertainty and the operators' adoption of new pricing structures, the number of households who subscribe to mobile phone plans increases by 19.5% from 41 million to 49 million. The total number of monthly calling minutes remains largely unchanged, while the total number of monthly overage minutes increases as the new pricing structures (which feature increased per-minute prices for free minutes and reduced overage prices) induce more consumers to knowingly and optimally incur overage calling minutes.

Following the regulation and the resulting pricing changes, consumers incur lower fixed fee payments but higher overage payments. In addition, due to diminishing marginal utility of calling minutes for individual households, the fact that roughly the same number of calling minutes are now spread over a larger number of households results in an increase in the total utility from those calling minutes. The combined effect of those factors is an increase of $132 million in monthly total consumer surplus. The operators also benefit from the regulation and the resulting pricing changes, with monthly total operators profit increasing by $175 million. An increase in mobile penetration explains the joint increase in firm profits and consumer surplus. Overall, monthly total surplus increases by $307 million, which contrasts with the total surplus loss in the case without price adjustments and highlights the importance of accounting for firms' strategic responses when evaluating the effects of a regulation.

Theoretical work by Grubb (2015a) shows that the welfare effects of bill shock regulation are ambiguous. The seminal empirical work by Grubb and Osborne (2015) predicts that the regulation would lower average consumer welfare by about $33 per year. The panel nature of data in Grubb and Osborne (2015) allows them to address consumers' beliefs and learning: similar to Grubb (2009), consumers have biased beliefs, and furthermore consumer calling thresholds are less sensitive to the overage rate if individuals are overconfident, as their overconfidence leads them to underestimate the likelihood of an overage. In comparison, consumers have rational expectations in our model (the cross-sectional nature of our data prevents us from estimating consumers' beliefs), and consumers would increase overage calling as a result of a reduction in overage prices in our counterfactual simulations. In addition, the data used in Grubb and Osborne (2015) pertains to a specific type of consumer—university students who were enrolled with a single mobile network operator, whereas the data used in this paper is nationally representative and covers all carriers, which allows us to assess the effects of the regulation on a broader population. We therefore view our paper as complementary to Grubb and Osborne (2015), and together these papers allow a more comprehensive understanding of the important yet nuanced effects of the bill shock regulation.

Our paper also relates to studies on optimal three-part tariff plans. Fibich et al. (2017) consider a theoretical model in which a service provider sells to risk-neutral and rational consumers. They characterize the optimal three-part tariff plan under general conditions, focusing on the monopoly case and abstracting from competition among service providers. Baek and Brueckner (2015) consider a theoretical model of three-part tariffs in the presence of heterogeneous users. They find that the monopoly outcome yields under-consumption of the service by both high- and low-demand consumers, while the duopoly outcome under Bertrand competition is efficient. Our empirical model takes into account the strategic interactions among the major mobile-phone carriers when computing the new price equilibrium following the bill shock regulation. By incorporating the strategic interactions among the firms when computing the new equilibrium, our counterfactual results give a more realistic picture of what would happen if the regulation comes into effect.

In this paper, we apply our empirical framework to mobile telecommunication markets in the U.S., but our approach and findings have useful implications for other countries and industries as well, in light of similar bill shock regulations in those settings. For example, Roaming Regulation in EU requires operators to send free text messages to their travelling customers to notify them about roaming charges, in order to protect consumers from bill shock arising from those charges (McGowan, 2018). Likewise, U.K. Retail Banking Market Investigation Order 2017 requires that banks send customers a text alert before charging them for entering into an un-arranged overdraft (Feikert-Ahalt, 2019). The results in the present paper can serve as one benchmark for future studies in this area, and our empirical framework can be modified to fit other settings for investigating how firms' strategies will respond to such regulations and what the implications are for consumer welfare and industry profit.

The remainder of the paper is organized as follows. Section 2 presents the intuition of the model using diagrams. Section 3 proposes an empirical industry model with consumer usage and price uncertainty. Section 4 describes the data used for the estimation of the model. Section 5 discusses the identification and estimation results of parameters in the model. Section 6 discusses the effects of bill shock regulation via counterfactual simulations. Section 7 concludes.

2 Intuition of the model

Before introducing the formal model, we first use Figures 1-a and 1-b to show the intuition of the model proposed in this paper. These figures show one consumer's behavior under one particular plan with a monthly allowance of 120 minutes and an overage price of \$0.60/min (if this consumer uses fewer than 120 minutes this month, the marginal price for each calling minute is 0; if this consumer uses more than 120 minutes this month, the marginal price jumps to \$0.60/min.)

Before the implementation of bill shock regulation, this consumer has uncertainty about her actual usage and the actual marginal price for the next calling minute. Figure 1-a demonstrates the existence of perception error ω as the ratio between this consumer's perceived usage x and her actual usage $q = x\omega$; she never observes the actual realization of ω, so she is never sure about what her actual usage $q = x\omega$ is and can make her consumption decision based only on her perceived usage x instead. Figure 1-b shows the impact of the perception error on this consumer's calling decision and overage payment. At any perceived usage x, there is strictly positive possibility that this consumer's actual usage $q = x\omega$ is already longer than 120 minutes and that she has to unintentionally pay an overage price of $0.60/min; hence, this consumer's expected overage payment is strictly positive at any perceived usage x and is smoothed out at around 120 minutes.

3 An empirical industry model with usage and price uncertainty

In this section, we propose an empirical industry model in which consumers have price uncertainty when they make their usage decision on their mobile phones. This price uncertainty is caused by consumers' uncertainty regarding their exact usage relative to the number of free minutes (allowance) included in the plan.

3.1 Model setup

We make the following assumptions in the model. Consumers cannot perfectly recall their exact mobile phone usage, and their perceived (estimated) usage may differ from their actual usage. However, on average, consumers have a correct perception of their usage, and their perception error (actual usage/perceived usage) follows a log-normal distribution. Results from a field study, reported in the next subsection, support this assumption.

The industry model consists of three stages. In stage 1, mobile network operators set the pricing structures of their plans; in stage 2, consumers make subscription decisions (choose a plan from all the plans available in the market); in stage 3, consumers decide on their number of monthly calling minutes conditional on the plan chosen.

We begin with the last stage and work backwards.

3.2 Stage 3: consumers' calling decision

We consider consumers indexed by $i = 1, 2, \ldots, N_m$ in $m = 1, 2, \ldots, M$ markets. Consumers first decide whether to subscribe to a mobile phone service. Conditional on subscribing to the mobile service, consumer i chooses a plan from the set of available plans, indexed by $j = 1, 2, \ldots, J_m$, offered by carriers $k = 1, 2, \ldots, K_m$, and the number of calling minutes x_i using the plan.[1] To

[1] In reality, consumers use cell phone plans for a long time, and because of inertia or switching cost, the plan used by consumers may not be optimal with respect to the choice set available. In this model, we abstract from such inertia or switching cost.

use plan j, consumers must pay a monthly fixed fee, F_j; A_j minutes are included in plan j; once consumers use more than A_j minutes in a given month, they must pay a per-minute overage price of p_j.

Consumer i faces a time constraint T. She allocates her time to either talking on her mobile phone or outside activities (the marginal utility of which is normalized to 1) subject to the time constraint T.[2] Conditional on choosing plan j, consumer i chooses the number of calling minutes x_{ij} and the quantity of time spent on the outside activities x_{i0} to maximize her surplus.

We model consumer price uncertainty by including a perception error in consumers' consumption decisions. With this perception error, consumers cannot keep track of their exact usage and recall past usage incorrectly. The presence of the perception error can be interpreted as limited consumer attention to keeping track of exact usage. For simplicity, we assume that this perception error does not change during the billing month. To be specific, a perception error is randomly determined in the beginning of the month without being observed by the consumer. We assume that the consumer knows the distribution of the perception error but does not know its exact realization. The consumer realizes her perception error at the end of the month by receiving the bill. Under this specification, consumers' perceived usage is modeled as x_{ij}, while their actual usage is $q_{ij} = x_{ij}\omega_i$. Here, ω_i is consumer i's perception error that measures the ratio of actual usage over perceived usage. Since ω_i is not observed, the consumer maximizes her expected utility conditional on the distribution of ω_i:

$$\max_{x_{ij}} v_{ij}(x_{ij}) = \int_{\omega_i} \overbrace{\theta_i ln(x_{ij}\omega_i)}^{\text{utility from calling}} + x_{i0} + \overbrace{\alpha_i p_j \max\{(x_{ij}\omega_i) - A_j, 0\}}^{\text{disutility from payment}} dF(\omega_i). \quad (1)$$

$$\text{subject to} \quad x_{ij}\omega_i + \underbrace{x_{i0}}_{\text{outside activity}} \leq \underbrace{T}_{\text{time constraint}}$$

Let x^*_{ij} be the value of x_{ij} that solves equation 1 (see Appendix A.1 for more details). [3] The realized usage is the product of the optimal perceived usage and the perception error: $q_{ij} = x^*_{ij}\omega_i$. The maximum monthly utility from calling using plan j for consumer i is therefore

$$v_{ij}(x^*_{ij}; \theta_i, \alpha_i, A_j, p_j) = \int_{\omega_i} \theta_i ln(x^*_{ij}\omega_i) + \alpha_i p_j \max\{(x^*_{ij}\omega_i) - A_j, 0\} + T - (x^*_{ij}\omega_i) dF(\omega_i). \quad (2)$$

[2]The time constraint ensures that the number of calling minutes is bounded at a marginal price of zero.

[3]In the paper, we assume that consumers could increase calling minutes by answering inbound calls or otherwise by making outbound calls. Also it is assumed that consumers could, if necessary, stop increasing calling minutes by not answering inbound calls or stop making outbound calls.

Income effect and the preference parameter We allow α_i, the price coefficient, to vary as a function of a household's monthly income per person: [4]

$$\alpha_i = \bar{\alpha} + \alpha_D D_i^{\alpha}. \tag{3}$$

The preference parameter θ_i measures how many minutes consumer i will call monthly if the marginal price of calling is zero. θ_i varies as a function of consumers' observable and unobservable characteristics. We restrict θ_i to be positive by specifying it as an exponential function of consumers' characteristics

$$\theta_i = exp(\bar{\theta} + \theta_D D_i + \nu_i), \tag{4}$$

where $\{\bar{\theta}, \theta_D\}$ are parameters and D_i is a column vector of consumers' key demographic characteristics.[5] ν_i represents consumers' unobservable heterogeneity. We assume that ν_i has a normal distribution with mean 0 and variance σ^2. As discussed in Berry et al. (1995), the observable and unobservable heterogeneity in θ_i ensures that consumers who have a strong preference for calls (high θ_i) will tend to attach high utility to all plans with large minutes allowances. This specification allows plans with similar minutes allowances to be close substitutes for each other.

Discussion: perception error We choose to incorporate the perception error in consumers' consumption choice to reflect the fact that consumers have uncertainty about their actual usage relative to the allowance included in the three-part tariff plan. This in turn translates into consumers' uncertainty about the exact marginal price for the next calling minute in the context of the three-part tariff plan. Different from the marketing literatures on two-part tariffs (Danaher (2002), Essegaier et al. (2002), and Kumar and Rao (2006)), this modeling choice is specific to a three-part tariff context (as in Lambrecht and Skiera (2006), Iyengar et al. (2007), Lambrecht et al. (2007), and Fibich et al. (2017)).

The model proposed here differs from those in the previous literature on three-part tariff in the sense that it incorporates a new dimension of consumer usage uncertainty and price uncertainty that is consistent with the bill shock regulation. The same modeling approach could be applied to the context with a block-pricing structure, in which the marginal price changes according to the cumulated usage, as with electricity pricing. In a two-part tariff context in which the marginal price does not change according to the usage, the inclusion of a perception error as formulated in this paper does not affect the actual usage q on average (see Appendix A.2 for detail), but does

[4] Monthly income per person = $\frac{\text{monthly income}}{\text{household size}}$, D_i^{α} is the high income dummy which equals to 1 if household i has monthly income per person higher than \$1000 per month.

[5] Consumers' key demographic characteristics include family dummy, the age of the head of the household is over 55 dummy, and renting dummy.

add noise to q which reduces welfare. In particular, even on a two-part tariff with known marginal price, introduction of the perception error still means that consumers will be making suboptimal consumption choices due to not knowing their exact marginal value for an additional perceived minute. This will reduce consumer surplus and affect the extensive margin (resulting in lower penetration or lower fixed fees or both).

We would also like to point out that in our model, the effect of the bill-shock regulation changes discontinuously as the plan allowance A_j decreases to zero. For A_j close to zero (e.g., 1 minute), the consumer receives an alert at the beginning of the month and immediately infers the entire month's perception error ω. The consumer can then perfectly implement her optimal calling minutes. However, if $A_j = 0$ as in a two-part tariff, then she does not infer ω and makes suboptimal consumption choices due to the perception error.

In reality, the perception error likely evolves over the course of a billing period, and the smaller is A_j, the less about ω can be inferred from an alert. In such a setting, the effect of alerts would be continuous as A_j hits zero. The discontinuity in our model arises from the simplification that ω is realized once for the entire month at the beginning of the month and can be perfectly inferred from arbitrarily small usage, but not from zero usage. This assumption may overestimate the benefits of the bill shock regulation in cases in which firms set very small allowances. This issue does not appear to be particularly relevant for our empirical application, though, as even in the counterfactual scenario in which allowances are reduced (Subsection 6.2), the overage probability only goes up to 51%, indicating that the allowance is large enough to accommodate all calls half the time.

Field study supporting the model assumption on the perception error To assess the model assumption on the perception error, we conducted a field study at a major university in North America. We asked people passing by a hot spot of the university during lunchtime to fill out a survey in exchange for a chocolate bar. On the first page of the survey, we asked respondents to estimate their current usage of voice, text and data during this monthly billing cycle. We then asked them to turn over the page and check their actual usage of voice, text and data this month, either from their phones or by logging on to their online account.

We collected 100 surveys from this field study. In the end, 86 respondents completed the information on estimated voice usage and actual voice usage; 82 completed the information on estimated text usage and actual text usage; 75 completed the information on estimated data usage and actual data usage. We define people's perception error as the ratio of their actual usage over their estimated (perceived) usage. The results presented here regarding respondents' perception errors on voice, text and data usage come from surveys with completed information on estimated and actual usage of voice, text and data.

Table 1 presents summary statistics of respondents' perception errors on their voice usage, text usage and data usage. The table shows that the mean of perception errors on voice, text and data is close to 1; i.e., on average, people have a correct perception about their real usage. Figure 2 presents the histogram of the perception errors of respondents' voice, text and data usage from this field study. The figure shows that the distribution of respondents' perception errors on their voice, text and data usage resembles a log-normal distribution.

Furthermore, the correlation between perceived voice usage and the absolute difference between perceived voice usage and actual voice usage is 0.89, which means that the bigger is the perceived voice usage, the bigger is the absolute difference between perceived voice usage and actual voice usage. This finding lends support to the multiplicative (as opposed to additive) formulation of the perception error.

Additionally, the correlation between perceived voice usage and actual voice usage is 0.43. For comparison, at our model estimates (obtained later in the paper), the model predicted correlation between perceived voice usage and the absolute difference between perceived voice usage and actual voice usage is 0.51, and the model predicted correlation between perceived voice usage and actual voice usage is 0.90. The differences between the correlations from the field study and those predicted by the model are in part due to the fact that the field study and the observational data come from very different time periods and pertain to different populations (see more detail below), and we consider our assumption a useful though imprecise way of characterizing the perception error.

Discussion: rational expectations We would like to note that the survey results described above, which we use to support the assumption of consumers' rational expectations about their real usage, have some limitations. Specifically, although the wireless data is from 2000-2001, the field study was conducted in 2011. The consumers surveyed in 2011 likely have a better sense of their phone usage than the consumers in 2000-2001 for at least two reasons. First, cell phones had been around for much longer in 2011 than in 2000-2001, and so consumers in 2011 might be more familiar with cell phones and therefore have a more accurate perception of their phone usage. Second, in 2011, the presence of smartphones (which weren't around in 2000-2001) means consumers can track their phone usage more easily. Because of these differences, consumers' perception of their phone usage might be less biased in 2011 than in 2000-2001. As a result of these limitations, the survey results should be interpreted with some caution, and we think of the rational expectations assumption based on the survey results as an approximation, which is imprecise but nonetheless a useful step toward modeling and understanding consumers' perception errors.

Our survey was conducted on university students, and the empirical study in Grubb and Osborne (2015) also uses data on university students. While Grubb and Osborne (2015) model biased beliefs, we assume rational expectations, and the differences in the two papers' approaches and findings are in part attributable to the differences between the time periods in which the two studies were conducted as well as the differences between the sample taking our survey and the student population of cell phone users studied in Grubb and Osborne (2015).

Discussion: perfect foresight In our model, consumers have perfect foresight in the sense that at the beginning of each month they know the optimal number of calling minutes to make in that month, and the only uncertainty they have is about how to implement it. In reality, consumers likely do not have such perfect foresight and therefore also face some uncertainty about the optimal number of calling minutes, but this kind of uncertainty is not captured in our model. Incorporating such uncertainty into the model would smooth out bunching even without a perception error. To the extent that the perception error is identified based on the lack of bunching, we may be overestimating the perception error's variance and thus the effect of the bill shock regulation.

Discussion: all calls have the same value For simplicity, our model implicitly assumes that all calls have the same value. In reality, diminishing marginal utility of call minutes can come from two sources, one of which we capture in the model and the other we don't. First, different kinds of calls may have different inherent values (depending on who we are calling to, for example). Marginal utility of calling minutes is then decreasing if we order the calling minutes by value from high to low (as opposed to chronologically). Second, even for calls of the same kind (talking to the same person, for example), there may be diminishing returns to talking. Our paper assumes the latter (while Grubb and Osborne (2015) assume the former).

When calls have different values, under bill shock regulation, when an individual reaches the allowance and therefore receives an alert, she may wish she could go back and unmake some low-value calls in order to have enough minutes to make more high-value calls during the remainder of the month. By assuming away this type of inefficiency under bill shock regulation, our assumption that all calls have the same value may result in an overestimation of the bill shock regulation's benefit for consumers, which could partially explain why our paper predicts a larger increase in consumer surplus from the bill shock regulation than Grubb and Osborne (2015).

3.3 Stage 2: consumers' subscription decision

Utility from calling is only part of the consumer's utility from subscribing to a plan. In particular, the consumer suffers from the disutility of paying the plan's monthly fixed fee. We assume that the total monthly utility that consumer i enjoys from subscribing to plan j in market m is:

$$u_{ijm} = v(x^*_{ijm}; \theta_i, \alpha_i, A_j, p_j) + Z'_{jm}\lambda + \alpha_i F_{jm} + \xi_{jm} + \sigma_\epsilon \epsilon_{ijm}, \tag{5}$$

where $v(x^*_{ijm})$, defined in equation 2, is the maximum monthly utility from using plan j for consumer i, and λ are taste parameters for plan j's attributes Z_{jm}. We include dummy variables in Z_{jm} such as year, firm, and whether roaming and long distance minutes are included in the monthly allowance.

We assume that the utility from the outside option in market m is $T + \sigma_\epsilon \epsilon_{im0}$, which is the utility that consumers get by spending all of their time on outside activities. The interpretation of the utility that consumer i derives from plan j is the difference relative to the outside option. Given the distribution of utility function parameters and the plan's attributes in a given market, we can compute the model's predicted market shares by aggregating over utility-maximizing households.

Finally, for computational simplicity, we assume that the idiosyncratic errors ϵ_{ijm} have an i.i.d. type I extreme value distribution. We denote the standard deviation of idiosyncratic errors to be σ_ϵ, which we estimate. The normalization of the marginal utility of outside activities $T - (x^*_{ij}\omega_i)$ to 1, stated previously in Subsection 3.2 and used in equation (2), is what fixes the model's scale of utility and thereby allows the identification of the error term's standard deviation σ_ϵ.

Let F_i^m be the distribution of consumer preferences and demographics in market m. Given the distribution assumption on ϵ_{ijm}, the model's predicted market share for plan j in market m is:

$$s_{jm} = \int \{ \frac{exp((\delta_{jm} + \mu_{ijm})\sigma_\epsilon^{-1})}{1 + \sum_k exp((\delta_{km} + \mu_{ikm})\sigma_\epsilon^{-1})} \} dF_i^m, \tag{6}$$

where $\delta_{jm} = Z'_{jm}\lambda + \bar{\alpha}F_{jm} + \xi_j$ and $\mu_{ijm} = v(x^*_{ijm}; \theta_i, \alpha_i, A_j, p_j) + (\alpha_i - \bar{\alpha})F_{jm}$. We aggregate the demand at the plan level. In the estimation, we take a "Micro BLP" approach and match the model prediction with the data both at the aggregate level-market shares and micro level-moments of monthly calling minutes.

3.4 Stage 1: mobile network operators' pricing decision

A mobile network operator's gross profit (i.e., profit before fixed costs) in a market is

$$\pi_{fm}(\overrightarrow{F_{fm}}, \overrightarrow{A_{fm}}, \overrightarrow{p_{fm}}, \overrightarrow{J_{fm}}) = N_m \sum_{j \in \overrightarrow{J_{fm}}} s_{jm}(\overrightarrow{F_{fm}}, \overrightarrow{A_{fm}}, \overrightarrow{p_{fm}}, \overrightarrow{J_{fm}}) \Bigg[F_{jm} - C_{fm} \tag{7}$$

$$+ \sum_{i \in I_{jm}} \left\{ \int_{\omega_i} \left(p_{jm} \max\{x^*_{ijm}\omega_i - A_{jm}, 0\} - c_{fm}(x^*_{ijm}\omega_i) \right) \, dF(\omega_i) \right\} \frac{s_{ijm}(\overrightarrow{F_{fm}}, \overrightarrow{A_{fm}}, \overrightarrow{p_{fm}}, \overrightarrow{J_{fm}})}{\sum_{i \in I_{jm}} s_{ijm}(\overrightarrow{F_{fm}}, \overrightarrow{A_{fm}}, \overrightarrow{p_{fm}}, \overrightarrow{J_{fm}})} \Bigg],$$

where m denotes market, f firm, and j plan. $\overrightarrow{J_{fm}}$ is the set of plans offered by firm f in market m, with corresponding sets of monthly fixed fees $\overrightarrow{F_{fm}}$, allowances $\overrightarrow{A_{fm}}$, and overage prices $\overrightarrow{p_{fm}}$; N_m is the number of consumers in market m; s_{jm} is the market share of plan j in market m; C_{fm} is firm f's cost of serving one consumer in market m; I_{jm} is the set of consumers in market m choosing plan j; c_{fm} is firm f's marginal cost per minute in market m; $x^*_{ijm}\omega_i$ is the number of minutes used by consumer i choosing plan j; and s_{ijm} is the probability of consumer i choosing plan j in market m.

Mobile network operators compete by choosing plans' pricing structures to maximize profits. A complete pricing-strategy profile for one mobile network operator in one market includes the fixed fee, allowance, and overage price for each plan that the operator offers in that market. In our counterfactual analysis, reported below in Section 6, we allow the mobile network operators to re-optimize their pricing strategies in response to the regulation. To make the problem tractable, we simplify each mobile network operator's pricing strategy in each market to three variables: the level of fixed fees LF_{fm}, the level of overage prices Lp_{fm}, and the level of allowances LA_{fm}. The initial levels of the pricing structures correspond to $LF_{fm} = 1$, $LA_{fm} = 1$, and $Lp_{fm} = 1$. If mobile network operator f decides to increase the level of fixed fees LF_{fm} in market m by 20 percent, then that means the fixed fees of all plans offered by this mobile network operator f in market m are increased by 20 percent, and LF_{fm} is increased from 1 to 1.2. Similarly, if mobile network operator f decides to decrease the level of overage prices Lp_{fm} in market m by 20 percent, then that means the overage prices of all plans offered by this mobile network operator f in market m are decreased by 20 percent, and Lp_{fm} is decreased from 1 to 0.8. Similarly for LA_{fm}.

The above simplifying assumption makes the model tractable. If we instead assume that each firm solves a multi-product profit maximization problem in each market by simultaneously choosing all three choice variables F_{jm}, p_{jm}, and A_{jm} for each plan it offers in that market, taking all other firms' choices as given, there will be a large number of choice variables jointly chosen by each firm, and it quickly becomes infeasible to solve for the Nash equilibrium among all the firms.

Moreover, anecdotal evidence suggests that in the telecommunication industry, firms often make pricing changes that are proportional across multiple plans that they offer, lending some empirical support to our assumption. For example, Pressman (2019) documents that following similar price cuts at Verizon, in October 2019 AT&T announced new unlimited data plans that were about 10% cheaper than the offerings they replaced, representing proportional price changes applied to multiple plans at the same time. Nonetheless, given the restriction that we have had to make in the model to make it tractable, the counterfactual results we report below should be viewed as an approximation and should be interpreted with some caution.

The cost structure For each mobile network operator f in market m, the total monthly cost (TMC) is defined as

$$TMC_{fm} = N_{fm}^{cus} C_{fm} + N_{fm}^{min} c_{fm} + FMC_{fm}, \tag{8}$$

where N_{fm}^{cus} is the total number of consumers served by firm f in market m; N_{fm}^{min} is the total number of calling minutes by all consumers of firm f in market m; C_{fm} is the cost of serving one consumer for firm f in market m; c_{fm} is the marginal cost per minute for firm f in market m; and FMC_{fm} is the fixed monthly operating cost that is not affected by the number of consumers served or the total monthly calling minutes. The per-consumer cost includes the cost of customer service, billing, etc. The per-minute cost includes the usage payment for (often long distance) communication infrastructure owned by other companies, shadow cost due to the network's capacity constraint, and so on. In particular, if the demand of calling minutes for a given network exceeds its capacity, then some calls need to be dropped. Each additional calling minute induces a shadow cost due to the increase in the probability of exceeding the capacity constraint.

4 Data

The main data source for this paper is the bill-harvesting data collected by TNS Telecoms.

4.1 TNS national survey

TNS conducts a quarterly national survey of U.S. households. The sample used in the paper includes the years 2000-2001, or eight quarters in total. The time period of these data has two advantages: (1) In 2000-2001, voice was the major function of mobile phones, which provides a cleaner setting in which to focus on the voice usage of mobile phones only; (2) Mobile phones were more homogeneous in 2000-2001 than they are today due to the absence of smart phones. Naturally, the older a dataset is, the more difficult it is to apply it to current issues. That said, a study based on such historical data can still provide useful implications for the present day; in particular, even though text messages and data usage have become important functions of mobile phones, three-part tariffs apply to text messages and data usage as well and therefore remain relevant. Of course, the present dataset has the limitation that it is not very recent to reflect the exact magnitude of changes caused by the regulation. A newer dataset that includes the observations after the changes would allow a more precise evaluation of the impact of the regulation.

In its survey, TNS asks about households' characteristics and ownership of mobile phones. Among 263,707 observations appearing in the survey in 2000-2001, 262,826 have complete key demographic information. Among these 262,826 households, 130,259 (50%) of them own at least one mobile phone. 16,914 of these 130,259 households provide their mobile phone bills.

These data give us information on the mobile phone penetration rate in each market. In estimation, the market share of the outside good is constructed as "1 - penetration rate".

4.2 TNS mobile phone bills

As mentioned in the previous section, around 16 percent of households in the TNS national survey handed in their mobile phone bills. There are, in total, 17,155 mobile phone bills; we call these the bill data. In a separate file, 11,051 bills have detailed information on each outgoing and incoming call during the month; we call this the call detail data. See Appendix B.1 for more details on the bill data. Table 2 presents the counts of bills and market shares of major mobile network operators. Table 3 describes the summary statistics of the bill data and the call detail data.

4.3 Tariff data

MyRatePlan.com collects pricing plans charged by different mobile operators.[6] We use tariffs offered in the same period as the sample period (year 2000-2001) of the bill data to construct the choice set of consumers in each market. A plan is uniquely defined by five key characteristics: monthly fixed fee; allowance; overage price; long distance price, and roaming fee. The plan's coverage is directly associated with the long distance and roaming fees: local plans charge a strictly positive price for both long distance calls and roaming calls; regional plans offer free long distance calls and charge a strictly positive price for roaming calls; national plans offer free long distance and roaming calls.

4.4 Estimation sample

Bill data is matched with tariff data to construct the estimation sample, a process that drops 73% of the data (from 11,051 bills with call detail information to 2,992 matched bills) and results in a

[6] http://www.myrateplan.com/.

sample that may not be representative. See Appendix B.2 for more details on the data matching process.

Table 4 shows the summary characteristics of key variables of aggregate-level and micro-level data in the estimation sample. At the aggregate market level, Table 4 shows summary statistics on the number of providers, the market shares of the biggest and smallest provider, and the average number of plans offered per provider in each market. At the aggregate plan level, there are in total 577 plans in the 26 markets considered, and Table 4 shows the summary statistics on the key characteristics of the plans offered. At the micro level, there are in total 1,987 cell phone bills in all 26 markets considered, and Table 4 shows summary statistics on key demographic variables and cell phone usage variables associated with the cell phone bills. Altogether, 1,987 cell phone bills are used to compute the market shares of plans and providers.

A strength of the data that we use is the fact that it includes rich information on individual billing details from multiple firms, while a limitation is that the matching process results in the dropping of a large percentage of data and the remaining sample may not be representative. In some cases the consumer's chosen plan is not in the set of currently available plans because a consumer has been grand-fathered in on plan terms chosen earlier which are no longer offered, and such observations are dropped. The sample is then skewed toward new customers or customers who have switched plans recently. A comparison between Tables 3 and 4 shows that the estimation sample differs somewhat from the bill data on the statistics of monthly calling minutes (the mean and standard deviation are 154 and 329 for the bill data and 185 and 298 for the estimation sample).

5 Identification and estimation results

In this section, the model developed in section 4 is estimated. Table 5 shows the estimates of all parameters in the model and their standard errors. In the estimation of standard errors, we take into account both sampling error and simulation error as in Berry et al. (2004). The parameters are estimated using the following algorithm.

5.1 Estimation algorithm

For a given value of nonlinear parameters, $\{\alpha_D, \sigma_\epsilon, \bar{\theta}, \sigma, \sigma_\omega, \theta_D\}$, we construct the model prediction on monthly calling minutes and on the market share of plans.

Step 1: Simulate the price coefficient α_{im} and the preference parameter θ_{im} for each simulated consumer i in market m.

We simulate $i = 1, 2, \ldots, N_m$ in $m = 1, 2, \ldots, M$ markets. The demographics of each simulated consumer D^{α}_{im} and D_{im} in market m are drawn from the observations in the corresponding market in the national survey data. We also draw one realization of usage shocks ν_{im} for each simulated consumer from the assumed distribution (normal with mean 0 and variance σ^2). Each simulated consumer i's price cofficient α_{im} and calling preference θ_{im} are computed according to equations (3) and (4), respectively.

Step 2: Given α_{im} and θ_{im} for each simulated consumer i in market m, compute consumer i's perceived optimal usage under plan j, x^*_{ij}. Then compute the utility each simulated consumer i gets from plan j and the model prediction on the market share of plans.

Consumer i's perceived optimal usage under plan j, x^*_{ij}, can be obtained by solving equation (11) in Appendix A.1. The utility each simulated consumer i gets from plan j can be computed using equation (5). And the model prediction on each plan's market share can be computed using equation (6).

Step 3: Recover the model's prediction on the probability of consumer i choosing plan j in market m, $\hat{s}_{ijm}$, and use $\hat{s}_{ijm}$ as a weighting measure to construct moments of the model predicted monthly calling minutes conditional on plan choices and subscribing to mobile phone services. Then construct moments used in the estimation by taking the difference between the model predicted moments and the moments in the data.

5.2 Moment conditions

We use four sets of moment conditions in the estimation. (1) Match the model predicted probability that the monthly calling minutes fall in between 90% and 110% of the allowance to the observed probability, for each of the two income groups (low-income and high-income). Two moments in this set. (2) Match the model predicted mean of monthly calling minutes to the observed mean, for each of the eight combinations of the three demographic groups (family, age, and rent). Eight moments in this set. (3) Match the model predicted coefficient of variation in monthly calling minutes to the observed coefficient of variation. One moment in this set. (4) Match the covariance of demand-side instruments Z^d_{jm} with the unobserved demand shock ξ_{jm} to zero. Our instruments follow standard practice in demand estimation using aggregate data. First, we allow observed product characteristics, Z_{jm}, to instrument for themselves. Observed product characteristics (eight in total) include dummy variables for non-minutes plan characteristics such as firm, year, etc. Second, we account for price endogeneity by instrumenting for it using the average price of plans with similar allowance levels offered in the same economic area groupings, but outside the same economic area. We have seven such Hausman price instruments. So there are $8 + 7 = 15$ moments in this last set. Combining all four sets, altogether there are $2 + 8 + 1 + 15 = 26$ moment conditions.

In each set, we match the model predicted moments to the corresponding moments observed in the data, where the model prediction is constructed as the weighted average of the average per market using the number of observations per market in the data as the weight. The average in each market is computed by averaging the weighted average of each simulated individual, using the probability of each individual choosing a particular plan conditional on choosing mobile phone services as the weight.

5.3 Identification and estimation results of consumers' preference parameters

We first estimate the distribution of the preferences for calling on mobile phones θ_i and the distribution of the perception error ω_i using individual calling data. We then estimate jointly with the price coefficient α_i and the non-price preference parameters λ_j using market share, price, and plan characteristics data. Recall that consumers make a choice of plan based on the preference parameter θ_i, which is observed fully by consumers but not fully by the econometrician. For this reason, when observing consumption patterns, we need to take into account the bias created by selection into plans. We correct for this selection bias by constructing moments of the model's prediction on monthly calling minutes conditional on plan choices and subscribing to mobile phone services. The conditioning on plan choices requires knowing the parameters of the model of plan choices (stage two in the model, given in equation (5)). [7]

Consumer i's monthly calling minutes on plan j, x_{ijm}, is obtained by solving equation 11. Hence x_{ijm} depends on the calling preference θ_i, the distribution of perception error $F(\omega_i)$, plan j's monthly allowance A_j, and plan j's overage price p_j. The calling data are the measurement of monthly calling minutes at the individual level. We estimate the distribution of θ_i and ω_i by matching moments of the model's prediction of monthly calling minutes to moments in the calling data.

The constructed moments are sensitive to model parameters. For example, the mean of monthly calling minutes for the eight combinations of three demographic groups (family, age, and rent) is sensitive to the demographic parameters in the preference parameter θ_i. And the coefficient of variation in monthly calling minutes is sensitive to the standard deviation of the unobserved heterogeneity, σ: because of the exponential specification of the preference parameter (see equation 4), the unobserved heterogeneity is in fact multiplicative of the mean of the preference parameter, which makes the coefficient of variation (i.e. the ratio between standard deviation and the mean) a more suitable moment for the identification of this parameter.

[7] We jointly estimate the parameters of the distribution of calling preferences, the price coefficient, and the perception errors, together with the plan choice parameters, as in Lee (2013).

5.3.1 Identification and estimation results of the perception error

The perception error is identified using the smoothness of the distribution of the usage ratio around 1 and the parametric assumption on the perception error. Figure 3 shows the histogram of the usage ratio in the data: there is no clear mass point in the distribution of the usage ratio around 1. The perception error is assumed to follow a log-normal distribution with parameters μ and σ_ω, which are the mean and standard deviation of ω_i's natural logarithm. In addition, we also assume that consumers have, on average, correct perception of their actual usage, i.e., $E(x_{ij}\omega_i) = x_{ij}E(\omega_i) = x_{ij}$. Given that ω_i is log-normal, this implies that $\mu = \frac{-\sigma_\omega^2}{2}$. Under these parametric assumptions, the parameter σ_ω determines the distribution of the perception error ω_i. Here, we assume that the variance of perception error is different for high income and low income consumers and stationary over time. In reality, consumers could learn about the variance of perception error over time; however, without panel data that tracks the same consumer's usage over multiple months, we cannot incorporate the learning process in the current setting.

σ_ω is identified by the key moment in the data: the probability of the usage ratio being between 0.90 and 1.10, that is, the probability that consumers' actual usage level is between 90 percent and 110 percent of the monthly allowance. Table 6 compares this moment in the data and the same moment simulated from the model by setting the key parameter σ_ω at different levels for high income and low income consumers. Table 6 also shows the value of the key moment for two other values of the variance of ω_i—specifically, 25 percent and 50 percent of the estimated value. Table 6 confirms the intuition discussed above. When σ_ω is zero, i.e., when consumers have a precise perception of their actual usage level, a large proportion of consumers end up using between 90 and 110 percent of their allowance because of the discontinuity in the marginal price. The larger the variance in consumers' perception error is, i.e., the larger σ_ω is, the lower is the proportion of consumers who end up using between 90 and 110 percent of their allowance.

5.3.2 Identification and estimation results of the price coefficient

The price coefficient is identified using moment (4), the covariance of demand-side instruments, Z^d_{jm}, with the unobserved demand shock, ξ_{jm}. Our instruments for the monthly fixed fees of plans follow standard practice in demand estimation on aggregate data. First, we allow observed product characteristics, Z_{jm}, to instrument for themselves. Observed product characteristics include dummy variables for non-minutes plan characteristics such as firm, year, etc. Second, we account for price endogeneity by instrumenting for it with the average price of plans with similar allowance levels offered in the same economic area groupings, but outside the same economic area.[8] We discretize the allowance level into six different groups: smaller than 100 minutes; 100 to 200 minutes; 200 to 300 minutes; 300 to 400 minutes; 400 to 600 minutes; and more than 600 minutes. Following Hausman (1997), these are often called Hausman instruments. These instruments have been used for demand estimation in settings such as Nevo (2001). We have tested for weak instruments. The concentration parameter μ^2 equals 10281, and the number of instruments K is 7. $\mu^2/K = 1469$, large enough that we can conclude that the instruments are not weak.

The second and third rows of Table 7 present the mean and median of the price elasticities of major mobile network operators's market shares with respect to the fixed fee levels. The fourth and fifth rows of Table 7 present the mean and median of the price elasticities of overage minutes (the number of minutes that are charged with overage prices) of major mobile network operators with respect to the overage price levels. The overage minutes' price elasticity of a particular firm in each market is computed as the percentage change of the firm's number of overage minutes when the overage prices of all plans offered by this firm in this market are increased by one percent, while holding the prices of the other firms constant. For example, Sprint was operating in 20 out of 26 markets in the estimation sample; if Sprint increases the overage prices of all its plans in one market by one percent, while holding the prices of the other firms in the same market constant, Sprint's number of overage minutes would, on average (average across markets where this firm is present), decrease by 1.93 percent.

5.4 Identification and estimation results of costs

As mentioned in the discussion of cost structure in the previous section, there are two sources of cost for mobile network operators: the cost per customer and the cost per minute, which are denoted by C_{fm} and c_{fm}, respectively, for firm f in market m. We use the profit-maximization problem of mobile network operator f in market m to identify these two components of cost for each firm-market pair. This specification gives a linear approximation to firms' cost structure; it does not account for economies of scale.

As discussed in Subsection 3.4, to make the problem tractable, we simplify the pricing strategy of each mobile network operator in each market to three variables: the level of fixed fees LF_{fm}, the level of overage prices Lp_{fm}, and the level of allowances LA_{fm}. The first-order conditions of profit with respect to these levels provide optimization conditions for identifying the cost per customer and the cost per minute.

[8]The Economic Area Groupings are also known as Regional Economic Area Groupings for 220 MHz. Created by Federal Communications Commission staff, the economic area groupings are an aggregation of economic areas into 6 regions excluding the Gulf of Mexico.

The estimation results for costs are reported in Table 8. The results show that among the major operators, the mean cost per customer ranges from \$7.06 per month to \$11.51 per month, and the mean cost per minute ranges from \$0.14 per minute to \$0.20 per minute.

Discussion: cost-per-minute estimates As discussed in the industry model section, the marginal cost of one minute is a linear approximation of the cost structure imposed by the capacity constraint. Before hitting the network's capacity constraint, the marginal cost of one minute is low; once the capacity constraint is reached, there is a sudden jump in the marginal cost per minute. The effect of the capacity constraint is presented in the form of dropped calls in reality.

During the sample period, unlimited plans were uncommon, and most plans were associated with high overage prices. This pricing pattern was partially due to the high capacity constraint that mobile network operators were facing at that time. The cost per minute is estimated through the profit-maximization problem of mobile network operators with respect to overage prices; thus, the cost-per-minute estimates reflect the level of the capacity constraint that made the overage prices observed in the data optimal. As mobile network operators acquired more spectrum and lessened the capacity constraint with respect to voice traffic, they began to offer more unlimited plans in later years.

6 The effects of bill shock regulation

As discussed above, the lack of bunching of call minutes at the monthly allowance level is indicative of consumer uncertainty regarding price. We model consumers' perception error in recalling past usage as the source of such price uncertainty. Consistent with this modeling strategy, we now simulate the effects of the bill shock regulation on mobile network operators' profits and pricing decisions by running the following counterfactual.

Under the bill shock regulation, cell phone companies are required to alert consumers when their usage hit the allowance. Therefore, when the consumer does not receive the alert from her cell phone company, she at least knows that her actual usage is below the allowance even if she does not keep track of it precisely. Such knowledge helps her update the distribution of perception error and her expected utility from calling minutes. On the other hand, if she receives the alert when her perceived number of calling minutes is x_i, she realizes that her actual call minutes $x_i\omega_i$ is identical to the allowance A_j, and therefore she learns the realization of $\omega_i = A_j/x_i$.

Let $x_i^\dagger$ denote consumer i's optimal choice of perceived calling minutes under the bill shock regulation. In solving for $x_i^\dagger$, we consider two cases.

In the first case, suppose that $\theta_i > A_j$. In this case, the consumer has no incentive to stop calling until she receives the alert, because it is optimal for her to make calls up to θ_i minutes when there is no overage fee. But once she receives the alert, she learns the realization of ω_i and thus chooses perceived calling minutes $x_i^\dagger$ so that the actual calling minutes $x_i^\dagger\omega_i$ maximizes her utility.

$$x_i^\dagger(\omega_i) = \arg\max_{x_i \geq A_j/\omega_i} \theta_i \ln(x_i\omega_i) + x_{i0} + \alpha_i p_j(x_i\omega_i - A_j)$$
$$\text{subject to } x_i\omega_i + x_{i0} \leq T,$$

which can be solved as

$$x_i^\dagger(\omega_i) = \begin{cases} \frac{A_j}{\omega_i} & \text{if } \theta_i \in (A_j, (1+\alpha_i p_j)A_j], \\ \frac{\theta_i}{\omega_i(1+\alpha_i p_j)} & \text{if } \theta_i > (1+\alpha_i p_j)A_j. \end{cases}$$

In the second case, suppose that $\theta_i \leq A_j$. First note that in this case, the consumer has no incentive to increase her calling minutes when she receives the alert, because she then knows that her actual usage is identical to A_j. The consumer's utility when she receives the alert is

$$u_{A_j} = \theta_i \ln A_j + T - A_j.$$

On the other hand, when the consumer does not receive the alert, she learns that her perception error is in the range $(0, A_j/x_i)$. Therefore, her expected utility from making x_i minutes of perceived usage is

$$v_{na}(x_i) = \int_{\omega_i} \{\theta_i \ln(x_i\omega_i) + x_{i0}\}\, dF(\omega_i|\omega_i < A_j/x_i),$$

where $x_{i0} = T - \int_{\omega_i}(x_i\omega_i)dF(\omega_i|\omega_i < A_j/x_i)$ and $F(\omega_i|\omega_i < A_j/x_i)$ is the distribution function of ω_i conditional on $\omega_i < A_j/x_i$. The consumer, who has not received the alert at the perceived usage level x_i, will increase her calling minutes if and only if the expected utility from doing so exceeds her current utility $v_{na}(x_i)$. In other words, using $\rho^\Delta(x_i) = \text{Prob}\{\omega_i > A_j/(x_i+\Delta)|\omega_i < A_j/x_i\}$ to denote the probability of receiving the alert at usage level $x_i + \Delta$ conditional on $\omega_i < A_j/x_i$, the consumer will increase her calling minutes if and only if

$$v_{na}(x_i) < \rho^\Delta(x_i)u_{A_j} + (1-\rho^\Delta(x_i))v_{na}(x_i+\Delta). \quad (9)$$

As shown in Appendix C, by taking the limit of $\Delta \to 0$ in (9), we can obtain the desired necessary and sufficient condition as

$$v'_{na}(x_i) + (u_{A_j} - v_{na}(x_i))\frac{f(A_j/x_i)}{F(A_j/x_i)}\frac{A_j}{x_i^2} > 0. \quad (10)$$

Let $x_i^{\ddagger}$ be the perceived usage level that satisfies (10) with equality. Then the consumer's optimal choice of perceived calling minutes, when the realization of perception error is ω_i, can be written as

$$x_i^{\dagger}(\omega_i) = \begin{cases} x_i^{\ddagger} & \text{if } x_i^{\ddagger} < A_j/\omega_i, \\ A_j/\omega_i & \text{if } x_i^{\ddagger} \geq A_j/\omega_i. \end{cases}$$

Combining the above two cases ($\theta_i > A_j$ and $\theta_i \leq A_j$), consumer i's optimal choice of perceived calling minutes under the bill shock regulation can be succinctly summarized as

$$x_i^{\dagger}(\omega_i) = \begin{cases} \frac{1}{\omega_i}\max\{A_j, \frac{\theta_i}{1+\alpha_i p_j}\} & \text{if } \theta_i > A_j, \\ \min\{x_i^{\ddagger}, A_j/\omega_i\} & \text{if } \theta_i \leq A_j. \end{cases}$$

Corresponding to consumer i's perceived calling minutes $x_i^{\dagger}(\omega_i)$, her actual calling minutes is

$$q_i^{\dagger}(\omega_i) = x_i^{\dagger}(\omega_i)\omega_i = \begin{cases} \max\{A_j, \frac{\theta_i}{1+\alpha_i p_j}\} & \text{if } \theta_i > A_j, \\ \min\{x_i^{\ddagger}\omega_i, A_j\} & \text{if } \theta_i \leq A_j. \end{cases}$$

Before proceeding to the counterfactuals, we assess model fit by comparing the model simulated prices without regulation to the observed prices. Table 9 reports the ratios of the model simulated levels of fixed fees, overage prices, and allowances to those of the observed pricing structures of the operators. The table shows that there is a good fit between the model simulated prices and the observed prices, as the ratios are close to 1.

In the first counterfactual below, we show what would happen to consumer surplus and mobile network operators profit when the pricing structures of calling plans, as predicted by our model, remain unchanged following the regulation. This counterfactual is not realistic in terms of predicting actual changes, but serves as a useful benchmark for comparison purposes. In the second counterfactual, we allow mobile network operators to adjust their pricing structures in response to the regulation. We solve for the operators' pricing strategies in the new equilibrium and examine how the regulation and the new pricing structures would affect consumer surplus and mobile network operators profit.

6.1 The effects with unchanged pricing structures

Table 10 reports the monthly per-household and total effect of the bill shock regulation, keeping the simulated pricing structures unchanged. When consumers receive information on when their usage hits the allowance and no longer have uncertainty regarding marginal price, the probability of making overage payments decreases from 0.21 to 0.09, and 46% of subscribing consumers use exactly their allowance. The total number of monthly calling minutes goes up from 4396 million to 5557 million, while the total number of monthly overage minutes goes down from 249 million to 189 million.

Consumers benefit from more calling minutes and lower overage payments, and monthly total consumer surplus increases by $53 million. Monthly total operators profit, on the other hand, decreases by $117 million mainly due to higher costs associated with the larger number of calling minutes and lower overage payments. Overall, monthly total surplus decreases by $64 million, indicating that the increase in costs due to the larger number of calling minutes dominates the increase in consumers' utility from those additional calling minutes.

Here's some intuition regarding the decrease in total surplus. Without bill shock regulation, many consumers use much less than their free allowance due to fear of bill shock. The regulation alleviates such fear and results in many consumers using their entire allowance—assumed to be unchanged in this counterfactual analysis. Because of diminishing marginal utility of calling minutes for individual households[9], consumers' benefits from these additional calling minutes are lower than the firms' costs of providing them, leading to a reduction in total surplus. If consumers' under-consumption before the regulation is instead driven by, for example, shocks to month-to-month communication needs, then the regulation may be less likely to increase consumers' calling minutes to the allowance level and the welfare effect of the regulation may be less negative.

6.2 The effects with price responses

To compute the new price equilibrium under the bill shock regulation, we let each mobile network operator f in each market m readjust the model-predicted level of fixed fees LF_{fm}, level of overage prices Lp_{fm}, and level of allowances LA_{fm} for all of its plans, as discussed in Subsection 3.4. The new price equilibrium is the Nash equilibrium of the firms' pricing strategies in the counterfactual case.

Table 11 shows the changes in the major operators' pricing structures in the new equilibrium. This table compares simulated prices under bill shock regulation to simulated prices without bill shock regulation. All major operators decrease the allowances, the fixed fees, and the overage prices in response to the regulation. The decrease in fixed fees is proportionally less than the decrease in allowances, which means the per-minute price for the free minutes in the plan increases.

[9]In our model, diminishing marginal utility of calling minutes for individual households is reflected in the log specification of the calling utility function in equation 1.

When consumers' usage and price uncertainty is eliminated, the operators can no longer profit from consumers' "accidental" overage minutes (the source of bill shock). In response, they make the free minutes more expensive while making the overage minutes less expensive, thereby reducing the gap between the two prices. Those changes, combined with the reduction in monthly allowances, induce more consumers to knowingly and optimally incur overage calling minutes, as described below.

Note that the free allowances chosen by the operators are positive both before and after the bill shock regulation, indicating that the operators optimally choose three-part tariffs when two-part tariffs are an option (by setting the allowances to zero). The literature on two-part tariffs and three-part tariffs shows that there are multiple potential sources for firms' choice of three-part tariffs over two-part tariffs. For example, in Grubb (2009) the optimality of three-part tariffs over two-part tariffs arises from biased beliefs (which we don't model in this paper as we assume rational expectations), while Bagh and Bhargava (2013) show that three-part tariffs can be more efficient than two-part tariffs as price-discriminating mechanisms in the presence of consumer preference heterogeneity. More recently, Fibich et al. (2017) derive optimal three-part tariff plans under quite general conditions, including rational consumers, a general utility function, homogeneous or heterogeneous consumers, etc.

Table 12 reports the effects of the bill shock regulation when we take into account the operators' pricing responses to the regulation, using simulated prices for both "before regulation" and "after regulation" for an apples to apples comparison. Following the elimination of usage and price uncertainty and the pricing changes implemented by the operators, on the extensive margin, the number of households who subscribe to mobile phone plans sees a significant increase of 8 million (19.5%) from 41 million to 49 million. On the intensive margin, the average number of monthly calling minutes per subscribing household drops from 106.4 to 88.3, while the average number of monthly overage minutes per subscribing household increases substantially from 6 to 37.4. Overall, the total number of monthly calling minutes is largely unchanged, dropping slightly by 69 million (1.57%) from 4396 million to 4327 million.

Following the regulation and the resulting pricing changes, the consumers incur lower fixed fee payments but higher overage payments. In addition, due to diminishing marginal utility of calling minutes for individual households, the fact that roughly the same number of calling minutes are now spread over a larger number of households, with each household on average having fewer calling minutes, results in an increase in the total utility from those calling minutes. The combined effect of the above factors is an increase of \$132 million in monthly total consumer surplus from \$2008 million to \$2140 million.

The operators also benefit from the regulation and the resulting pricing changes, with monthly total operators profit increasing by $175 million from $248 million to $423 million, mainly because the increase in total overage payments dominates the reduction in total fixed fee payments. Overall, monthly total surplus increases by $307 million from $2256 million to $2563 million. These results are in contrast to those reported in Table 10 for the case with unchanged pricing structures, which shows a decrease in total operators profit and a decrease in total surplus following the regulation. One key difference between the two cases is that with unchanged pricing structures (Table 10), consumers make a significantly larger number of calling minutes following the regulation, and the resulting increase in the operators' costs drives total operators profit and ultimately total surplus down. When we incorporate the operators' pricing changes in response to the regulation (Table12), those pricing changes result in the consumers making roughly the same number of calling minutes post-regulation (although the divide between free minutes and overage minutes changes), thereby avoiding the increase in costs. The contrast between the results from the two different cases highlights the importance of accounting for firms' strategic responses when evaluating the effects of a regulation.

Here's some intuition for the above results. Without the bill shock regulation, consumers face usage and price uncertainty, and the mobile network operators design their plans in a way to exploit such uncertainty and extract surplus at consumers' expense. With the bill shock regulation, consumers' usage and price uncertainty is eliminated. If the operators keep their pricing structures unchanged, they would incur a drop in their profit levels, as there is no longer usage and price uncertainty on consumers' part to be exploited and the original pricing structures would backfire. When they are free to respond to the regulation, we find that the operators would optimally lower the fixed fees, the allowances, and the overage prices. More consumers would subscribe to mobile phone plans but on average monthly calling minutes would fall. The operators would earn more revenue and incur a lower cost per customer. With the pricing responses, both the operators and the consumers would benefit from the regulation. The overall welfare effect is a sizable increase in total surplus, with the welfare gain roughly evenly divided between the consumers and the operators.

Biased beliefs In this paper we have assumed rational expectations, i.e. the expectation of the perception error is 1. Since we are unable to estimate biases in beliefs in the empirical model due to data limitations, we perform two additional counterfactual exercises to explore whether potential biases in beliefs would qualitatively change the results. In the first exercise, we assess the effects of the bill shock regulation with adjusted prices when the expectation of the perception error is 1.1 (so that on average consumers underestimate their usage), and in the second exercise we assess the effects when the expectation of the perception error is 0.9 (so that on average consumers overestimate their usage).[10] The results of the two counterfactual exercises are given in Tables 13 and 14, respectively. These tables show that when we introduce moderate levels of biases in beliefs, the effects of the regulation remain qualitatively similar to those with unbiased beliefs, even as the magnitudes of the effects vary. For instance, same as in the case with unbiased beliefs, the regulation increases consumer surplus, operators profit, and overall surplus both when the expectation of the perception error is 1.1 and when it is 0.9. The results in these two tables indicate that our main findings are robust to moderate levels of biases in beliefs.

Real-world outcomes following bill shock regulation This paper has tried to predict what would have happened had the bill shock agreement been implemented in 2001, not 2013 when it was actually implemented. We now briefly look back to see what actually happened following the bill shock agreement implementation in April 2013 and compare the changes to our model predictions, recognizing that the wireless industry went through massive changes between 2001 and 2013 and so this comparison should be taken with a grain of salt.

Table 15 reports some industry statistics before and after the actual implementation of the bill shock agreement. According to the Pew Research Internet Project[11], the mobile phone penetration rate in the U.S. increased following the regulation (from 87% in December 2012 to 91% in December 2013), in line with our model prediction (keeping in mind that the industry was much closer to saturation in 2013 than in 2001 and that other factors such as a time trend may also be a reason for the actual changes we see). Additionally, according to CTIA (a trade association representing the wireless communications industry in the U.S.)[12], the total number of wireless subscriptions in the U.S. grew by 3% from 326.5 million at the end of 2012 to 335.7 million at the end of 2013. The average monthly revenue per unit (ARPU) remained roughly the same during that period, with \$48.99 in December 2012 and \$48.79 in December 2013. Different from our model prediction, the average monthly calling minutes increased from 587 minutes in December 2012 to 649 minutes in December 2013.

We also look at the profitability of the wireless industry before and after the implementation of the agreement. In the absence of data on economic profits, we instead consider accounting profits using EBITDA (Earnings before Interest, Taxes, Debt, and Amortization). In line with our model prediction, the Federal Communications Commission (FCC)[13] found that EBITDA per subscriber per month increased between 2012 and the first half of 2014 for each of the top three wireless service

[10] Recall that $q_{ij} = x_{ij}\omega_i$, where q_{ij} is the actual usage, x_{ij} the perceived usage, and ω_i the perception error.
[11] https://www.pewresearch.org/internet/fact-sheet/mobile/#mobile-phone-ownership-over-time.
[12] CTIA Wireless Industry Indices, Year-End 2013.
[13] 17th Annual Mobile Wireless Competition Report, at https://www.fcc.gov/17th-annual-mobile-wireless-competition-report.

providers: from \$22.21 to \$24.19 for Verizon, from \$18.64 to \$19.67 for AT&T, and from \$6.11 to \$11.13 for Sprint. The fourth and fifth providers saw decreases in their EBITDA, though: from \$12.09 to \$8.64 for T-Mobile and from \$11.51 to \$6.13 for US Cellular.

7 Conclusion

In this paper, we tackle the substantive question of what will happen to consumer welfare and firm profits when a consumer protection policy is enacted to provide consumers with more information. In the context of the mobile telecommunication industry, we assess the effects of a bill shock regulation that would give consumers accurate information about their actual usage. Our counterfactual analysis allows firms to adjust their pricing structures in response to the regulation and recomputes the new pricing equilibrium. We find that both consumers and operators would be better off in the new equilibrium. Overall, the regulation would lead to a sizable increase in total surplus which would be divided roughly evenly between consumers and operators.

Academic researchers can benefit from the modelling approach used in this paper. Understanding the effects of consumer protection policies on firms' strategies is not an easy task. The challenge lies in the fact that these policies are often proposed because consumers' behavior deviates from the standard economic model with fully informed and rational agents. A model that attempts to incorporate the source of consumers' limited information and bounded rationality runs the risk of deviating so far from the standard economic model that it is not estimable and cannot be used to make reasonable predictions on firms' strategies in response to the regulation. We overcome that challenge by finding the right degree of deviation in the modeling approach: in our model, consumers make judgment mistakes, but they know that they make these mistakes and take this factor into account in their utility-maximization problem.

We also reiterate the caveat that we have had to make some simplifying assumptions in order to make the model tractable, and therefore the results should be viewed as an approximation and should be interpreted with some caution. Nevertheless, one general policy implication arising from these results is the importance of accounting for firms' strategic responses to a consumer protection regulation, which are shown to have a big influence on the market outcome, including the welfare effects.

Future research could extend the model proposed in this paper to study the effect of consumer protection policies in other industries and settings. It would be interesting to see how firms' strategies will respond differently in other settings and what the implications of such differences would be for consumer welfare and industry profit.

References

Baek, J. and J. Brueckner (2015). Three-part tariffs with heterogeneous users: monopoly and duopoly cases. *Review of Industrial Organization 47*(2), 155–165.

Bagh, A. and H. K. Bhargava (2013). How to price discriminate when tariff size matters. *Marketing Science 32*(1), 111–126.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: the new car market. *Journal of Political Economy 112*(1), 68–105.

Danaher, P. J. (2002). Optimal pricing of new subscription services: Analysis of a market experiment. *Marketing Science 21*(2), 119–138.

Essegaier, S., S. Gupta, and Z. J. Zhang (2002). Pricing access services. *Marketing Science 21*(2), 139–159.

Feikert-Ahalt, C. (2019). United kingdom: Two banks found in breach of overdraft alert requirement, must refund customers. Library of Congress web page. https://www.loc.gov/item/global-legal-monitor/2019-12-20/united-kingdom-two-banks-found-in-breach-of-overdraft-alert-requirement-must-refund-customers/.

Fibich, G., R. Klein, O. Koenigsberg, and E. Muller (2017). Optimal three-part tariff plans. *Operations Research 65*, 1177–1189.

Freking, K. and R. Alonso-Zaldivar (2019). Trump calls on congress to end 'surprise medical bills'. Associated Press News, accessed May 29, 2019, https://apnews.com/bd3dfd133dc349bcb76f346fcdb29fda.

Grubb, M. D. (2009). Selling to overconfident consumers. *The American Economic Review 99*(5), 1770–1807.

Grubb, M. D. (2015a). Consumer inattention and bill-shock regulation. *Review of Economic Studies 82*, 219–257.

Grubb, M. D. (2015b). Overconfident consumers in the marketplace. *Journal of Economic Perspectives 29*(4), 9–36.

Grubb, M. D. and M. Osborne (2015). Cellular service demand: Biased beliefs, learning and bill shock. *American Economic Review 105*, 234–271.

Hausman, C. (2019). Shock value: Bill smoothing and energy price pass-through. *Journal of Industrial Economics 67*(2), 242–278.

Hausman, J. (1997). *The Economics of New Goods*, Chapter Valuation of new goods under perfect and imperfect competition. University of Chicago Press.

Iyengar, R., A. Ansari, and S. Gupta (2007). A model of consumer learning for service quality and usage. *Journal of Marketing Research 44*(4), 529–544.

Kumar, N. and R. Rao (2006). Research note—using basket composition data for intelligent supermarket pricing. *Marketing Science 25*(2), 188–199.

Lambrecht, A., K. Seim, and B. Skiera (2007). Does uncertainty matter? consumer behavior under three-part tariffs. *Marketing Science 26*, 698–710.

Lambrecht, A. and B. Skiera (2006). Paying too much and being happy about it: Existence, causes and consequences of tariff-choice biases. *Journal of Marketing Research 43*, 212–223.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review 103*(7), 2960–3000.

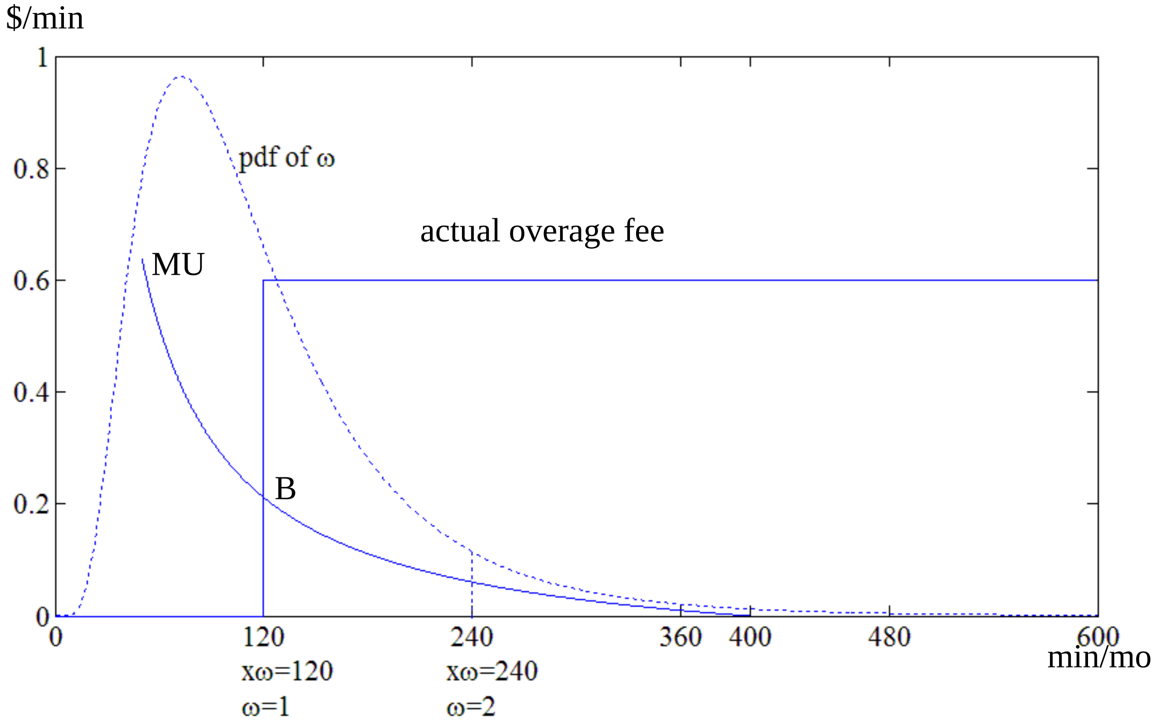
Littlechild, S. (2006). Mobile termination charges: calling party pays versus receiving party pays. *Telecommunications Policy 30*(5-6), 242–277.

McGowan, F. (2018). The roaming regulation and the case for applying behavioural industrial organisation to EU competition policy. ESRI Working Paper No. 598.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

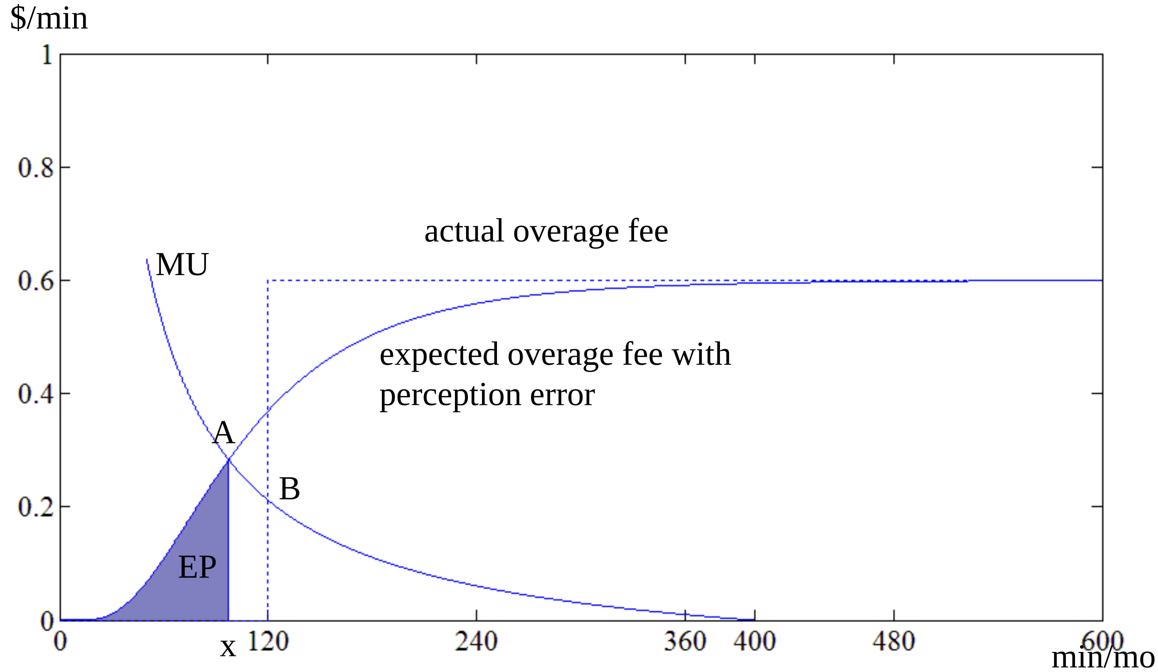
Pressman, A. (2019). The mobile price wars are on. Here's how much you can save. *Fortune*, accessed January 24, 2020, https://fortune.com/2019/10/31/att-verizon-t-mobile-unlimited-price-cut-cheapest-cell-phone-plans/.

Figure 1-a: Demonstration of the perception error



This graph demonstrates the existence of perception error of one particular consumer under one particular plan. This consumer's calling preference is represented by the curve MU (marginal utility of calling). If the marginal price of calling is zero for the whole month (unlimited plan), this consumer would like to stop calling at 400 minutes. However, the actual overage price that this consumer faces is 0 when she calls fewer than 120 minutes and jumps to \$0.60/min when she calls more than 120 minutes. This consumer has uncertainty about her actual usage $q = x\omega$: At any perceived usage x, she never knows what the exact realization of her perception error $\omega = \frac{q}{x}$ is (it could be the case that $\omega = 1$, which means that she has the correct perception of her actual usage; it could also be the case that $\omega = 2$, which means that even though her perceived usage is $x = 120$, her actual usage is $x\omega = 240$.).

Figure 1-b: Demonstration of the impact of the perception error on consumer's calling decision and overage payment



In this graph, the real marginal price per minute is 0 when the monthly calling minutes are fewer than 120 minutes and then jumps to \$0.60/min. If the consumer is perfectly certain about her usage, she will stop calling at exactly 120 minutes in this example. With the perception error, the perceived usage is different from the actual usage: At any perceived usage, there is a chance that the actual usage passes the 120 minutes threshold and the \$0.60/min overage price applies; taking this fact into account, the perceived marginal price per minute is positive at any perceived usage. This consumer will stop calling at point x. The expected overage payment is represented by the area EP.

Table 1: Summary statistics of the perception error from field study

	Number of obs	Mean	Std. dev.	Min	Max
Voice	86	0.96	0.83	0	3.8
Text	82	1.17	1.10	0	5
Data	75	0.91	0.69	0	3.3

Table 1 presents summary statistics of respondents' perception error on their voice usage, text usage and data usage.

Table 2: Count of bills and market share of major mobile network operators

Bill count in bill data			
Mobile network operator	Bill count	Percentage	Market shares (Kagan)
Sprint	1,327	8%	9%
AT&T	2,218	13%	12%
Verizon	3,540	21%	25%
Cingular	1,636	10%	19%
Others	8,434	48%	35%
Total	17,155	100%	100%
Bill count in call detail data			
Mobile network operator	Bill count	Percentage	Market shares (Kagan)
Sprint	1,100	10%	9%
AT&T	1,618	15%	12%
Verizon	2,430	22%	25%
Cingular	856	8%	19%
Others	5,047	45%	35%
Total	11,051	100%	100%

Table 3: Summary statistics of bill data and call detail data

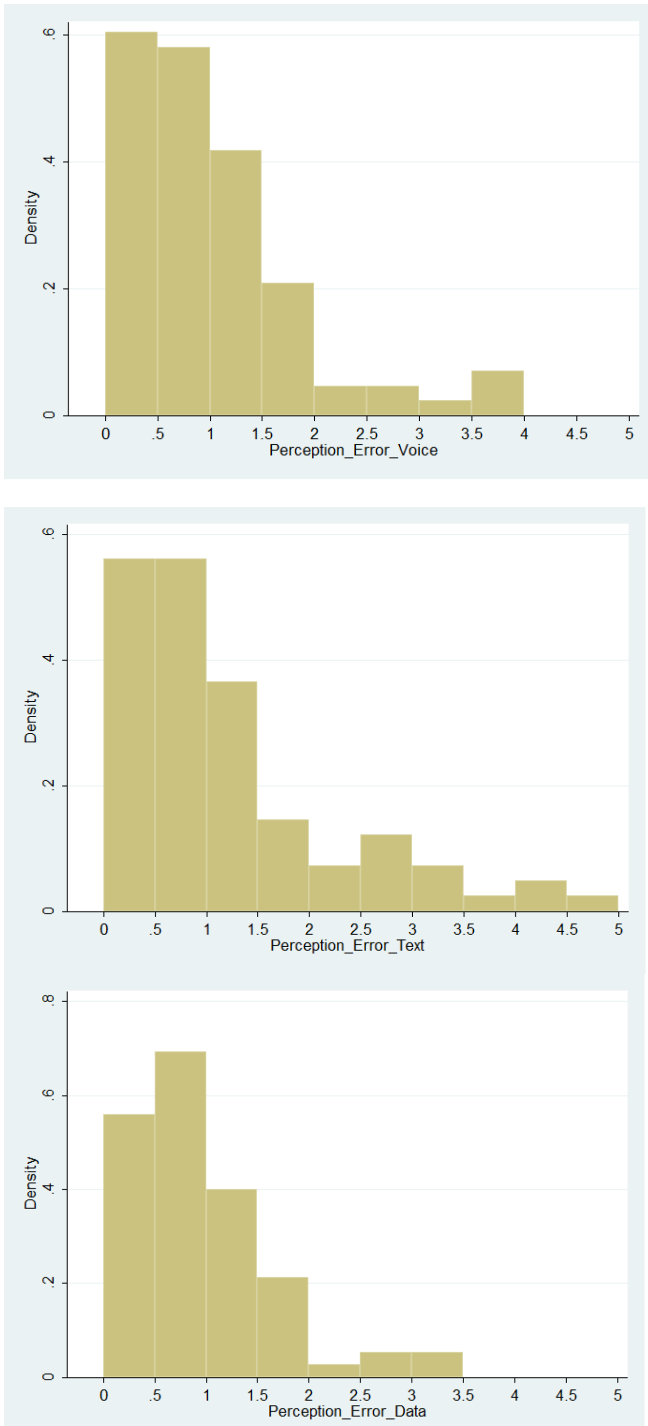
Bill data					
Variable	Number of bills	Mean	Std. dev.	Min	Max
fixed fee $/mo	16,894	32.20	22.80	0.00	349.98
free min used/mo	16,986	115.84	269.91	0.00	5473
billed min/mo	16,986	36.67	148.03	0.00	3256
total min used/mo	16,986	154.29	329.03	0	5715
Call detail data: billed calls					
Bills with billed calls	Number of bills	Percent	Average payment	Max payment	
roaming only	1,359	22%	6.58 $/mo	521.14 $/mo	
long distance only	2,681	44%	4.57 $/mo	165.69 $/mo	
roaming and long distance	1,057	17%	9.02 $/mo	287.14 $/mo	
non-roaming overage	3,495	57%	20.06 $/mo	438.90 $/mo	
bills with billed calls	6,100	100%	16.53 $/mo	620.45 $/mo	
Billed calls	Number of calls	Percent	Average duration	Average payment	
roaming only	6,591	7%	2.58 min	1.36 $/call	
long distance only	13,217	11%	3.28 min	0.93 $/call	
roaming and long distance	4,098	4%	2.83 min	2.33 $/call	
non-roaming overage	87,242	77%	2.66 min	0.80 $/call	
total billed calls	111,148	100%	2.74 min	0.91 $/call	

Fixed fee is the monthly fixed fee of the three-part tariff plan. Free minutes used indicates total number of minutes used during the month for which there is no charge (either free minutes included in the allowance or free off-peak minutes). Billed minutes are minutes that are charged a strictly positive price (either minutes outside of the allowance, or roaming/long distance calls that are usually not free for local and regional plans). Total minutes used are the sum of free minutes used and billed minutes.

Table 4: Summary statistics in the estimation sample

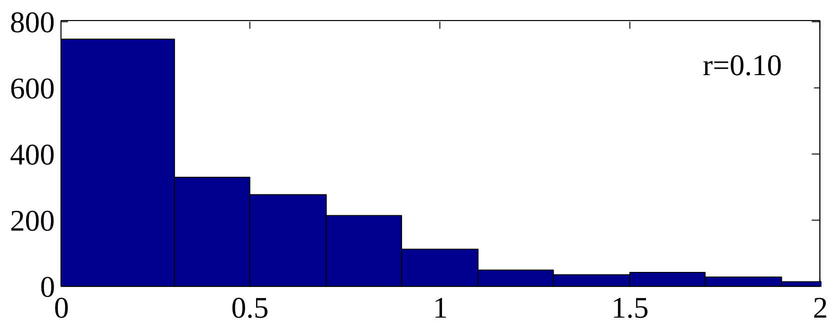
Aggregate data: market Level					
Variable	Number of markets	Mean	Std. dev.	Min	Max
number of providers per market	26	4.19	0.69	3	5
market share of the biggest provider	26	0.26	0.06	0.16	0.39
market share of the smallest provider	26	0.05	0.02	0.002	0.10
average number of plans per provider	26	5.22	1.58	2.40	8.75
Aggregate data: plan Level					
Variable	Number of plans	Mean	Std. dev.	Min	Max
fixed fee ($/mo)	577	40	17	15	128
allowance (min/mo)	577	311	238	0	1400
overage price ($/min)	577	0.35	0.07	0.18	0.65
free long distance calls	577	0.48	0.50	0.00	1.00
free roaming calls	577	0.27	0.44	0.00	1.00
market share	577	3%	2%	0%	16%
Micro data					
Variable	Number of hh	Mean	Std. dev.	Min	Max
monthly calling minutes (min/mo)	1,987	185	298	0	3169
family	1,987	0.77	0.42	0.00	1.00
head age over 55	1,987	0.35	0.48	0.00	1.00
renting	1,987	0.21	0.41	0.00	1.00
high income	1,987	0.85	0.36	0.00	1.00
prob. of incurring overage charges	1,987	17%	38%	0%	100%
overage charges ($/mo)	1,987	17	62	0	729

Figure 2: Histogram of perception error of voice, text, and data usage from field study



This graph shows the distribution of respondents' perception error on their voice (top panel), text (middle panel), and data (bottom panel) usage.

Figure 3: Histogram of the usage ratio in the data



This graph shows the histogram of the usage ratio in the data. Usage ratio is defined as the ratio of monthly calling minutes over total number of free minutes included in the plan. r is the radius of interval used for the histogram: the distance between the boundary of the interval and the center of the interval. In this histogram, r=0.10: the histogram of the usage ratio around 1 is approximated by the number of observations with usage ratio between 1-0.10 and 1+0.10.

Table 5: Estimates of parameters in the model

	Symbol in the paper	Parameter estimate	Standard error	Interpretation
Logit standard error	σ_ϵ	87.48	36.36	
Mean preference parameter	$\bar{\theta}$	5.00	0.37	
Standard deviation of unobservable heteogeneity	σ	0.98	0.12	
Standard deviation of log of perception error for high income consumers	σ_{ω_h}	0.27	0.05	
Standard deviation of log of perception error for low income consumers	σ_{ω_l}	0.22	0.10	
Price coefficient	α	-16.15	16.84	
Income effect	α_D	2.97	11.46	
Preference shifter: family dummy	θ_D	0.04	0.15	
Preference shifter: hh head age over 55 dummy	θ_D	-0.05	0.14	
Preference shifter: renting dummy	θ_D	0.01	0.27	
Year dummy: year 2000	λ	-3.99	3.26	compared with year 2001
AT&T dummy	λ	-12.66	5.25	compared with Verizon
Cingular dummy	λ	-10.71	4.10	compared with Verizon
Sprint dummy	λ	-11.84	4.34	compared with Verizon
Other carrier dummy	λ	-13.77	4.90	compared with Verizon
Free long distance dummy	λ	-7.89	2.25	
Free roaming dummy	λ	1.13	0.80	

Table 6: Comparison of the moments from model simulations and from data

	Prob. of usage ratio is between 0.90 to 1.10	
	High income consumers	Low income consumers
Data	0.06	0.08
Model simulation: variance of omega as estimated	0.06	0.08
Model simulation: variance of omega=0	0.49	0.53
Model simulation: variance of omega 1/4 as estimated	0.26	0.31
Model simulation: variance of omega 1/2 as estimated	0.18	0.22

This table compares the probability of the usage ratio being between 0.90 and 1.10 (the probability that consumers' actual usage level is between 90% and 110% of the number of free minutes included in the chosen plan) in the data and the same moment simulated from the model by setting the key parameter σ_ω at different levels for high income and low income consumers.

Table 7: Estimates for price elasticities

	Sprint	AT&T	Verizon	Cingular
Number of markets	20	26	20	12
Mean own price elasticities wrt fixed fee	-4.55	-4.19	-3.11	-4.02
Median own price elasticities wrt fixed fee	-4.37	-4.11	-3.09	-3.93
Mean own price elasticities wrt overage price	-1.93	-1.85	-2.01	-1.99
Median own price elasticities wrt overage price	-1.90	-1.88	-1.95	-1.92

Number of markets is the number of markets where a particular firm is present. The subscription price elasticity is defined as the percentage change in the penetration rate (the percentage of the population with mobile phones) in one market with a one percentage point increase in fixed fees of all plans in the market. A particular firm's price elasiticity in each market is computed as the percentage change in the firm's market share when the fixed fees of all plans offered by this firm in this market are increased by 1 percent, while holding prices of other firms constant. For example, Sprint was operating in 20 out of 26 markets in the estimation sample; if Sprint increases the fixed fee of all plans in one market by one percent while holding prices of other firms in the same market constant, the market share of Sprint would, on average (average across markets where this firm is present), decrease by 4.55 percent.

Table 8: Estimates for costs

	Sprint	AT&T	Verizon	Cingular
Number of markets	20	26	20	12
Mean cost per customer \$/mo	11.51	8.71	10.74	7.06
Std. dev. \$/mo	11.01	12.45	14.44	10.32
Mean cost per minute \$/min	0.14	0.17	0.14	0.20
Std. dev. \$/min	0.08	0.09	0.11	0.09
Industry estimate cost per minute \$/min*	0.11			

The 2nd and 3rd rows of table 8 present the mean and standard deviation of cost per consumer of major mobile network operators across markets they served in the estimation sample. For example, Sprint served in 20 out of 26 markets in the estimation sample. The mean cost per consumer across these 20 markets is \$11.51, and the standard deviation of cost per consumer across these 20 markets is \$11.01. The 4th and 5th row of table 8 present the mean and standard deviation of cost per minute of major mobile network operators across markets they served in the estimation sample. For example, Sprint served in 20 out of 26 markets in the estimation sample. The mean cost per minute across these 20 markets is \$0.14, and the standard deviation of cost per minute across these 20 markets is \$0.08. The last row of table 8 shows the costs per minute Sprint reported to FCC in the year 2003 in order to obtain a ruling from the FCC that it was entitled to seek reciprocal compensation based on its own wireless network's traffic-sensitive costs rather than the wireline carrier's costs. (Please refer to table 2 in Littlechild (2006).)

Table 9: Ratios of the model simulated price levels to the observed price levels, without bill shock regulation

Operator	Number of markets	Mean fixed fees level	Mean overage prices level	Mean allowances level
Sprint	20	1.03	0.97	1.02
AT&T	26	0.94	1.03	0.98
Verizon	20	0.97	0.93	1.06
Cingular	12	0.94	0.99	0.99

Table 10: The effect of the bill shock regulation assuming no price adjustments

Monthly per-household effect	Before regulation	After regulation	Change
Non-welfare outcomes			
Penetration of mobile	52%	56%	4%
Mean monthly calling minutes (cond. on subscription) (min/mo)	106.4	123.5	17.1
Mean monthly overage minutes (cond. on subscription) (min/mo)	6	4.2	-1.8
Prob. of using exactly free minutes in the plan (cond. on subscription)	0.00	0.46	0.46
Prob. of making overage payment (cond. on subscription)	0.21	0.09	-0.12
Mean monthly overage payment (cond. on subscription)	\$2.1	\$1.5	-\$0.6
Welfare Outcomes			
Mean consumer surplus	\$25.1	\$25.8	\$0.7
Mobile operators profit	\$3.1	\$1.7	-\$1.4
Total surplus	\$28.2	\$27.5	-\$0.7
Monthly total effect			
Non-welfare outcomes			
Number of hh with mobile	41 million	45 million	4 million
Total monthly calling minutes	4396 million	5557 million	1161 million
Total monthly overage minutes	249 million	189 million	-60 million
Number of hh making overage payment	8.6 million	3.9 million	-4.7 million
Monthly overage payment	\$87.1 million	\$65.2 million	-\$21.9 million
Welfare outcomes			
Total consumers surplus	\$2008 million	\$2061 million	\$53 million
Total mobile operators profit	\$248 million	\$131 million	-\$117 million
Total surplus	\$2256 million	\$2192 million	-\$64 million

Table 10 presents the monthly per-household and total effect of the bill shock regulation, keeping the simulated pricing structures unchanged. We first compute the monthly per-household effect in each market (defined as economic-year pair), then multiply the per-household effect in each market by the number of households reported in Census in the year 2000. The numbers reported are the sums of the total effects of the 26 markets in the estimation sample. The per household effect reported in table 10 is computed by dividing the total effect reported in the table by the total number of households in 26 markets in the estimation sample. The monetary values reported in Table 10 are in year 2000 dollars.

Table 11: Changes in fixed fees, overage prices and allowances after the bill shock regulation (comparing simulated prices to simulated prices)

Mean changes

	Number of markets	Mean change in fixed fees	Mean change in overage prices	Mean change in allowances
Sprint	20	-24.00%	-50.75%	-53.50%
AT&T	26	-30.77%	-37.69%	-66.15%
Verizon	20	-11.50%	-48.50%	-23.50%
Cingular	12	-31.25%	-40.00%	-69.17%

Overall levels before and after the regulation

	Before regulation	After regulation	Change
Fixed fees ($/month)	40	30	-24.58%
Overage prices ($/minute)	0.35	0.18	-47.52%
Allowances (min/month)	311	147	-52.66%

Table 12: The effect of the bill shock regulation with adjusted prices

Monthly per-household effect	Before regulation	After regulation	Change
Non-welfare outcomes			
Penetration of mobile	52%	61%	9%
Mean monthly calling minutes (cond. on subscription) (min/mo)	106.4	88.3	-18.2
Mean monthly overage minutes (cond. on subscription) (min/mo)	6	37.4	31.4
Prob. of using exactly free minutes in the plan (cond. on subscription)	0.00	0.34	0.34
Prob. of making overage payment (cond. on subscription)	0.21	0.51	0.30
Mean monthly overage payment (cond. on subscription)	$2.1	$6.4	$4.2
Welfare Outcomes			
Mean consumer surplus	$25.1	$26.8	$1.7
Mobile operators profit	$3.1	$5.3	$2.2
Total surplus	$28.2	$32.1	$3.9
Monthly total effect			
Non-welfare outcomes			
Number of hh with mobile	41 million	49 million	8 million
Total monthly calling minutes	4396 million	4327 million	-69 million
Total monthly overage minutes	249 million	2308 million	2059 million
Number of hh making overage payment	8.6 million	26.4 million	17.8 million
Monthly overage payment	$87.1 million	$383 million	$296 million
Welfare outcomes			
Total consumers surplus	$2008 million	$2140 million	$132 million
Total mobile operators profit	$248 million	$423 million	$175 million
Total surplus	$2256 million	$2563 million	$307 million

Table 13: The effect of the bill shock regulation with adjusted prices when the expectation of the perception error is 1.1

Monthly per-household effect	Before regulation	After regulation	Change
Non-welfare outcomes			
Penetration of mobile	54%	68%	14%
Mean monthly calling minutes (cond. on subscription) (min/mo)	117.8	87.5	-30.3
Mean monthly overage minutes (cond. on subscription) (min/mo)	7.8	39.96	32.2
Prob. of using exactly free minutes in the plan (cond. on subscription)	0.00	0.34	0.34
Prob. of making overage payment (cond. on subscription)	0.29	0.58	0.29
Mean monthly overage payment (cond. on subscription)	$2.8	$6.7	$3.9
Welfare Outcomes			
Mean consumer surplus	$24.1	$26.7	$2.6
Mobile operators profit	$2.1	$6.2	$4.1
Total surplus	$26.2	$32.9	$6.9
Monthly total effect			
Non-welfare outcomes			
Number of hh with mobile	43 million	54 million	11 million
Total monthly calling minutes	4830 million	4725 million	-105 million
Total monthly overage minutes	319 million	2158 million	1839 million
Number of hh making overage payment	11.5 million	32.8 million	21.3 million
Monthly overage payment	$114.8 million	$361.8 million	$247 million
Welfare outcomes			
Total consumers surplus	$1928 million	$2135 million	$207 million
Total mobile operators profit	$167 million	$496 million	$329 million
Total surplus	$2095 million	$2631 million	$536 million

Table 14: The effect of the bill shock regulation with adjusted prices when the expectation of the perception error is 0.9

Monthly per-household effect	Before regulation	After regulation	Change
Non-welfare outcomes			
Penetration of mobile	51%	59%	8%
Mean monthly calling minutes (cond. on subscription) (min/mo)	99.7	87.6	-12.1
Mean monthly overage minutes (cond. on subscription) (min/mo)	4.77	41.20	36.4
Prob. of using exactly free minutes in the plan (cond. on subscription)	0.00	0.33	0.33
Prob. of making overage payment (cond. on subscription)	0.17	0.51	0.34
Mean monthly overage payment (cond. on subscription)	$1.7	$6.8	$5.1
Welfare Outcomes			
Mean consumer surplus	$25.2	$26.2	$1.0
Mobile operators profit	$3.3	$4.8	$1.5
Total surplus	$28.5	$31.0	$2.5
Monthly total effect			
Non-welfare outcomes			
Number of hh with mobile	40 million	47 million	7 million
Total monthly calling minutes	4088 million	4117 million	29 million
Total monthly overage minutes	196 million	1936 million	1740 million
Number of hh making overage payment	7.1 million	27.5 million	20.4 million
Monthly overage payment	$70 million	$319 million	$261 million
Welfare outcomes			
Total consumers surplus	$2018 million	$2087 million	$79 million
Total mobile operators profit	$267 million	$386 million	$119 million
Total surplus	$2285 million	$2483 million	$198 million

Table 15: Industry statistics before and after the actual implementation of the bill shock agreement (April 2013)

	December 2012	December 2013	Change
Penetration of mobile[a]	87%	91%	4%
Total number of wireless subscriptions[b]	326.5 million	335.7 million	9.2 million
Average monthly revenue per unit (ARPU)[b]	$48.99	$48.79	$0.20
Average monthly calling minutes[b]	587 minutes	649 minutes	62 minutes
	2012	1st half 2014	Change
Verizon EBITDA per subscriber per month[c]	$22.21	$24.19	$1.98
AT&T EBITDA per subscriber per month[c]	$18.64	$19.67	$1.03
Sprint EBITDA per subscriber per month[c]	$6.11	$11.13	$5.02
T-Mobile EBITDA per subscriber per month[c]	$12.09	$8.64	-$3.45
US Cellular EBITDA per subscriber per month[c]	$11.51	$6.13	-$5.38

These statistics are collected from (a) Pew Research Internet Project, (b) CTIA Wireless Industry Indices, and (c) FCC 17th Annual Mobile Wireless Competition Report.

Appendix

A Details on Stage 3 of the empirical model: consumers' calling decision

A.1 Solution for the optimal perceived calling minute

The consumer's expected overage payment is given by $\int_{\omega_i} p_j \max\{x_i\omega_i - A_j, 0\}\, dF(\omega_i)$. ω_i follows a distribution with the probability density function $f(\omega_i)$ and the cumulative distribution function $F(\omega_i)$. Taking the derivative of the expected overage payment with respect to x_i, we can derive the expected marginal price of the next calling minute as follows.

$$
\begin{aligned}
&\frac{\partial}{\partial x_i}\left(\int_{\omega_i} p_j \max\{x_i\omega_i - A_j, 0\} dF(\omega_i)\right) \\
&= \frac{\partial}{\partial x_i}\left(p_j \int_{x_i\omega_i - A_j > 0} (x_i\omega_i - A_j) f(\omega_i) d\omega_i\right) \\
&= \frac{\partial}{\partial x_i}\left(p_j\left(\int_{\omega_i > \frac{A_j}{x_i}} x_i\omega_i f(\omega_i) d\omega_i - A_j\left(1 - F\left(\frac{A_j}{x_i}\right)\right)\right)\right) \\
&= p_j \int_{\omega_i > \frac{A_j}{x_i}} \omega_i f(\omega_i) d\omega_i \\
&= p_j \frac{\int_{\omega_i > \frac{A_j}{x_i}} \omega_i f(\omega_i) d\omega_i}{(1 - F(\frac{A_j}{x_i}))}\left(1 - F\left(\frac{A_j}{x_i}\right)\right) \\
&= p_j(E(\omega_i | x_{ij}\omega_i > A) prob(x_{ij}\omega_i > A))
\end{aligned}
$$

The consumer's optimal choice is to equate expected marginal utility to expected opportunity cost for the next calling minute, as illustrated in Figure 1-b. Formally, we have

$$
\underbrace{\frac{\theta_i}{x_{ij}}}_{\text{EMU for the next calling minute}} = \underbrace{\alpha_i p_j (E(\omega_i | x_{ij}\omega_i > A) prob(x_{ij}\omega_i > A)) + 1}_{\text{Expected opportunity cost of the next calling minute}} \qquad (11)
$$

The solution for the optimal perceived calling minute x^*_{ij} can be obtained by numerically solving equation 11.

A.2 The impact of perception error on consumers' decision under two-part tariff

If consumer i is under a two-part tariff plan with monthly fixed fee of F_j and per minute price p_j, with perception error, consumer i tries to choose a perceived number of calling minutes x_{ij} to maximize her expected utility conditional on the distribution of ω_i:

$$\max_{x_{ij}} v_{ij}(x_{ij}) = \int_{\omega_i} \overbrace{\theta_i ln(x_{ij}\omega_i)}^{\text{utility from calling}} + x_{i0} + \overbrace{\alpha_i p_j (x_{ij}\omega_i)}^{\text{disutility from payment}} dF(\omega_i) \tag{12}$$
$$\text{subject to} \quad x_{ij}\omega_i + \underbrace{x_{i0}}_{\text{outside activity}} \leq \underbrace{T}_{\text{time constraint}} .$$

Let x^*_{ij} be the optimal number of perceived minute. F.O.C. implies that

$$\frac{\theta_i}{x^*_{ij}} - (1 + \alpha_i p_j) \int_{\omega_i} \omega_i dF(\omega_i) = 0. \tag{13}$$

Because $\int_{\omega_i} \omega_i dF(\omega_i) = 1$, we obtain $x^*_{ij} = \frac{\theta_i}{1+\alpha_i p_j}$. The actual number of calling minutes is then equal to $x^*_{ij}\omega_i = \frac{\theta_i \omega_i}{1+\alpha_i p_j}$.

B Details on data

B.1 Details on bill data

Bill data can be uniquely matched with call detail data using quarter-household ID-bill number. The bill data and call detail data show the name of household's mobile operator. Table 2 shows the count of bills in bill data and call detail data by major mobile network operators in 2000-2001.

As a test of sample representativeness, the last column of Table 2 reports the aggregate market shares reported in Kagan's *Wireless Telecom Atlas & Databook* 2001 Volume 2. We find that the bill data and call detail data are roughly representative, with a few exceptions. The difference between the market share of Cingular in bill-level data and that reported by Kagan may be explained by the fact that Cingular was established only at the beginning of 2001, as a joint venture between SBC Communications and BellSouth; the bill-level data include only Cingular bills in the year 2001 (not in the year 2000), while Kagan reports the market share of SBC and BellShouth as that of Cingular in the year 2000.

Additionally, according to CTIA's semi-annual Industry Survey Results (CTIA Wireless Industry Indices, Year-End 1999), during June-December 1999 the average monthly calling minutes for a mobile phone subscriber in the U.S. was 185 minutes, which is in the same ballpark as the 154.29 minutes reported in Table 3 for the bill data.

Table 3 shows the summary statistics of key variables included in the bill data. The bill data have two shortcomings. First, billed minutes reported in the data do not distinguish minutes that are charged because of overage fees (minutes that are over the allowance) from roaming and long distance minutes that are charged in the form of linear pricing for local and regional plans. Second, 4,057 bills recorded zero usage, which is inconsistent with the call detail data. The call detail data overcome these shortcomings.

The 11,051 bills with call detail information report, in total, 748,391 calls. For each call, we can see the time of the call, whether it is roaming, what charges apply to this call, and what long distance charges apply to this call. We refer to 111,148 calls that were charged a strictly positive price outside of the allowance as billed calls. Table 3 also shows the composition, duration, and charges of billed calls. Non-roaming overage calls are calls that have been billed, but are neither roaming calls nor billed long distance calls.

Based on this information, we can overcome the shortcomings in the bill data discussed above.

We can compute how many billed minutes are due to overage fees (additional minutes that are over the allowance), how many are roaming minutes, and how many are long distance minutes. Finally, by adding together the duration of all calls placed, we get the total number of minutes used.

Table 3 shows the average and maximum monthly charges of bills with billed calls outside of the allowance. Among 11,051 bills with call detail information, 6,100 bills (around 55%) have billed calls. Among these 6,100 bills, 3,495 bills (around 57%) have billed calls due to non-roaming overage charges.

B.2 Details on data matching process

This section presents the process of matching the bill data with the tariff data. We define a market as an economic area-year pair.[14] We match bills with the tariff data using the market-operator-fixed fee listed in both sources and use 2,992 matched bills to construct the market share of plans.[15] This data matching process assumes that consumers chose the optimal plan from the choice set available in the same year; consumers who chose plans that were not available from the same year have been excluded from the estimation sample. Bills that belong to the four major operators account for close to 80 percent of matched bills in the sample. The four major operators are: Sprint, AT&T, Verizon, and Cingular. We aggregate all plans offered by operators other than the four major operators in each market. Among plans offered by the four major operators, 95 percent of matched bills have fixed fees in the \$19.99-\$59.99 range. For each major operator in each market, we aggregate plans with fixed fees higher than \$59.99 into one plan.

The matched 2,992 bills cover 108 markets: 46 economic areas in the year 2000 and 62 economic areas in the year 2001. 39 economic areas in the year 2000 have fewer than 30 matched bills, and 41 economic areas in the year 2001 have fewer than 30 matched bills. These 80 markets have too few matched bills to approximate the market shares of plans and, thus, are excluded from the estimation sample. Among the remaining 28 markets, we also exclude one market from 2000 and one market from 2001 where the majority of bills belong to non-major carriers and non-major plans.

In the data matching process, a small number of plans in the tariff data could not be matched with the billing data and are dropped from the estimation sample. Table 16 shows the number of matched plans as well as the total number of plans in the tariff data for each market. On average, each market has 24.3 plans, out of which 2.1 plans are dropped.

C Details on counterfactual

In this section, we will clarify the necessary and sufficient condition for the consumer, who has not received the alert, to increase her calling minutes. Suppose that $\theta_i \leq A_j$, and that the consumer has not received the alert at her current perceived usage level x_i. When she increases her calling

[14] The Economic Area service areas are based on the Economic Areas delineated by the Regional Economic Analysis Division, Bureau of Economic Analysis, U.S. Department of Commerce February 1995 (1-172), with the following additions: Guam and the Northern Mariana Islands (173), Puerto Rico and the U.S. Virgin Islands (174), and American Samoa (175). The Federal Communication Commission has also designated the Gulf of Mexico (176) as an additional Economic Area.

[15] We distinguish between plans with the same market-operator-fixed fee, but that differ in terms of whether they offer free roaming calls or long distance calls. We double check the accuracy of matching by looking at whether free roaming calls or long distance calls recorded in the bill are consistent with the corresponding characteristics of the plan. We also check whether the usage level recorded in the bill is consistent with the allowance level of the plan and drop the matches that are inconsistent.

Table 16: Number of matched plans in each market

Market	**No. of Plans Matched**	**No. of Plans in Tariff Data**	**Market**	**No. of Plans Matched**	**No. of Plans in Tariff Data**
1	23	24	14	22	24
2	22	25	15	13	16
3	35	38	16	25	25
4	38	41	17	19	21
5	18	20	18	14	15
6	40	44	19	12	15
7	10	12	20	13	15
8	21	23	21	16	19
9	17	19	22	35	37
10	24	25	23	30	33
11	20	20	24	13	16
12	31	33	25	13	15
13	29	32	26	24	25

minutes by $\Delta > 0$, the conditional probability of receiving the alert is

$$\begin{aligned}\rho^{\Delta}(x_i) &= \text{Prob}\{\omega_i > A_j/(x_i+\Delta) | \omega_i < A_j/x_i\} \\ &= 1 - F(A_j/(x_i+\Delta) | \omega_i < A_j/x_i) \\ &= 1 - \frac{F(A_j/(x_i+\Delta)}{F(A_j/x_i)}.\end{aligned}$$

The consumer will increase her calling minutes if and only if

$$\begin{aligned}&\rho^{\Delta}(x_i)u_{A_j} + (1-\rho^{\Delta}(x_i))v_{na}(x_i+\Delta) > v_{na}(x_i) \\ &\Leftrightarrow v_{na}(x_i+\Delta) - v_{na}(x_i) + (u_{A_j} - v_{na}(x_i+\Delta))\rho^{\Delta}(x_i) > 0 \\ &\Leftrightarrow \frac{v_{na}(x_i+\Delta) - v_{na}(x_i)}{\Delta} + (u_{A_j} - v_{na}(x_i+\Delta))\frac{\rho^{\Delta}(x_i)}{\Delta} > 0.\end{aligned}$$

By taking the limit of $\Delta \to 0$, we can obtain the desired necessary and sufficient condition as

$$v'_{na}(x_i) + (u_{A_j} - v_{na}(x_i))\frac{f(A_j/x_i)}{F(A_j/x_i)}\frac{A_j}{x_i^2} > 0$$

since

$$\begin{aligned}\lim_{\Delta\to 0}\frac{\rho^{\Delta}(x_i)}{\Delta} &= \frac{1}{F(A_j/x_i)}\lim_{\Delta\to 0}\frac{F(A_j/x_i) - F(A_j/(x_i+\Delta))}{\Delta} \\ &= \frac{1}{F(A_j/x_i)}\left(\frac{A_j}{x_i^2}f(A_j/x_i)\right).\end{aligned}$$