Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

An Empirical Analysis of Mobile Voice Service and SMS: A Structural Model

Youngsoo Kim, Rahul Telang, William B. Vogt, Ramayya Krishnan,

MANAGEMENT SCIENCE
Vol. 56, No. 2, February 2010, pp. 234–252
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5602 | 0234

DOI 10.1287/mnsc.1090.1091

An Empirical Analysis of Mobile Voice Service and SMS: A Structural Model

Youngsoo Kim
School of Information Systems, Singapore Management University, Singapore 178902, yskim@smu.edu.sg

Rahul Telang
H. J. Heinz III College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, rtelang@andrew.cmu.edu

William B. Vogt
RAND Corporation, Pittsburgh, Pennsylvania 15213, william.b.vogt@gmail.com

Ramayya Krishnan
H. J. Heinz III College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, rk2x@andrew.cmu.edu

In addition to the wireless telephony boom, a similar exponentially increasing trend in wireless data service—for example, short message service (SMS)—is visible as technology advances. We develop a structural model to examine user demand for voice service and SMS. Specifically, we measure the own- and cross-price elasticities of these services. The cross-price elasticity is of significant importance because marketing activities are critically influenced by whether the goods are substitutes or complements. The research context poses significant econometric challenges due to three-part tariffs and sequential discrete plan choice and continuous quantity choice decisions. Using detailed individual consumption data of more than 6,000 customers, we find that SMS and voice service are small substitutes. A 10% increase in the price of voice minutes will induce about a 0.8% increase in the demand for SMS. The own-price elasticity of voice is also low, to the order of approximately −0.1. Younger users' demand is far more inelastic than that of older users. We then conduct counterfactual policy experiments that fully capture the effects of changes in key parameters on the firm's revenues. Finally, we discuss the generalizability of our framework.

Key words: wireless communication; price elasticity; short message service (SMS); structural model; nonlinear tariff; substitutes versus complements; policy experiments
History: Received February 5, 2008; accepted September 2, 2009, by Wallace J. Hopp, information systems. Published online in *Articles in Advance* November 25, 2009.

1. Introduction

In most countries, mobile (cellular) telephones have grown to be a central part of the telecommunication network—mobile penetration rates exceed fixed line penetration rates in all regions of the world, and by a margin of three to four times in some regions (ITU World Telecommunication 2009). Along with the growth of mobile telephony, the use of wireless data services—specifically short message service (SMS)—has also grown exponentially. The growth of mobile services has been even more spectacular in some of the less developed countries in Africa, Asia, and Europe. Many users are giving up their traditional land lines and relying solely on their cellular phones as primary communication devices. A recent McKinsey study found that mobile services have generated billions of dollars in consumer welfare (Enriquez et al. 2007). Currently, revenues from mobile data services are about 15%–20% of average revenues per user (*Comms Dealer* 2008).[1] However, going forward, it is expected that the mobile data will become a key revenue source for mobile operators. Most cellular operators are betting on an increasing uptake of mobile data services and investing billions in infrastructure. Despite tremendous growth in mobile voice and data services, our understanding of how users consume these services is still limited. In particular, little work exists that examines the interaction between voice and data services. In this paper, using a rich panel data set, we estimate demand elasticities for these services with a particular focus on the cross-price effects of voice service and SMS.

This study has important implications for practitioners and academicians. Wireless communication tariffs tend to be highly nonlinear, being composed of a fixed fee component and a marginal fee

[1] Most of the nonvoice revenue is from SMS.

component. Although these price schemes are intuitive to describe, an empirical estimation of demand parameters under a nonlinear price structure is nontrivial. In particular, from an academic perspective, one has to carefully model user choices for both the service plan and then subsequent consumptions. For managers, understanding user response to a typical nonlinear pricing scheme is important because it allows them to optimally price their products. For example, a manager might want to know how changes in the fixed fee or the marginal price would impact her firm profitability. Similarly, the cross-price elasticity between two services has important implications for optimal pricing and promotion decisions. For example, if managers comprehend how voice and SMS interact (whether SMS is a substitute or a complement to voice), they can better predict the direct and indirect impact of pricing and promotion and set appropriate marketing strategies.

However, the impact of SMS on voice (and vice versa) is ambiguous. On the one hand, SMS can be a close substitute for voice if it is perceived to serve the same purpose. In that case, a price increase in voice may lead to increased usage of SMS (and vice versa). However, SMS and voice may serve possibly different purposes, and at best they are weak substitutes. Andersson et al. (2006) argue that SMS and voice may even exhibit a complementary relationship. For example, a price decrease in SMS may lead to more incoming messages for a user, who in turn may respond by increasing voice consumption. Given these conjectures, how voice and SMS interact is an open empirical question. However, except for a couple of recent papers, which we describe in detail in the next section, the literature examining this mutual relationship is sparse, presumably because of the lack of individual-level consumption data or the relatively recent nature of this phenomenon. Our study fills this gap in the literature. We have access to individual-level consumption data for these services, which allows us to build a richer model and draw useful conclusions.

In addition to the unique data set, our paper also makes important methodological contributions. As noted, the telecom tariffs (especially in the wireless industry) tend to be two-part (or three-part) nonlinear tariffs, which require careful modeling to avoid bias in demand estimations. In our case, we not only have a three-part tariff structure with consumers' sequential decisions (first choosing a plan and then choosing the consumption level), we also deal with a two-goods problem—voice and SMS. In this paper, we offer a structural model by specifying a utility function and deriving the demand curves for these services. This model allows us to jointly specify the sequential consumption decision at an individual user level for these goods, and allows us to estimate the own- and cross-price elasticities for these services. To our knowledge, this is the first study that offers a methodology to estimate the own- and cross-price elasticities for voice and SMS services based on detailed individual-level consumption data. Another advantage of the structural model is that it allows us to perform some policy experiments. Thus, with our estimates, we can simulate and specify how changes in key strategic variables (e.g., fixed fee, marginal price, etc.) would affect the firm demand and revenues (or profitability).

Our paper is composed of two parts. In the first part, we specify a utility structure and develop a general model of individual consumption behavior for voice and SMS. In the second part, we estimate our model using a real-world data set, explicitly allowing for cross-effects and endogeneity in the choice of plans and consumption levels. Our framework incorporates four crucial elements that could affect users' optimal consumption behavior: (1) the interrelationship of voice and SMS in demand through the cross-price elasticities, (2) estimation of the own-price elasticities, (3) the service-specific satiation points, and (4) the potential for endogeneity through covariance between the voice and SMS satiation points. The fourth point is critical. If users who consume large numbers of voice minutes are also likely to send large numbers of messages, the demand for the two services would be positively correlated *even if* the two services are not complementary.

The rest of this paper is organized as follows. We provide a background on the relevant literature in §1.1. Our research context (data and pricing scheme) is provided and analyzed in §2. We outline our analytic model in §3 and estimation strategy in §4. In the subsequent section, we discuss our results along with the policy experiments. Finally, we conclude with a summary of our findings, limitations, and suggestions for future research.

1.1. Relevant Literature

The telecommunication demand modeling literature has a rich history (see Taylor 1994 for a summary). Generally, telecommunication pricing and consumption behavior has been an intriguing research area because the industry has unique characteristics such as high fixed costs, low short-run marginal costs, and network externalities. These issues pose interesting econometric challenges. More importantly, the telecommunication industry traditionally has been a regulated industry, and much work has been conducted to examine the impact of regulatory changes on user demand and social welfare.

The relevant telecommunication literature for our study can be classified into two research topics: (1) evaluating the fixed line telecommunication demand structure and (2) evaluating the mobile communication demand. Many studies have estimated price elasticity of demand in various contexts for fixed line telephony. Taylor and Kridel (1990) make a distinction between telephone access and telephone use while calculating the price elasticity of telephone demand. Distinguishing the use-based (metered) and the flat-rate pricing, some researchers investigate how users choose one over the other and how their demand changes under these pricing regimes (Kling and Van der Ploeg 1990, Park et al. 1983). Miravete (2002) explicitly incorporates users' uncertainty in the number of minutes demanded in the next month when choosing a plan today. Most of the studies find low price elasticities for local voice services. Kling and Van der Ploeg (1990) estimate the elasticity to be on the order of −0.17, whereas Park et al. (1983) estimate about −0.1.

In the domain of wireless telephony, one main research stream has been to examine the penetration of mobile services and how the mobile communication interacts with fixed line communication (Ahn and Lee 1999, Hausman 1999, Rodini et al. 2003, Sung and Lee 2002). The results have been conflicting. Some have found a substitution effect between the fixed lines and mobiles (Ahn and Lee 1999), whereas others have found a complementary effect (Gruber and Verboven 2001). More recently, Economides et al. (2008) reported that the fixed line usage (in both local and regional services) for a household is greater if the household owns a cellular phone, and suggested that the cellular service proxies for the intensity of household communication needs.

Another research stream has analyzed market structure, mark-ups, and competition between mobile providers. Hausman (1999) estimated the welfare impact of diffusion of cellular phones. Garbacz and Thompson (2007) estimated demand for mobile services in developing countries. Miravete and Röller (2004) developed a model of competition between the U.S. cellular operators and recovered key cost parameters from aggregate demand data. Gruber (2001) analyzed how competition affects diffusion of mobile services.

Much of the previous work analyzing mobile communication is focused only on mobile voice service, leaving out data services such as SMS. For example, Iyengar (2004) used individual-level data similar to ours to estimate consumer demand but he does not have information on SMS usage. Given SMS's wide use, the interaction of voice and SMS is an interesting and understudied area. Recently, two papers looked at the interaction between voice and SMS. Andersson et al. (2006) analyzed the relationship of the two services in the context of network size. They estimated the cross-price elasticity of voice and SMS and showed that, depending on the network size, the relationship could be either substitutive or complementary. However, their data were highly aggregate. They used quarterly average messages sent in Norway, and used the data over 33 quarters to estimate a reduced-form model. Such a high level of aggregation precluded them from modeling and interpreting user behavior at a micro level. In a recent paper, Grzybowski and Periera (2008) showed that SMS and voice have a complementary relationship. Although their data were at the individual level, they did not observe the user's plan choice, and hence did not observe the marginal prices paid by the user (or even the average price paid by the user). This again forced data aggregation. Our data are richer in this regard because we can observe plan characteristics and consumers' sequential choices (plan choice and consumption) in detail. This allows us to build a more realistic model of consumer decision making.

Our work is also closely related to the "multiple category purchasing behavior" line of research in the marketing discipline. Researchers have begun to understand cross-category relationships in consumers' decision making using multi category models (see Seetharaman et al. 2005 for a review of this topic). The recognition of cross-category dependencies implies that a consumer's purchase decisions across categories are not independent. In other words, a consumer's decision of whether or how much to buy in one category depends on her corresponding decision in the related category (Chiang 1991, Seetharaman et al. 2005, Song and Chintagunta 2007, Niraj et al. 2008). In our setting, voice and SMS are two such categories that impact the consumption decisions of each other.

We now provide some details on our data and outline key econometric challenges.

2. Research Context

2.1. Data

We collected detailed consumption data on 6,847 subscribers of a cellular service provider in an Asian country (the operator was the third largest in the county with more than two million consumers at that time) from April 2002 through December 2002. The firm offers two kinds of communication services: voice and SMS.[2] We have the following consump-

[2] Because of a disclosure agreement, the name of the firm is not disclosed. The available data were from about 10,000 subscribers for the whole year of 2002. However, the first three months of the data were for a trial period. About a third of the consumers had signed up for a "family plan," which was not specific to an individual, so we dropped those observations. Some observations were dropped when we did not have good demographic information on users. We also dropped those months of data when the users used fewer than 10 minutes of voice. This led to a final sample of about 6,847 consumers for a total of 59,866 observations.

Table 1 Pricing Scheme (Three-Part Nonlinear Tariff)

Plan no.	Fixed fee (units)[a]	Free minutes	Overtime charge of voice service	Charge of SMS
1	350	350	3 units/min	3 units/message
2	800	517		
3	1,100	917		
4	1,500	1,217		
5	2,000	2,117		

[a]In currency units of the country.

tion information at an individual level: (1) the voice service use measured in the number of minutes for each consumer for each month, (2) the number of messages sent for each month, and (3) demographic information about the users (gender, age, and marital status). Users select among five different three-part tariffs offered by the firm. The plan is selected at the beginning of a month. Table 1 shows the pricing scheme offered; the pricing scheme did not change during our research period.

In Table 2, we offer descriptive statistics of our sample. In particular, we show the consumption statistics of voice and SMS across demographic characteristics and across the five plans. The statistics indicate that plan 1 is heavily subscribed to. The user consumption seems to follow a reasonable pattern. The consumption of both voice and SMS is fairly heterogenous. Younger users consume higher volumes of voice minutes as well as of messages. Across the plans, as one moves from plan 1 to plan 5, the consumption of both voice and SMS increases (except for the last plan, where the consumption of SMS decreases). Recall from Table 1 that the free minutes available increase as one moves from plan 1 to plan 5.

2.2. "One-Side" and "Step" Nonlinear Pricing

Because the firm offers a menu of three-part tariffs (second-degree price discrimination), consumer budget constraints are characterized not only by quantities consumed but also by the plan selected (which determines the amount of fixed fee to be paid). In our case, although the quantity of voice service is continuous, the quantity discount is discrete because users realize the benefit of nonlinear pricing only by choosing a higher plan. Here we will call this "step" nonlinear pricing. By contrast, SMS is sold on a flat, metered basis (linear pricing). Thus, the pricing is "one-side" nonlinear pricing. "One-side" and "step" nonlinear pricing affect consumer utility maximization by generating unique budget constraints as shown in Figure 1. The x-axis is the expenditure on SMS and the y-axis is the expenditure on voice. The budget line is kinked. Consider a user with a budget constraint of 2,000 units (enough to select plan 5). The user can choose plan 5, which will give her 2,117 free minutes.

Table 2 Average Usage Statistics

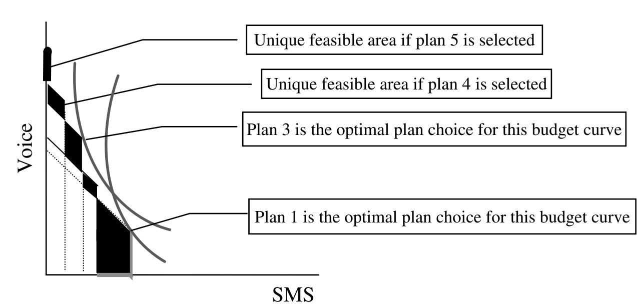
	Mean	SD	Min	Max	N
Whole sample					
Voice	281.3	252.6	10	11,845	59,866
SMS	11.8	34.2	0	3,106	
Users with age < 30					
Voice	329.8	299.5	10	11,845	22,483
SMS	16.7	36.7	0	1,558	
$30 \leq$ age < 40					
Voice	265.1	223.5	10	6,602	29,520
SMS	9.2	33.7	0	3,106	
Age ≥ 40					
Voice	203.7	168.1	10	2,917	7,863
SMS	7.3	26.8	0	1,079	
Female					
Voice	287.1	257.9	10	11,845	32,592
SMS	13.2	31.6	0	1,558	
Male					
Voice	274.4	246.0	10	6,602	27,274
SMS	10.1	37.1	0	3,106	
Plan 1					
Voice	249.0	167.7	10	2,887	56,634
SMS	10.2	31.6	0	3,106	
Plan 2					
Voice	642.6	316.9	20	4,612	2,103
SMS	39.1	59.0	0	932	
Plan 3					
Voice	1,011.4	505.8	53	5,170	750
SMS	44.3	60.2	0	633	
Plan 4					
Voice	1,556.4	994.4	169	11,845	286
SMS	43.1	54.8	0	562	
Plan 5					
Voice	1,994.4	889.3	430	4,335	93
SMS	27.5	41.0	0	188	

Notes. Voice is measured as the amount of minutes consumed per month. SMS is measured as the number of messages per month.

Then the user cannot consume any SMS because it will exceed the given budget. If the user wants to consume SMS, the user needs to move to a lower plan and save some money for SMS consumption, hence the vertical downward shift in the budget line leading to a kink.

Under the same budget, every plan is characterized by distinctive areas indicating available amounts for voice and SMS. In Figure 1, the shaded area under the budget line shows the feasible selection area of

Figure 1 Feasible Area Conditional on a Voice Plan and Optimal Plan Choice



voice and SMS within the chosen plan. To choose the optimal mix of voice and SMS, consumers search for the highest indifference curve that touches the feasible set. The graph draws the indifference curve of one consumer who chooses plan 1 and an another consumer who chooses plan 3, which maximize their utilities, respectively. In our fully articulated model described below, the total amount spent on voice plus SMS is endogenously chosen by the consumer.

This kinked budget line is consistent with the nonlinear pricing models (Moffitt 1990) and leads to an interesting and unusual structure of the budget set, which complicates our estimation task but enables us to identify our model, as we explain later.

3. Econometric Model

3.1. Conceptual Background

In the case of many goods, a change in the price of a good induces income and substitution effects that affect the demand for the other goods. Two goods are Marshallian substitutes if a price increase in one increases demand for the other. On the other hand, the goods are complements if a price increase in one decreases demand for the other good.

To prove that two goods are complements, one cannot simply show that their consumptions positively covary. That is, if we find that high voice users also tend to be high SMS users, this *does not* imply that they are complements. Two goods are complements if an increase of price in one yields a reduction in demand for the other. As pointed out by Manchanda et al. (1999), multiple categories can be bought together on one shopping trip because of (1) complementarity (or cross-effect), (2) consumer heterogeneity, and (3) coincidence. Because the main focus of this study is to examine cross-effect, we need to control the other two confounders, heterogeneity and coincidence. To control for heterogeneity in inherent preferences for voice and SMS, we allow user types (as determined by the preferences for both services) to follow a probability distribution (discussed later). We also let the preferences for voice and SMS covary to account for preference homogeneity. Furthermore, we use a user demographic profile to account for the remaining consumer heterogeneity. The varieties of coincidence that are most likely to apply (e.g., habit persistence) can also be accommodated within the correlated user type (or equivalently via the satiation points), as we model below.

3.2. Analytical Framework

Our analytical model is based on (discrete) plan choice and (continuous) consumed quantities across two services under the given one-side and step nonlinear pricing. Consumers are assumed to maximize utility by making two decisions at different points on the time horizon. The consumer makes a plan choice decision among discrete alternatives at the first stage. This choice is based on the expected optimal consumption bundle. The consumer decides continuous quantities of service usage in the second stage. As a result, the analytic framework takes the form of a two-stage discrete/continuous mixture choice model.

The important early papers that laid the blueprint for these modeling contexts used discrete choice models such as logit and probit (e.g., Albert and Chib 1993), or the multinomial brand choice models (e.g., Rossi et al. 1996). But the continuous quantity decision at the second stage may not be well addressed by the (nested) logit model because transforming continuous variables into discrete variables causes loss of information (Niraj et al. 2008). Similarly, the approach to treat different combinations—three element vectors of discrete plan choice and two continuous consumptions in our context—as different multinomial choice alternatives leads to an immense number of alternatives, making multinomial choice unviable. Hanemann (1984) developed a unified framework for formulating econometric models of discrete/continuous choices in which both discrete and continuous choices aim at the same underlying utility maximization decision. But in their model, the alternatives for discrete choices are essentially (perfect or near perfect) substitutes, and thus a consumer prefers to buy only one option (e.g., brand A over brand B, C, and D) at any time, and the continuous choice is the magnitude of the option chosen. However, in our context, the assumption that only one alternative has to be selected on a discrete choice occasion is not appropriate, because the majority of consumers select both services simultaneously. Finally, and more importantly, given that both services are selected in a mixed discrete/continuous context, separate modeling of the two decisions—using a logit model for the discrete choice and regression models for the continuous decision—is incorrect due to the endogeneity problem (ultimately causing inefficient estimation or the omitted variable problem) if the quantity decision is not statistically independent of the choice decision and vice versa.[3]

In our context, we deal with the two-goods problem, whereas the literature has problems focused on one good. It is obvious that the two decisions (which plan to select and how many minutes in voice and

[3] To solve the problem of the two-step approach, Krishnamurthi and Raj (1988) first estimated the choice model with a discrete choice model and then estimated the regression model based on the parameters estimated after correcting selection bias. But, their research context was also about the choice of one brand and its subsequent consumption.

how many messages in SMS to use) are statistically dependent. This violates the basic assumption of the two-step approach (limited-information maximum likelihood estimation (MLE)), so we have to estimate the model with full-information MLE. In the following, we derive the joint likelihood function for the observed consumption data from general distributional assumptions regarding consumer heterogeneity and random error terms.

3.3. User Utility Specification

Because this paper contains a large number of notations, for readability we first provide a table with notations. We will continue to refer to Table 3 throughout.

We consider individuals indexed by $i = 1, 2, \ldots, N$. The individuals choose a plan from the set of available plans, indexed by $k = 1, 2, \ldots, 5$ over time $t = 1, 2, \ldots, T$. They then choose continuous quantities of voice q_{ikt} and SMS s_{ikt}, conditional on the plan choice. When a user selects plan k, the user must pay a fixed fee, T_k, to sign up for the plan, and is allowed to use a free quota of voice minutes, FQ_k.

Table 3 Notations Used in this Paper

Index i, k, t	i represents a user, k represents a plan, and t represents time (month)
q_{ikt}	Number of voice minutes consumed by user i under plan k during month t
s_{ikt}	Number of messages sent by user i under plan k during month t
FQ_k	Number of free quota minutes provided in plan k
T_k	Fixed fee to enrol in plan k
y_{it}	Normalized outside good consumed by user i during month t
I_{it}	Income of user i during month t.
θ_{iqt}, θ_{ist}	User types indicating user i's inherent preference for voice service and SMS during month t
μ_q, σ_q	Estimated mean and standard deviation of θ_{iqt}
μ_s, σ_s	Estimated mean and standard deviation of θ_{ist}
ρ	Correlation in preference for voice and SMS
p_q	Marginal price of each voice minute after free quota is exceeded
p_s	Marginal price of each short message
ε^q_{ikt}, ε^s_{ikt}	Error terms that capture the difference between expected and actual consumption of voice and SMS
$\sigma_{\varepsilon q}$, $\sigma_{\varepsilon s}$	Estimated variance of the error terms above
b_q, b_s	Estimated price response parameters for voice and SMS
b_{int}	Estimated parameter indicating impact of voice consumption on SMS consumption
PBC_k	Plan boundary curve: Indifference curve that separates plan $k-1$ from plan k. Users below PCB_k choose plan k.
FQC_k	Free quota curve: Indifference curve that separates satiated users from others. Users below FQC_k are satiated and do not expect to use marginal minutes, whereas users above the curve are either at the kink or use marginal minutes.
MMC_k	Marginal minutes curve: Indifference curve that separates the users exceeding free quota minutes from the rest. Users above MMC_k are expected to use marginal minutes, whereas users below the curve are either at the kink or satiated.

Once the user exceeds the free quota, she has to incur a marginal cost, p_q, per minute for voice. Thus, each plan is a three-part tariff that can be characterized by T_k and FQ_k. Users always pay the marginal price p_s for each SMS. We assume that the consumers spend the remainder of their income on an outside good, y_i, at a normalized unit price.

We assume that the utility obtained by an individual i from selecting plan k and consuming (q_{ikt}, s_{ikt}) of voice and SMS is quadratic in quantities consumed. The utility structure is of the form

$$\begin{aligned} &U_{ikt}(q_{ikt}, s_{ikt}, y_{it} \mid \theta_{iqt}, \theta_{ist}, b_q, b_s, b_{int}) \\ &\quad = \frac{1}{b_q}\left(\theta_{iqt} q_{ikt} - \frac{q_{ikt}^2}{2}\right) + \frac{1}{b_s}\left(\theta_{ist} s_{ikt} - \frac{s_{ikt}^2}{2}\right) \\ &\qquad + b_{int} q_{ikt} s_{ikt} + y_{it}\, b_q, b_s, \theta_{iqt}, \theta_{ist} \geq 0. \end{aligned} \tag{1}$$

The quadratic utility structure is widely modeled in the telecommunication demand literature (Economides et al. 2008, Miravete and Röller 2004). Note that the utility function is quasi-linear in the outside good so that the choice of plan and consumption do not depend on income. In short, we assume that although change in price will induce both substitution (or complementary) and income effect, the income effect is negligible. This assumption is reasonable and widely used because a relatively small portion of income is assigned to the telephony services (Park et al. 1983, footnote 5; Economides et al. 2008).[4] Low income effects are highlighted in many other goods as well (see Tammo et al. 2005). Though it must be stressed that unavailability of income data is a limitation of our analysis.

In Equation (1), the first term represents utility from voice consumption, the second term represents utility from SMS consumption, and the fourth term captures utility from outside goods. As is commonly done, we assume that the utility from outside goods and the utility from mobile services are separable. Thus, consumption in mobile services does not affect the marginal utility obtained from the outside good, and vice versa.[5] However, utilities from voice service and SMS are inseparable. Thus, a change in consumption of one service can influence consumption of the other service. This interaction is modeled through the third term in Equation (1), and b_{int} captures the substitutive

[4] In our setting, most users spend less than 0.3% of their average per-capita income on monthly cellular tariffs.

[5] Some utility from outside goods (e.g., fixed phone service and broadband services) might be dependent on wireless services. However, the portion of utility from fixed phone service and broadband services is very small (or negligible) compared to the utility from all other human activity, and there is no evidence of a consistent relationship between them.

effect (if negative), the complementary effect (if positive), no effect (if zero). Thus, our specification is very flexible. Our utility structure assumes diminishing marginal utility of voice service and SMS and is the same for all users.[6]

The parameters b_q and b_s represent price responses and can be used to calculate the own- and cross-price elasticities. The set $(\theta_{iqt}, \theta_{ist})$ represents a user's inherent preferences for two services and varies across users. Higher values of $(\theta_{iqt}, \theta_{ist})$ represent higher preferences for the services, and $(\theta_{iqt}, \theta_{ist})$ is closely related to the concept of consumer satiation point—the maximum consumption level desired for a good at any price (consumers will never demand more than the satiation point, even if the price is zero).[7] The notion of satiation point is widely modeled in the telecommunication demand literature (Economides et al. 2008, Kling and Van der Ploeg 1990). Notice that we allow the satiation points to vary over time, which is consistent with variation in consumption across time for a user.

Consumers first choose a plan and then choose quantities for voice and SMS. We solve the problem with backward induction, starting with the second stage, where an individual user selects optimal quantities for voice and SMS conditional on the plan choice k in the first stage and subject to the budget constraint I_{it}:

$$I_{it} \geq T_k + p_q(q_{ikt} - FQ_k)^+ + p_s s_{ikt} + y_{it},$$

where

$$(q_{ikt} - FQ_k)^+ = \begin{cases} q_{ikt} - FQ_k & \text{if } q_{ikt} > FQ_k, \\ 0 & \text{if } q_{ikt} \leq FQ_k. \end{cases} \tag{2}$$

Equation (2) indicates that the marginal price of voice kicks in only if the free quota of minutes is exceeded. Substituting the budget constraint into the utility maximization objective function yields

$$\begin{aligned} &\max_{q_{ikt}, s_{ikt}} U_{ikt}(q_{ikt}, s_{ikt} \mid k) \\ &\quad = \frac{1}{b_q}\left(\theta_{iqt} q_{ikt} - \frac{q_{ikt}^2}{2}\right) + \frac{1}{b_s}\left(\theta_{ist} s_{ikt} - \frac{s_{ikt}^2}{2}\right) \\ &\qquad + b_{int} q_{ikt} s_{ikt} + I_i - T_k - p_q(q_{ikt} - FQ_k)^+ - p_s s_{ikt}. \end{aligned}$$

[6] The second derivative of the utility function with respect to q and s is negative.

[7] The satiation points are $((\theta_{iq} + b_{int} b_q \theta_{ist})/(1 - b_{int}^2 b_s b_q),\ (\theta_{is} + b_{int} b_s \theta_{iq})/(1 - b_{int}^2 b_s b_q))$. The satiation point is composed of three parameters to be estimated and a consumer type. If there is no cross-effect ($b_{int} = 0$), $(\theta_{iq}, \theta_{ist})$ is individual i's satiation points. Because of interaction, the satiation point in SMS affects a satiation point in voice service and vice versa. The mutual influence of the cross-effect on satiation points critically depends on the sign of b_{int}.

The first-order condition is

$$\begin{cases} \dfrac{\partial U_{ikt}}{\partial q_{ikt}} = \dfrac{\theta_{iqt} - q_{ikt}}{b_q} + b_{int} s_{ikt} - p_q = 0 & \text{if } q_{ikt} > FQ_k, \\[2ex] \dfrac{\partial U_{ikt}}{\partial q_{ikt}} = \dfrac{\theta_{iqt} - q_{ikt}}{b_q} + b_{int} s_{ikt} = 0 & \text{if } q_{ikt} \leq FQ_k \end{cases}$$

and

$$\frac{\partial U_{ikt}}{\partial s_{ikt}} = \frac{\theta_{ist} - s_{ikt}}{b_s} + b_{int} q_{ikt} - p_s = 0.$$

Solving these simultaneously yields the following demand equations:

$$\begin{aligned} &(q_{ikt}, s_{ikt})_{b_{int}=0} \\ &\quad = \begin{cases} (\theta_{iqt} - b_q p_q, \theta_{ist} - b_s p_s) & \text{if } q_{ikt} > FQ_k, \\ (\theta_{iqt}, \theta_{ist} - b_s p_s) & \text{if } q_{ikt} \leq FQ_k; \end{cases} \end{aligned} \tag{3}$$

$$\begin{aligned} &(q_{ikt}, s_{ikt})^*_{b_{int} \neq 0} \\ &\quad = \begin{cases} (\theta_{iqt} - b_q p_q + b_q b_{int} s_{ikt}, \theta_{ist} - b_s p_s + b_s b_{int} q_{ikt}) \\ \qquad\qquad\qquad\qquad \text{if } q_{ikt} > FQ_k, \\ (\theta_{iqt} + b_q b_{int} s_{ikt}, \theta_{ist} - b_s p_s + b_s b_{int} q_{ikt}) \\ \qquad\qquad\qquad\qquad \text{if } q_{ikt} \leq FQ_k. \end{cases} \end{aligned} \tag{4}$$

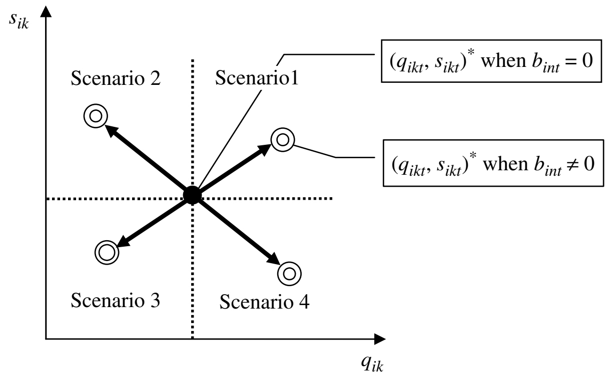
Equations (3) and (4) present the demand curves for voice service and SMS in the absence and presence of the cross-effects, respectively. They show that the optimal bundle in one service shifts because of the consumption level of the other service, reflecting the interaction effect of two services. Therefore, comparison of the consumption bundles of two services when there is no cross-effect and when there is a cross-effect enables us to see the impact of the cross-effect on usage level as well as how to intuitively interpret b_{int} embedded in the interaction term.

The middle point in Figure 2 is the optimal consumption when there is no interaction between voice and SMS ($b_{int} = 0$). The consumption bundle can move to one of the four directions on two dimensions of q_{ikt} and s_{ikt} when $b_{int} \neq 0$ (see Figure 2)—i.e., the comparison of Equations (3) and (4) presents four possible scenarios regarding the change of the optimal consumption bundle of voice and SMS.

If the interaction effect exists, we can intuitively divide it into two groups—the symmetric scenario (1 and 3) versus the asymmetric scenario (2 and 4). Given the definition and the utility specification, the cross-effect must be symmetric, whatever its direction is: increasing in scenario 1 or decreasing in scenario 3.[8] Because b_q, b_s, q_{ikt}, and s_{ikt} are positive, the

[8] Given our utility function, this symmetry can be shown by Young's theorem: $\partial U/(\partial q_{ikt} \partial s_{ikt})$ is constant.

Figure 2 Satiation Points and the Cross-Effect



direction of the change depends on the sign of b_{int} (positive or negative). In scenario 1 ($b_{int} > 0$), consumption in one service induces more consumption in other service. As a result, positive b_{int} provides evidence that one service plays a role of complementing the other service. In scenario 3 ($b_{int} < 0$), voice service is a substitute for SMS and vice versa. Consequently, the role of b_{int} is clear: the presence or absence of b_{int} as well as its sign captures the potential interaction effect of these two services.

3.4. Demand Function

Because of the nonlinearities of three-part tariffs, the demand functions need close inspection. In particular, there are three distinct possibilities: (i) a user's satiation point is less than the free quota, (ii) the satiation point is more than the free quota but the marginal price is high enough that the user does not want to exceed the free quota minutes, and (iii) the satiation point is high enough that the user consumes marginal minutes.

Thus, within each plan, a user can be in three different regions (R_k^1, R_k^2, R_k^3) depending on their type ($\theta_{iqt}, \theta_{ist}$). These regions are characterized by indifference curves, PBC_k, FQC_k, and MMC_k, which are derived in detail in the next section. These indifference curves are the threshold values of ($\theta_{iqt}, \theta_{ist}$) that separate users in different regions. Figure 3 explains these regions more intuitively. We draw these for plans 1 and 2. The rest are analogous. PBC_k (the plan boundary curve) separates plan k from plan $k+1$. Notice that PBC_1 separates plan 1 from plan 2. So there are four PBC curves. PBC_0 is simply ($\theta_{iqt} = 0$, $\theta_{ist} = 0$).

FQC (the free quota curve) and *MMC* (the marginal minute curve) are the indifference curves that separate the regions within a plan. *FQC* separates the users who are satiated from the users who are at the kink. Thus, the users between PBC_0 and FQC_1 (R_1^1) are satiated. Similarly, *MMC* separates the users who are at the kink from the users who are willing to use marginal minutes by exceeding the free quota. Thus, users between FQC_1 and MMC_1 (R_1^2) are at the kink, and the users between MMC_1 and PBC_1 (R_1^3) are using marginal minutes.

More formally, the demand curve within each region can be derived from Equations (3) and (4):

$$\text{if } PBC_{k-1} \leq (\theta_{iqt}, \theta_{ist}) < FQC_k$$
$$\begin{cases} q_{ikt} = \dfrac{\theta_{iqt} + b_{int} b_q \theta_{ist} - b_{int} b_q b_s p_s}{1 - b_{int}^2 b_s b_q} \\[2ex] s_{ikt} = \dfrac{\theta_{ist} + b_{int} b_s \theta_{iqt} - b_s p_s}{1 - b_{int}^2 b_s b_q} \end{cases} \qquad ------R_k^1$$
$$\text{if } FQC_k \leq (\theta_{iqt}, \theta_{ist}) < MMC_k$$
$$\begin{cases} q_{ikt} = FQ_k \\ s_{ikt} = \theta_{ist} + b_{int} b_s FQ_k - b_s p_s \end{cases} \qquad ------R_k^2$$
$$\text{if } MMC_k \leq (\theta_{iqt}, \theta_{ist}) < PBC_k$$
$$\begin{cases} q_{ikt} = \dfrac{\theta_{iqt} + b_{int} b_q \theta_{ist} - b_{int} b_q b_s p_s - b_q p_q}{1 - b_{int}^2 b_s b_q} \\[2ex] s_{ikt} = \dfrac{\theta_{ist} + b_{int} b_s \theta_{iqt} - b_{int} b_q b_s p_q - b_s p_s}{1 - b_{int}^2 b_s b_q} \end{cases} \qquad ------R_k^3 \tag{5}$$

where

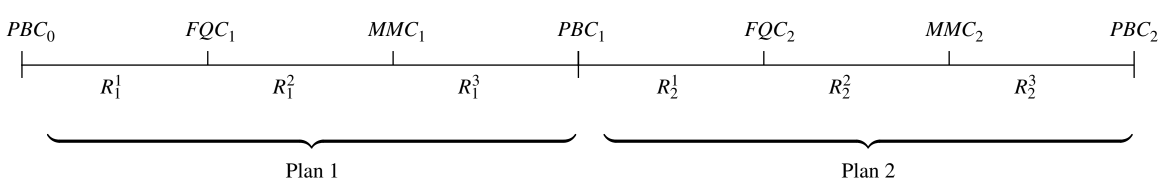
$$PBC_k = \{(\theta_{iqt}, \theta_{ist} \mid k) \mid U^*(R_k^3 \mid k) = U^*(R_k^1 \mid k+1)\},$$
$$FQC_k = \{(\theta_{iqt}, \theta_{ist} \mid k) \mid U^*(R_k^1 \mid k) = U^*(R_k^2 \mid k)\},$$
$$MMC_k = \{(\theta_{iqt}, \theta_{ist} \mid k) \mid U^*(R_k^2 \mid k) = U^*(R_k^3 \mid k)\}.$$

As noted above, the demand in R_k^1: (PBC_{k-1}, FQC_k) is for the users who are satiated, R_k^2: (FQC_k, MMU_k) is for the users at the kink, and R_k^3: (MMU_k, PBC_k) is for the users using marginal minutes.

3.5. User Heterogeneity and Indifference Curves

These indifference curves are derived by comparing the utilities after plugging the optimal (q_{ikt}^*, s_{ikt}^*) from Equation (5) into the utility function in Equation (1). It should be noted that the optimal consumptions in these two services are determined simultaneously.

Figure 3 Different Regions Across and Within Plans



We show that all users whose $(\theta_{iqt}, \theta_{ist})$ belongs to R_k: (PBC_{k-1}, PBC_k) select plan k. As the first step in deriving PBC_k, we classify users into five groups, which is equivalent to the number of plans available. The following axioms outline user heterogeneity based on their plan choice.

Axiom 1. *If a user is indifferent between two plans (plan k and plan $k+1$), the cost of plan k ("the fixed fee of plan k" + "the variable cost depending on additional usage beyond the free minutes of plan k") and the cost of plan $k+1$ (only the fixed fee of plan $k+1$) have to be identical at the expected optimal consumption point in voice service.*[9] *As a result, a consumer with a higher θ_{iqt} will choose a higher value plan to realize benefits of the nonlinear pricing scheme. Thus, the ranking of the plans is monotone in θ_{iqt}.*

Axiom 2. *Even if the marginal cost of SMS (p_s) is identical regardless of plan choice, the plan choice is affected by θ_{ist} due to the interaction. The expected optimal consumption of SMS is the monotone transformation of θ_{iqt}, which is monotone in plan choice. Therefore, the ranking of plans is monotone in θ_{ist} as well.*

Axiom 3. *From Axioms 1 and 2, the plan choice is ordered in θ_{iqt} and θ_{ist}, despite one-side nonlinear pricing. That is, the following equation always holds: $R_k = \{(\theta_{iqt}, \theta_{ist}) \mid U_{ik} \geq U_{ij}, \forall j \neq k\} = \{(\theta_{iqt}, \theta_{ist}) \mid U_{ik} \geq U_{ik+1}$ and $U_{ik} \geq U_{ik-1} \mid k \geq 2,\ U_{ik} \geq U_{ik+1} \mid k = 1\}$, where R_k indicates (PBC_{k-1}, PBC_k).*

Axiom 4. *The indifference curve between plan k and plan $k+1$ can be computed by comparing utilities generated from two plans: $PBC_k = \{(\theta_{iqt}, \theta_{ist} \mid k) \mid U^*(R_k^3 \mid k) = U^*(R_k^1 \mid k+1)\}$.*

Axiom 5. *The shape of indifference curves (here, the slope of indifference lines) is identical across plans under the assumption that the utility function is homothetic.*

Axiom 6. *Because the indifference curves do not overlap one another, indifference lines from the formula in Axiom 4 will be the boundaries dividing five plans.*

Comparing the indirect utility functions across five plans, we can compute the indifference curves separating the two adjoining plans (from Axiom 4). The indifference curves are given by

$$\begin{aligned} PBC_k\colon\ \theta_{ist} &= (2FQ_k p_q - 2b_{int}^2 b_q b_s FQ_k p_q + b_q p_q^2 + 2b_{int} b_q b_s p_q p_s - 2T_k \\ &\quad + 2b_{int}^2 b_q b_s T_k + 2T_{k+1} - 2b_{int}^2 b_q b_s T_{k+1}) \cdot (2b_{int} b_q p_q)^{-1} \\ &\quad - \frac{\theta_{iqt}}{b_{int} b_q}. \end{aligned}$$

[9] The actual consumption may differ from the optimal consumption due to some random shocks of zero mean. However, users can optimize plan choice only on the basis of expected optimal consumption.

We can see that the indifference curve is affected by (T_k, FQ_k, p_q, and p_s) and (b_q, b_s, and b_{int}). From Axiom 5, all indifference curves have the same slope, $-1/(b_q b_{int})$. The interpretation of the slope is identical to the interpretation of b_{int} as we described in §3.2: a positive value of b_{int} indicates a substitutive relationship, and a negative value a complementary relationship.

As a numerical illustration, we plug real numbers from our data into this equation to derive the indifference curve, PBC_1, separating plan 1 and plan 2:

$$PBC_1\colon\ \theta_{ist} = \frac{1{,}000 + 3b_q + 6b_{int} b_q b_s - 1{,}000 b_{int}^2 b_q b_s}{2b_{int} b_q} - \frac{\theta_{iqt}}{b_{int} b_q}.$$

FQC_k and MMC_k can be derived in the similar way. Plugging the optimal (q_{ikt}, s_{ikt}) from R_k^1 and R_k^2 into the indirect utility function will yield FQC_k, and the optimal (q_{ik}, s_{ik}) from R_k^2 and R_k^3 will yield MMC_k.

$$FQC_k\colon\ \theta_{ist} = \frac{FQ_k + 3b_{int} b_q b_s - FQ_k b_{int}^2 b_q b_s}{b_{int} b_q} - \frac{\theta_{iqt}}{b_{int} b_q},$$

$$MMC_k\colon\ \theta_{ist} = \frac{FQ_k + 3b_q + 3b_{int} b_q b_s - FQ_k b_{int}^2 b_q b_s}{b_{int} b_q} - \frac{\theta_{iqt}}{b_{int} b_q}.$$

4. Model Estimation

To operationalize the theoretical model, we derive a joint likelihood function. First, we formulate the probability of choosing a plan k and then the probability of observing the actual consumption bundle (q_{ikt}, s_{ikt}) conditional on the plan choice.

4.1. Probability of Choosing a Plan k

We calculate the probability of an individual i choosing a plan k from the fact that the individual chooses the plan that maximizes her utility conditional on the expected future consumptions (not real consumption) derived from $(\theta_{iqt}, \theta_{ist})$ and (b_q, b_s, and b_{int}).[10] The parameters (b_q, b_s, and b_{int}) are to be estimated, whereas $(\theta_{iqt}, \theta_{ist})$ are not observed by econometricians. We make a distributional assumption on unobserved user types (user heterogeneity).

We assume that the distribution of the user type, $f(\theta_{iqt}, \theta_{ist})$, follows a truncated bivariate normal distribution:

$$\begin{pmatrix} \theta_{iqt} \\ \theta_{ist} \end{pmatrix} \sim \mathrm{TBN}(\mu, \Sigma), \quad \mu = \begin{pmatrix} \mu_q \\ \mu_s \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \sigma_q^2 & \rho\sigma_q\sigma_s \\ \rho\sigma_q\sigma_s & \sigma_s^2 \end{pmatrix}.$$

[10] After selecting plan k, they cannot change the plan during the month.

Because the satiation point can never be less than zero, a truncated distribution is appropriate. Moreover, normal distribution is very flexible and tractable. We assume that $(\theta_{iqt}, \theta_{ist})$ follow a joint normal distribution rather than independent normal distributions—i.e., we allow ρ to capture the unobserved inherent association of two service consumptions, θ_{iqt} and θ_{ist}. Thus, $\rho > 0$ indicates that the preference for voice service is positively correlated with that for SMS. The assumption of a joint distribution allows us to examine the cross-effect by excluding the impact of inherent association (Manchanda et al. 1999). The structural parameters to be estimated are (b_q, b_s, and b_{int}), the means of the distribution (μ_q and μ_s), the variances (σ_q and σ_s), and the correlation (ρ).

Given the distributional assumption, we can write the probability of a consumer choosing plan:

$$\begin{aligned}\Pr_{ikt} &= \Pr(\text{plan}=k \mid \theta_{iqt}, \theta_{ist})\\ &= \iint_{R_k} f(\theta_{iqt}, \theta_{ist})\, d\theta_{iqt}\, d\theta_{ist}, \end{aligned} \tag{6}$$

where

$$f(\theta_{iqt}, \theta_{ist}) = \frac{e^{(-Q/2)}/2\pi\sigma_q\sigma_s\sqrt{1-\rho^2}}{1-\int_{-\infty}^{0}\int_{-\infty}^{0} f(\theta_{iqt}, \theta_{ist})\, d\theta_{iqt}\, d\theta_{ist}},$$

$$Q = \frac{1}{1-\rho^2}\left[\frac{(\theta_{iqt}-\mu_q)^2}{\sigma_q^2} - 2\rho\frac{(\theta_{iqt}-\mu_q)(\theta_{ist}-\mu_s)}{\sigma_q\sigma_s} + \frac{(\theta_{ist}-\mu_s)^2}{\sigma_s^2}\right],$$

$$R_k = (PBC_{k-1}, PBC_k).$$

4.2. Probability of Consumption Conditional on Plan Choice

In the second stage, users select a consumption bundle of voice and SMS. However, the actual demand observed would not be equal to the expected demand derived from Equation (5). The difference between the "actual" consumption and the "expected" consumption occurs due to unobserved noise that is unanticipated by consumers. Thus, the difference reflects a random shock (or measurement error) a user may experience in a given month. For the empirical estimation, we add an error term in the demand function as an additive form following Burtless and Hausman's (1978) approach. Absent an error term, all consumers in the R_k^2 region would always cluster at the kink—a fact that is inconsistent with the data. Thus, the realized demand is $q_{ikt} = E(q_{ikt}) + \varepsilon_{ikt}^q$, where the expected demand $E(q_{ikt})$ is derived in Equation (5). The SMS demand curve is derived analogously. We assume that the distribution of error terms, $h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)$, follows the independent bivariate normal distribution:

$$\begin{bmatrix}\varepsilon_{ikt}^q\\ \varepsilon_{ikt}^s\end{bmatrix} \sim BVN\left(\begin{pmatrix}0\\0\end{pmatrix}\begin{pmatrix}\sigma_{\varepsilon q}^2 & 0\\ 0 & \sigma_{\varepsilon s}^2\end{pmatrix}\right).$$

Assuming $f(\theta_{iqt}, \theta_{ist})$ and $h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)$ are stochastically independent, the probability of observing the consumption depends on plan choice k and the region R_k within the plan. Recall from Figure 3 that a user could be satiated, at the kink, or using marginal minutes. Therefore, the probability of observing (q_{ikt}, s_{ikt}) conditional on plan k is

$$g(q_{ikt}, s_{ikt} \mid k) = \frac{\iint_{R_k^1} h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)\,dF + \iint_{R_k^2} h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)\,dF + \iint_{R_k^3} h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)\,dF}{\iint_{R_k} dF}, \tag{7}$$

where $dF = f(\theta_{iqt}, \theta_{ist})d\theta_{iqt}d\theta_{ist}$. Regions R_k are as defined previously.

4.3. Joint Likelihood Function

We can now derive the joint distribution of a user i choosing a plan k and then choosing (q_{ikt}, s_{ikt}) conditional on user type. The joint likelihood function for a user i is $\Pr_{ikt}(k \mid \cdot) \times g_{ikt}(q_{ikt}, s_{ikt} \mid \cdot, k)$. The likelihood function across both individual i and time t is

$$\begin{aligned} L &= \prod_i \prod_t \left(\Pr_{ikt}(\cdot) \times g_{ikt}(\cdot)\right)\\ &= \prod_i \prod_t \left(\iint_{R_k} h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s) f(\theta_{iqt}, \theta_{ist})\, d\theta_{iqt}\, d\theta_{ist}\right). \end{aligned} \tag{8}$$

Any statistical or econometric model makes an important trade-off between an analytically tractable, but parametrically restrictive specification and an analytically complicated, but parametrically flexible specification. In similar vein, our formulation also imposes two nonlinear constraints derived from the concavity condition of the utility function and from the condition of monotone ordering of indifference curves:

(i) $0 < b_{int}^2 b_s b_q < 1$,
(ii) $b_q < (34/3)(1 - b_{int}^2 b_s b_q)$.

We cannot obtain a closed-form expression for the joint likelihood function. Thus, we need to compute it numerically. We estimate the model in two ways: (1) MLE with direct numerical approximation (quadrature integration) and (2) maximum simulated likelihood estimation (MSLE) using Monte Carlo methods. MLE imposes a significant computational burden involved in numerical integration restricting the sample size. MSLE, on the other hand, places less computational burden, and hence we report the estimates from MSLE. For MSLE, we selected $f(\theta_{iqt}, \theta_{ist})$ as an importance function. The simulated log-likelihood function is the average value of the function of $h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)$ over the importance function. We thus maximized the simulated log-likelihood function subject to the constraints:

$$\ln L \approx \sum_{i,t} \ln(E_{h(\varepsilon_{ikt}^q, \varepsilon_{ikt}^s)}[Tr \times f(\theta_{iqt}, \theta_{ist})]). \tag{9}$$

In Equation (9), *Tr* is a truncation compensation factor. To sample from a bivariate normal distribution of $(\theta_{iqt}^{rng}, \theta_{ist}^{rng})$, we began with draws from two standard normal distributions of (e_1^{rng}, e_2^{rng}) and then used the Cholesky decomposition to transform the random variables. By matrix algebra, it is possible to decompose the bivariate normal distribution in the following way:

$$\begin{pmatrix} \theta_{iqt}^{rng} \\ \theta_{ist}^{rng} \end{pmatrix} = \begin{pmatrix} \mu_q \\ \mu_s \end{pmatrix} + \begin{pmatrix} \sigma_q(\sqrt{1-\rho^2}) & \sigma_q\rho \\ 0 & \sigma_s \end{pmatrix} \begin{pmatrix} e_1^{rng} \\ e_2^{rng} \end{pmatrix}.$$

When we maximized the simulated log-likelihood, we used the same set of random draws for every computation to achieve continuity (1,000 draws). That is, each observation has its own unchanging vector of 1,000 draws. We used the BHHH (outer product of gradients) estimator to compute the asymptotic covariance matrix of the simulated maximum likelihood estimator.

4.4. Identification Issues

Because the three choice decisions, $[(k), (q_{ikt}, s_{ikt})]$, are the consequences of a single utility maximization for an individual, the model ensures that these decisions provide, in combination, the greatest possible utility to the individual. The parameters to be estimated in this study are grouped into two categories: (1) three structural parameters embedded in the utility function, (b_q, b_s, and b_{int}), and (2) distribution-related parameter groups consisting of (μ_q, μ_s, σ_q, σ_s, and ρ) for the user type distribution, and ($\sigma_{\varepsilon q}$ and $\sigma_{\varepsilon s}$) for the measurement error distribution.

First, note that all distributional parameters are readily identified because of the systematic variation in plan choices and consumptions across individuals. The mean values of voice (μ_q) and SMS (μ_s) with associated variance σ_q and σ_s, respectively, are identified because of variation in consumption across users and time. The parameters $\sigma_{\varepsilon q}$ and $\sigma_{\varepsilon s}$ are identified because of differences in expected and actual consumption. Analogously, ρ is identified.

The key price response parameters need closer examination. The parameter b_q (price response parameter to voice service) is readily identified as users face different marginal prices depending on whether they are satiated or not. In particular, consider a user type $\hat{\theta}_{iqt}$ who is indifferent between plan k and plan $k+1$. In plan k, the user is using marginal minutes, whereas in plan $k+1$ the user is satiated. Thus, for a slight variation around such a $\hat{\theta}_{iqt}$, the user faces different marginal costs affecting her voice consumption (a user of type $\hat{\theta}_{iqt} - \epsilon$ is using marginal minutes, whereas a user of type $\hat{\theta}_{iqt} + \epsilon$ is satiated, leading to different consumption amounts by potentially the same $\hat{\theta}_{iqt}$ type). By analogy, b_q is also identified off the kinks within a plan (recall that three different regions (R_k^1, R_k^2, and R_k^3) within a plan force a user to face different price schedules). In contrast, b_s is not directly identified because there is no change in SMS pricing—the marginal cost of SMS is constant even around the kinked budget constraint. However, variation in the price of voice minutes should impact SMS consumption through the hypothesized cross-effect, and thus b_s is (indirectly) identifiable. Our key parameter of interest, b_{int}, is identified through changes in marginal prices for voice as explained above. Moreover, whenever a user exceeds her free quota limits, her marginal price of voice changes from 0 to 3. This change in voice price provides an opportunity to identify how SMS consumption changes. Because identification of b_{int} is key to our paper, we explain in more intuitive detail how b_{int} is potentially identified in §5.2. It is noticeable that the key parameters (b_q, b_s, and b_{int}) are simultaneously estimated not just through likelihood maximization but also the unique pricing scheme causing the kinked budget constraint.

5. Results and Discussion

The estimation results by maximizing Equation (9) are given in Table 4. We estimate the model with a full sample and then split it along some demographic variables. The first three rows are the key structural parameters, and the rest are the estimates for distributional parameters.

We first focus on the distributional parameters (row 4 and below). Most of these estimates are consistent with our data. For example, μ_q and μ_s capture the mean voice and SMS consumptions, respectively, which match well with the means reported in the summary statistics in Table 2. These parameters are estimated precisely. $\sigma_{\varepsilon q}$ and $\sigma_{\varepsilon s}$ capture measurement error, and these estimates are fairly high, suggesting that the difference between actual and expected consumption is high—that is, users choose suboptimal plans (at least ex post). This is consistent with the prior literature (Miravete 2003). The intrinsic association between satiation points in the two services is significantly positively related ($\rho = 0.83$). That is, a heavy user of one service is likely to be the heavy user of the other service, regardless of substitution or complementary effect. This suggests that controlling for the intrinsic association of two services is important for an unbiased estimate on the cross-effects of two services.

The first three rows are the price response structural parameters. One of our key interests is the cross-effect estimate on b_{int}. The sign of b_{int} is negative, and it is statistically significant. The estimate is consistently negative across all demographic segments. This provides evidence that voice and SMS form a substitutive relationship in our data. The negative and significant

Table 4 Estimated Parameters

	Whole sample	Users with age < 30	30 ≤ age < 40	Age ≥ 40	Female	Male
b_{int}	−0.35**	−0.53**	−0.48**	−0.38**	−0.59**	−0.33**
	(0.07)	(0.11)	(0.05)	(0.07)	(0.19)	(0.03)
b_q	7.22**	5.27**	8.53**	9.24**	7.31**	8.51**
	(1.11)	(0.80)	(0.68)	(1.12)	(2.32)	(0.55)
b_s	0.12**	0.06**	0.08**	0.12**	0.04**	0.10**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
μ_q	283.11**	334.07**	272.81**	208.61**	290.01**	279.80**
	(1.26)	(2.35)	(1.14)	(2.76)	(1.54)	(1.40)
μ_s	11.58**	17.41**	10.49**	9.97**	15.24**	10.57**
	(0.27)	(0.31)	(0.73)	(0.47)	(0.25)	(0.21)
σ_q	134.82**	161.72**	147.39**	148.60**	136.75**	132.75**
	(1.03)	(5.29)	(5.17)	(2.95)	(2.06)	(1.19)
σ_s	23.03**	26.52**	17.37**	18.08**	22.44**	18.97**
	(0.11)	(0.09)	(0.18)	(0.29)	(0.05)	(0.05)
$\sigma_{\varepsilon q}$	200.45**	214.95**	165.30**	155.01**	174.13**	192.75**
	(0.14)	(0.15)	(0.24)	(0.44)	(0.06)	(0.12)
$\sigma_{\varepsilon s}$	17.83**	21.97**	11.58**	14.24**	17.12**	15.27**
	(0.003)	(0.01)	(0.003)	(0.01)	(0.004)	(0.003)
ρ	0.83**	0.84**	0.89**	0.63**	0.85**	0.71**
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
N	59,866	22,483	29,520	7,863	32,592	27,274
Log-likelihood	−606,520	−235,480	−299,250	−79,558	−332,790	−276,630

Note. Standard errors are shown in parentheses.
* Significant at $p < 0.05$; **significant at $p < 0.01$.

estimate indicates that (1) the real consumption levels, which we can observe, are lower than the satiation points, which we cannot see, in both services, and (2) the optimal consumption of voice (or SMS) decreases as the consumption of the other service goes up because the consumption in one service partially satisfies the satiation points in the other service. The estimates on b_q and b_s are also sensible and significant. Because the estimated b_q is significantly greater than b_s, it seems to suggest that the voice consumption is more sensitive to price change.

We next split the data based on demographic factors (age and gender) and analyze the difference in the key parameter estimates across subgroups. The results are shown in columns 3–7 of Table 4. Both price response parameters are significant in all groups though their magnitudes vary across groups. Younger users, on average, have higher estimated mean consumption in both services (μ_q, μ_s), consistent with the raw distribution of data. However, we also find a stronger substitution effect (bigger b_{int} in an absolute value) in the younger group than in the older group. Interestingly, there is less intrinsic correlation (ρ) between voice and SMS for older users than younger ones. Females are more likely to consume more SMS than males and have a higher estimated substitution effect. Also, notice that the estimated ρ for males is much smaller than females.

The estimated values of b_q, b_s, and b_{int} do not provide a clear picture of differences across groups. We now discuss the implication of b_{int} as well as the two price response parameters by calculating own- and cross-price elasticities.

5.1. Own- and Cross-Price Elasticities

To understand the true economic significance of estimated parameters, we calculate own- and cross-price elasticities using the estimates of b_q, b_s, and b_{int}. Because we need marginal prices for both voice and SMS to estimate elasticities, and a user experiences the marginal price of voice only if the user is in region R_k^3, we use the demand Equation (5) in region R_k^3 for estimating elasticities. Straightforward manipulation of the equation leads to the own-price elasticity of voice as $E_q = -3b_q/\bar{q}(1-b_{int}^2 b_s b_q)$ and the own-price elasticity of SMS as $E_s = -3b_s/\bar{s}(1-b_{int}^2 b_s b_q)$. The cross-price elasticity of voice with SMS (percentage changes in voice consumption in response to the percentage change in SMS price) can be calculated by $E_{q,s} = -3b_{int}b_q b_s/\bar{q}(1-b_{int}^2 b_s b_q)$, and the cross-price elasticity of SMS with voice can be calculated by $E_{s,q} = -3b_{int}b_q b_s/\bar{s}(1-b_{int}^2 b_s b_q)$.

Recall that our demand curves are linear, and therefore, the elasticities depend on quantities. Put another way, the reaction to proportional price changes will be different depending on the quantities consumed.

Table 5 Estimated Own- and Cross-Price Elasticities

	Mean-usage price elasticities		
	Mean	Own-price elasticity	Cross-price elasticity
Whole sample			
Voice	285	$E_q = -0.085$	$E_{s,q} = 0.078$
SMS	13	$E_s = -0.030$	$E_{q,s} = 0.003$
Users with age < 30			
Voice	330	−0.052	0.032
SMS	17	−0.011	0.001
30 ≤ age < 40			
Voice	265	−0.114	0.116
SMS	10	−0.028	0.004
Age ≥ 40			
Voice	204	−0.161	0.188
SMS	8	−0.053	0.007
Female			
Voice	287	−0.085	0.041
SMS	14	−0.009	0.002
Male			
Voice	274	−0.102	0.092
SMS	10	−0.033	0.003
Plan 1			
Voice	249	−0.097	0.100
SMS	10	−0.039	0.004
Plan 2			
Voice	643	−0.037	0.026
SMS	39	−0.010	0.001
Plan 3			
Voice	1,011	−0.023	0.022
SMS	44	−0.009	0.001
Plan 4			
Voice	1,556	−0.015	0.023
SMS	43	−0.009	0.0006
Plan 5			
Voice	1,994	−0.012	0.036
SMS	28	−0.014	0.0005

This is the reason we cannot directly infer the magnitude of substitution by just looking at the estimate for b_{int}. Also, note that the cross-price elasticities differ because of scaling. People use a significantly large number of voice minutes as opposed to the number of SMS. Because we observe the change in the marginal price of voice in our data (SMS prices remain unchanged), we will predominantly focus on the elasticity of SMS with voice ($E_{s,q}$). In Table 5, we report the elasticities at the mean values for voice and SMS consumption. Given the signs of the estimated parameters, the own-price elasticities are negative, and the cross-price elasticities are positive.

We start with the cross-price elasticities. The elasticity of SMS with voice is estimated to be small with a magnitude of about about 0.08. Roughly, this suggests that if the price of marginal voice minutes increases by 100%, the SMS consumption will increase by about 8%. Put another way, if the price of marginal minutes were to increase from three to six, the number of messages would increase by 1 message to about 14 messages (the mean number of messages is about 13, an 8% increase translates to about one message). Low values of cross-price elasticities suggest that although SMS acts as a substitute for voice, the magnitude of this substitution is *low*. It seems that the users in our sample view voice and SMS as differentiated services that serve different applications and needs. Therefore, a choice to send a message does not significantly diminish the need to make a voice call and vice versa.

The estimated own-price elasticities are small. The low values of own-price elasticities of voice and SMS suggest that users do not respond to price changes very aggressively (at least at the mean values). The numbers suggest that a 100% increase in the price of voice minutes would lead to about an 8.5% reduction in demand. Thus, if the price of voice minutes were to increase from three to four (a 33% increase), on average, it would lead to a reduction of about eight minutes. Low own-price elasticities are consistent with the previous research on land-line telephony usage. The price elasticities of fixed line demand in previous studies range from −0.1 to −0.5. Garbacz and Thompson (2003) report that the demand for local service is highly inelastic (−0.006 to −0.011). Park et al. (1983) use random experimental data and report that elasticity is about −0.1 in absolute value. Kling and Van der Ploeg (1990) show that the price elasticity of fixed phone service is in the range of −0.1. Andersson et al. (2006) report low own-price elasticity for mobile telephony (about −0.2). Generally, users exhibit low elasticity to more necessary services. The low number could be a reflection of the fact that mobile services are increasingly becoming necessary to users. In fact, in many countries, mobiles phones are becoming the only communication device users retain. Thus, low elasticities could be the manifestation of these changing structures.

SMS is far less elastic than voice service. This could be due to the relatively small consumption level—the price elasticity is calculated with not only price response parameters but also average usage (the average of SMS consumption is around 11 messages). That is, consumers are likely to be satiated with a relatively small number of messages and might respond to the marginal price of SMS far less than that of voice service.

At first blush, the low elasticity levels seem inconsistent with a familiar result from the neoclassical theory of the firm that a firm does not set prices in a region of its demand where the elasticity is less than one. However, there are a number of reasons why this is not so. First, the elasticity we are measuring is not the elasticity relevant to the standard result. Consumers might respond to an increase in our firm's prices both by reducing minutes consumed and by substituting the products of other firms. The standard result applies to the sum of these two effects, and we examine only the former of the two. It is easy to imagine the latter effect being the larger of the two. Another way of saying this is that we are not measuring the residual elasticity facing the firm (what the standard theory is about), but the market elasticity of the consumers who happen to have chosen this firm. Second, the standard result applies only to static models: in a setting with stickiness in demand or other interesting dynamics, the result need not apply, and firms may price at an inelastic point.

The results show that the price elasticities are different across age-based and gender-based segmented groups. The finding of a big variation in price elasticities across individual demographic profiles could be informative for practical managerial decision making. First, we find that the younger users are, on average, more inelastic than the older group for both the own- and cross-price elasticities. We do not have information regarding individual income. But, it is generally believed that the older users have a higher income, and they are expected to be more inelastic to price change. Therefore, this is somewhat counterintuitive and provides some insight into understanding mobile service. It could be that the parents are paying the bills for youngsters and that reduces their elasticity, or that the young users perceive mobile services as more necessary than the older users. Because the older users are more elastic, we conclude that when the marginal cost of SMS increases, the older users are more likely to change the communication tool from SMS to voice than the younger users.

The female users are more inelastic than the male users in both services. And the cross-price elasticity of voice in the female group is half that of the male group. Thus, it suggests that men are more likely to consider SMS as a substitute for voice than women.

Given our linear demand curve, the variation of own- and cross-price elasticities across plans is reasonable. As users select a higher plan, they show less sensitive price response behavior. This is to be expected because the average consumption is increasing in plans.

5.2. Alternative Evidence of Substitution Effect

One key goal of this paper is to identify the interaction of voice and SMS. We explained the details of our identification strategy behind b_{int} earlier. However, in a structural framework, there may be worries that nonlinearities in the likelihood function are driving identification. To provide more robustness to our analysis, we now provide more intuitive evidence of how SMS is a substitute for voice.

First, note that when users' voice consumption is below the free quota in their selected plans, users face zero marginal cost for voice but positive marginal cost for SMS. Once they exceed the free minutes, the voice minutes also incur a marginal cost. Thus, if SMS is a substitute (or complement), we should expect a significant jump (drop) in the SMS consumption when the free quota is exceeded. The free quota for plan 1 is 350 minutes. Thus, a user exceeding 350 minutes in plan 1 would pay for voice minutes. However, a user who signs up for plan 2 would not worry about the 350 minute limit because the free quota limit for plan 2 is 517 minutes. Therefore, we compare the user group in plan 1 with the user group in plan 2 and see how the SMS consumption changes when users exceed 350 minutes in both plans. More precisely, we run the following regression, where i indexes users and t indexes time:

$$SMS_{it} = \beta_0 + \beta_1 q_{it} + \beta_2 D_{FQ_k} + \beta_3 (q_{it} - FQ_k)^+ + \varepsilon_{it}, \quad (10)$$

where

$$D_{FQ_k} = 1 \quad \text{if } q_{it} \geq FQ_k, \quad \text{otherwise } 0;$$
$$(q_{it} - FQ_k)^+ = q_{it} - FQ_k \quad \text{if } q_{it} \geq FQ_k, \quad \text{otherwise } 0.$$

D_{FQ_k} captures the jump (drop) in SMS consumption before and after the free quota (or incurring a marginal cost per minute). Thus, a positive and significant estimate on this dummy for plan 1 users should provide the evidence of substitution effect. We report the results of this regression in Table 6. We run the regression for users selecting plan 1 and for users selecting plan 2. We confirm from Table 6 that there is an abrupt increase of SMS consumption at the kinked point (of 350 minutes) in only the user group selecting plan 1 (notice the significant coefficient ($p = 5\%$) of dummy D_{FQ_k} only for the plan 1 users and not for plan 2 users).

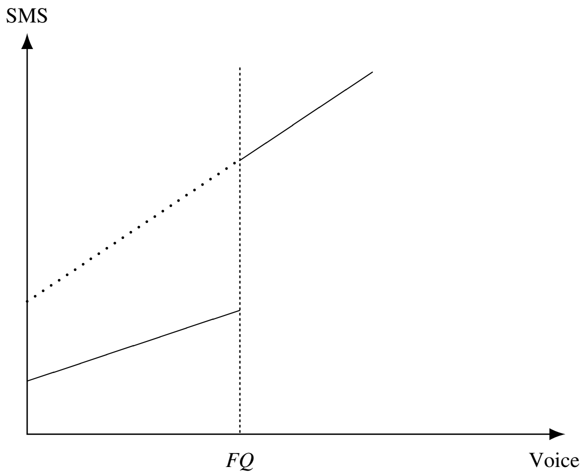
Figure 4 captures this jump in SMS consumption around the free quota boundary. SMS consumption increases with voice (positive correlation), but the

Table 6 Substitution Effect of SMS for Voice Service

	User group selecting plan 1	User group selecting plan 2
Voice	0.01**	0.12**
	(0.001)	(0.06)
D_{FQ_k}	1.03**	2.55
	(0.54)	(7.2)
(Voice-350)$^+$	0.013**	−0.12
	(0.002)	(0.06)
Constant	6.97**	−4.26**
	(0.38)	(15.66)
N	56,634	2,103

Note. Standard errors are shown in parentheses.
**Significant at $p < 0.05$.

Figure 4 Evidence of Substitutive Relationship of Voice and SMS



jump in SMS consumption around the free quota provides strong evidence of a substitutive relationship between voice and SMS.

We find the same results when considering the individual fixed effect. We tested the same model in the narrower range of voice consumption: $q_i = [150, 550]$. The dummy is still significant, providing robust support for substitution effect. In summary, we believe that the substitution effect identified in our structural model is robust and consistent with our data.

5.3. Policy Experiments

A key advantage of a structural model is that one can perform policy experiments and a "what-if" analysis. That is, we can analyze what happens to a firm's demand and, hence, revenue (or profit) if the fixed fee increases by 1% or if the firm decreases the free quota by 1%. One of the goals of this paper is also to provide some actionable recommendations to managers. Thus, conducting policy experiments provides useful information regarding the impact of change on the firm's revenues as well as profitability.

It should be noted that the nonlinear pricing and the kinked budget line make the interpretation of estimates difficult. In particular, a change in any of the key strategic (decision) variables cannot be analyzed in isolation because the change likely causes multiple changes in consumers' plan choices and consumption behavior. As a result, estimating price elasticity alone is not sufficient enough to highlight the impact of prices on consumer demand. For example, the change in the marginal cost of voice minutes will not only affect the demand for users consuming marginal minutes (region R_k^3) but will also force more users on kink point R_k^2. Therefore, the change will potentially affect the plan choice probability as well—i.e., a unit price increase in voice for a plan can impact the profitability of every plan. Thus, the change in marginal price

Table 7 Plan Choice Distribution

	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5
Data randomly generated (%)	0.96	0.04	0.00	0.00	0.00
Data collected (%)	0.95	0.04	0.01	0.00	0.00

of voice or SMS within or across plans is likely to have an indirect effect on the demand for all other plans and the associated consumption level.

Moreover, another key strategic variable available to the firm is the fixed fee of every plan. Similarly, the firm can increase or decrease the number of free minutes available in each plan. Because the fixed fee (or free minutes conditional on each plan) does not vary in our data, the impact of changes in fixed fee can *not* be readily estimated. But, our policy experiment based on a structural model outlined in our paper allows us to estimate the impact of a change in *all* the strategic variables more completely by simulating various counterfactual possibilities. Again, this analysis provides substantial advantages over a reduced-form model.

We perform various simulations. In particular, we examine how changes in the strategic variables affect the firm's revenue. Note that we outline only the impact on firm revenues. However, any other metrics, such as probability of plan choice or changes in demand, can also be readily calculated. The strategic variables we manipulate are (1) change in the marginal price of voice and SMS, both separately and simultaneously, (2) change in the fixed fee, and (3) change in the free quota minutes.

To carry out this exercise, we randomly generate around 5,000 of $(\theta_{iqt}, \theta_{ist})$ based on the estimated distribution-related parameters for policy experiments. Table 7 shows the plan choice distributions of both the generated data and the collected data, indicating that the model fits the data reasonably well. That is, the structural parameters are consistent with our large sample, and thus the data generated for the experiments are well representative of the "real-world" data we collected.

The first exercise is simply aimed at evaluating the effect of change in strategic variables without consideration of consumer differentiation across demographic segments. We calculated, for each generated $(\theta_{iqt}, \theta_{ist})$, the optimal plan choice and, subsequently, the optimal consumptions of both services. Given that, we can compute the firm's revenue from the expected consumption behaviors. We can then recalculate the revenues by changing any strategic variable. The upper part of Table 8 shows the change of revenue derived from one unit change of every strategic variable. But, a unit increase in the marginal price of voice is not directly comparable to a unit increase in

Table 8 Simulation of Change of Unit Pricing Scheme

	One-unit change of strategic variables				
	One-unit MP increase in both Services	One-unit MP increase in voice	One-unit MP increase in SMS	One-unit fixed fee increase	One-minute free quota decrease
Revenues increase (%)	2.94	0.87	2.06	0.12	0.15
	Normalized unit change of strategic variables				
		10% MP increase in voice	10% MP increase in SMS	5% fixed fee increase	5% free minute decrease
Revenue increase (%)		0.26	0.62	3.8	3.1

Note. MP, Marginal price.

the fixed fee due to scaling. Another possibility is to consider the percentage change in the prices. However, a 10% change in the marginal price of voice is not the same as a 10% change in the fixed fee. To keep the comparisons reasonable, we let the marginal prices of voice and SMS change by 10%, and the prices of the fixed fee and the free quota change by 5%.[11]

Users can respond to the change of the pricing scheme either by reducing the consumption level or by leaving the firm (i.e., switching mobile operators or giving up mobile communication). To incorporate the hypothetical defection behavior, we assume that consumers leave the firm when their surplus (utility level) from the mobile service goes below a certain threshold. To set this threshold, we rank order consumer surplus. The threshold is set at the lowest surplus level. As we change the parameters, if a consumer's surplus falls below the threshold, we consider the consumer to have left the firm. We applied this condition to all policy experiments we performed.[12] However, it must be noted that we cannot systematically and fully specify consumers' defection

[11] One could potentially calculate the equivalent fixed fee increase or free quota decrease for a unit price increase for marginal voice minutes. Based on the usage, individual users consume 39.75 additional voice minutes beyond the given free quota. A user pays an average price of 1.023 units for a voice minute. Therefore, a one-unit marginal price increase in the price of a voice minute is equivalent to increasing the fixed fee by 38.85 units or decreasing the free minutes quotes by about 39.75 minutes. Similarly, the average consumption of SMS is 11.8, and a one-unit increase of the per-minute marginal price for voice is comparable to about a 0.296 unit increase of per-message marginal price. We performed our simulations with this alternative specification as well, and results are comparable. They are available upon request from the authors.

[12] This approach is based on the assumption of cardinal utility and enables us to incorporate the negative repercussions of the change of strategic variables in the policy experiments. For example, a price change such as the increase of the fixed fee would make the mobile services more unattractive, at least for the users with low $(\theta_{iqt}, \theta_{ist})$.

Table 9 Policy Experiment Across Demographic Groups and Plans

	10% MP increase in voice (%)	10% MP increase in SMS (%)	5% fixed fee increase (%)	5% free minute decrease (%)
Age < 30				
Revenue increase	0.61	0.94	3.3	4
30 ≤ age < 40				
Revenue increase	0.17	0.44	3.80	2.91
Age ≥ 40				
Revenue increase	0.10	0.55	3.50	1.84
Female				
Revenue increase	0.08	1.07	3.80	1.54
Male				
Revenue increase	0.70	0.60	3.80	2.92
Plan 1				
Revenue increase			3.65	2.92
Plan 2				
Revenue increase			0.14	0.16

Notes. When conducting policy experiments in every segment, we generate $(\theta_{iqt}, \theta_{ist})$ based on the estimated parameters of the corresponding groups. MP, Marginal price.

behavior in our policy experiments because of lack of detailed data.[13]

Compared to change in the marginal prices of both services, change in the fixed fee (or free quota) has far greater impact on the firm's revenue. This is consistent with our general expectation. A change in the fixed fee or the free quota affects all consumers by increasing their fixed cost. However, change in the marginal cost of voice has little effect on the users not exceeding the given free quota—according to the data, only one-third of the observations (20,346) incur marginal costs across all plan. Furthermore, the results suggest that a decrease of free minutes can increase revenue, but the most effective strategy is to increase the fixed fee somewhat. A change in the marginal price of SMS has a bigger influence on the revenue than a change in the marginal price of voice. At a glance, this is inconsistent with the inference from the calculated price elasticities of demand. But the results are attributed to the difference of average consumption levels, which we highlighted earlier.

We further perform policy experiments across different demographic groups. We generated $(\theta_{iqt}, \theta_{ist})$ separately in every group based on the estimated parameters of the corresponding groups. Table 9 shows distinct differences in the effects of the changes of strategic variables across demographic segments as well as across selected plans.

[13] We essentially outline only small price changes, and one would expect that, for small price changes, consumers' defection probability is small. Thus, for small price changes, we treat the firm as a monopolist. So the policy experiments suggest how the change of pricing scheme will affect the direction of revenues.

The effect of a change in the marginal prices is far greater with younger users (age $<$ 30) than with older users (age $>$ 30), consistent with their calculated price elasticities. In particular, it is noticeable that the effect of an increase in the marginal price of voice is greater with younger users than with the others by around six times. We conjecture from this finding that younger users are more likely to incur marginal costs than older users.

A change in the marginal prices or the free minutes is more effective with younger users than with older users, whereas an increase in the fixed fee is the best strategy for older users. The simulation shows that the effect of a change in the free minutes is greatest in the younger user group. However, interestingly, the effect of the fixed fee increase is smallest in younger users. As a result, the firm can increase its revenue more by changing the fixed fee for the older user group than changing any other strategic variable. Younger users are likely to use as many free minutes as are available, and thus the decrease of free minutes turns out to be the most effective tool to increase revenue for that group.

Comparing the price elasticities between male and female groups, we found female users are more inelastic than male users in both services. However, the simulation shows that change in the marginal price for voice is more effective for males, whereas change in the marginal price of SMS is more effective for females. Thus, this result suggests that our policy experiments provide more useful information by accommodating the multiple changes caused by changing one variable.

Another exercise is to compare the impact of the changes in the fixed fee or the free quota across available plans. We increase (decrease) the fixed fee (free quota) of a selected plan by leaving the other plans unchanged. The simulation shows that an increase in the fixed fee is the most effective strategy to increase revenues, as shown in the lower part of Table 8. The majority of users select plan 1 or plan 2, and thus the effects of a change in the fixed fee (or the free minutes) in other plans are ignored.

5.4. Discussion and Implications of Our Findings

Our results offer two clear implications and contributions. The first contribution is academic. We offer a method to estimate own- and cross-price elasticities in the context of nonlinear tariffs involving two goods. Nonlinear tariffs are offered in many settings, and most popularly in wireless telephony. Firms are also increasingly offering multiple services. Because of the complexity of tariffs, the customer usage data available to firms and researchers is also complex. This precludes the use of simple demand estimates. Despite the complexity of the data, firms still need to tease out how multiple services interact and how elastic consumers are to prices. We believe that our paper extends the literature and provides a framework to analyze such data in a rigorous but tractable fashion. As the wireless market evolves and other 3G (third generation) data services become routine (e.g., Web browsing), our model can be extended to incorporate additional services and their interactive effect on voice and SMS.

Table 10 **Policy Experiment Regarding Bundling of Voice and SMS**

	Bundling of voice and free SMS					
Number of free messages	1	2	3	4	5	10
Revenue change (%)	−0.36	−0.71	−1.04	−1.36	−1.66	−2.99

In a similar fashion, our results have implications for firms offering a bundle of multiple services. Firms are increasingly offering bundled services and applications to users, especially in cellular markets. Bundling is used for price discrimination, increasing customer loyalty, as well as introducing new services. However, bundling multiple services also requires careful understanding of how those services interact. When introducing a new service, firms tend to give them away at no cost by bundling them with existing services. New services thus can be diffused effectively. However, such bundling requires careful considerations of whether the services are significant substitutes. If the services are complements or even independent, then such strategies can be highly effective (Andersson 2006). Otherwise, significant popularity of substitutive services offered at a lower price can have a large cannibalization impact on key revenue-earning services of a firm.

Table 10 shows the results of the policy experiments regarding the bundling strategy of voice and SMS. As the estimated substitutive relationship between two services indicates, the provision of "free" SMS cannot increase the consumption of voice, but free SMS decreases the consumption of voice. Therefore, the bundling of voice and free SMS reduces the firm's revenues.

Moving forward, these issues are going to be critical for firms as cellular platforms increasingly offer multiple services and applications, sometimes provided by third parties. Availability of multiple services and applications tends to increase the value of a cellular platform. However, they can also compete directly with the firm's own services. Understanding how a new application or a service would interact with the firm's portfolio is critical for a platform's profitability. This is one of the reasons a firm like Apple, for a long time, explicitly prohibited users from downloading voice over Internet protocol applications (like Skype), which directly cannibalize its voice services, onto their iPhones. Currently, iPhone and Google Voice are engaged in a similar battle. We believe that our model provides a useful framework to analyze these situations which are going to be increasingly commonplace.

Finally and more importantly, our results have useful managerial implications. As we pointed out, the complexity of nonlinear tariffs also makes it difficult to interpret how changes in key strategic variables such as marginal cost or fixed fee would impact firm demand and hence profitability. As outlined in Tables 8–10, our policy experiments provide a useful decision support to the firm. Specifically, our results can be used to design optimal tariffs by utilizing the results of the policy experiment. Given the user heterogeneity in our data set—most of the users have selected plan 1—the firm could have potentially exploited user heterogeneity better by offering more plans at the lower level. Considering everything, our results suggest that the firm's current tariff structure is potentially suboptimal.

6. Conclusion, Limitations, and Further Research Directions

Our paper seeks to examine the interaction of voice service and SMS provided by a wireless operator. The tariff structure offered by the operator is nonlinear, and the customer decision making is sequential (choice of a plan first and the consumption amounts of voice and SMS later). Moreover, customers consumes both voice and SMS conditional on a plan choice. We start by specifying a utility structure and derive consumer demand for these services. Our framework accounts for endogeneity of plan choice and subsequent consumptions. Thus, our model provides an estimable structure that takes into account the unique nature of wireless demand. We estimate the model with a unique and rich individual-level consumption data set, which has detailed information on each consumer's plan choice and the subsequent consumption quantities for voice and SMS. Using the data, we estimate, among other parameters, the cross-price effects of voice and SMS. Our model accounts for customer heterogeneity in terms of their inherent preferences for voice and SMS.

Our results show that voice service and SMS are Marshallian substitutes. However, the magnitude of substitution is small. We find that, on average, the cross-price elasticity of voice and SMS is low and about −0.08, that is, a 10% increase in the price of voice minutes would lead to about a 0.8% increase in SMS consumption. We also estimate the own-price elasticity of voice and SMS. We find that the own-price elasticity of voice is about −0.085, whereas the own-price elasticity of SMS (though it is not identified directly) is about −0.030. The low estimated elasticity for voice is comparable to what researcher have found in the land-line telephony studies (low price elasticities ranging from −0.1 to −0.2). We also find that, in many cases, consumers who prefer higher levels of voice consumption also have preferences for SMS consumption. In short, there is a strong inherent correlation between these services ($\rho = 0.83$). We highlight how controlling for this inherent preference is important for estimating the cross-effects. To test the robustness of our identification, we also perform an intuitive regression analysis. The regression uses the change in marginal price for voice service once the user exceeds the allowable free minutes of a selected plan. The estimates from the regression exercise confirm the substitutive relationship of voice and SMS.

We also explore how user sensitivity varies depending on demographic characteristics. We find that older users are more price sensitive in both services as well as exhibit higher cross-price elasticity. We find that there is little difference between the consumption patterns of males and females compared to age-based segmented groups. The extent of substitution effect and price elasticity somewhat depends on users' demographic factors. We then conduct some counterfactual policy experiments to examine how changes in key strategic variables affect firm profitability. We outline how our model allows us to estimate these effects. The last point is particularly important because the contribution of this paper goes beyond the rich data set we have collected. It is also the structural model we have built that allows us to identify and estimate key elasticity parameters that otherwise would not be possible with a naïve reduced-form model. Thus, our methodology provides us with insights into consumer behavior that otherwise would escape us. Thus, our research provides interesting avenues for future research in an important and growing field of cellular voice and data communication.

Although the richness of the data is a key strength, it is not without limitations. We do not have detailed information on consumer (disposable) income. Therefore, our analytical framework does not reflect how individual users allocate their monthly budget to mobile communication and outside goods. In reality, users are expected to make decisions of budget allocation by comparing the utility from mobile service and that from other goods. We cannot accommodate this feature in our model. Our data are also limited to one firm. In particular, the own-price and the cross-price elasticities should be interpreted with caution because of lack of competitor data. Similarly, our results should be interpreted as conditional on firm choice. Therefore, our measure is focused on only the consumer consumption reduction behavior in response to price increase—this limited research view is partly applied to the counterfactual policy experiments we conducted. Our data also lack variation in SMS pricing, making the estimates on price elasticity of SMS less credible. Our model also employs specific functional forms for the utility structure and assumes distributional forms for various parameters. Although many of these assumptions are widely used in the literature, our results should be interpreted appropriately.

Our current work is focused on voice and SMS in mobile service. As we outline in §5.4, wireless operators are increasingly offering multiple applications and services. Future work should analyze more prominent data services like wireless application protocol or application usage. Our model also ignores the network externalities. This would be another promising area to explore. We believe our paper takes a step in this direction and informs readers of the challenges of data in a complex setting and the models needed to analyze such data to provide insights into this market.

Acknowledgments

The authors thank the seminar participants at New York University, the University of Texas at Dallas, and the International Conference of Information Systems (2007) for valuable comments. Rahul Telang acknowledges generous support from the National Science Foundation through CAREER Award CNS-0540009. The authors also thank Atip Asvanund for help in data collection and tabulation and Ed Barr for editorial help.

References

Ahn, H., M. Lee. 1999. An econometric analysis of the demand for access to mobile telephone networks. *Inform. Econom. Policy* **11**(3) 297–305.

Albert, J. H., S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88**(422) 669–679.

Andersson, K., F. Steen, Ø. Foros. 2006. Text and voice: Complements, substitutes or both? CEPR Discussion Paper 5780, Centre for Economic Policy Research, London.

Burtless, G., J. Hausman. 1978. The effect of taxation on labor supply: Evaluating the gary negative income tax experiment. *J. Political Econom.* **86** 1103–1130.

Chiang, J. 1991. A simultaneous approach to the whether, what and how much to buy questions. *Marketing Sci.* **10**(4) 297–315.

Comms Dealer. 2008. Global revenues for non-voice services hits $188.7 bn in 2008, http://www.comms-dealer.com/mobile-zone/latest-news/global-revenues-non-voice-services-hits-1887bn-2008.

Economides, N., K. Seim, V. B. Viard. 2008. Quantifying the benefits of entry into local phone service. *RAND J. Econom.* **39**(3) 699–730.

Enriquez, L., S. Schmitgen, G. Sun. 2007. The true value of mobile phones to developing markets. *McKinsey Quart.* (September 17).

Garbacz, C., H. G. Thompson. 2003. Estimating telephone demand with state decennial census data from 1970–1990: Update with 2000 data. *J. Regulatory Econom.* **24** 373–378.

Garbacz, C., H. G. Thompson. 2007. Demand for telecommunication services in developing countries. *Telecomm. Policy* **31**(5) 276–289.

Gruber, H. 2001. Competition and innovation: The diffusion of mobile telecommunications in central and eastern Europe. *Inform. Econom. Policy* **13** 19–34.

Gruber, H., F. Verboven. 2001. The diffusion of mobile telecommunications services in the European Union countries. *Eur. Econom. Rev.* **45**(3) 577–588.

Grzybowski, L., P. Pereira. 2008. The complementarity between calls and messages in mobile telephony. *Inform. Econom. Policy* **20**(3) 279–287.

Hanemann, W. M. 1984. Discrete/continuous models of consumer demand. *Econometrica* **52** 541–561.

Hausman, J. 1999. Cellular telephone, new products and the CPI. *J. Bus. Econom. Statist.* **17** 188–194.

ITU World Telecommunication. 2009. ICT Indicators Database, 13th ed. International Telecommunication Union, Geneva. Accessed October 5, http://www.itu.int/ITU-D/ict/statistics/ict/index.html.

Iyengar, R. 2004. A structural demand analysis for wireless services under nonlinear pricing schemes. Working paper, Columbia University, New York.

Kling, J. P., Van der Ploeg. 1990. Estimating local call elasticities with a model of stochastic class of service and usage choice. A. de Fontenay, M. Shugard, D. S. Sibley, eds. *Telecommunications Demand Modeling: An Integrated View*. North-Holland, Amsterdam, 119–136.

Krishnamurthi, L., S. P. Raj. 1988. A model of brand choice and purchase quantity price sensitivities. *Marketing Sci.* **7**(1) 1–20.

Manchanda, P., A. Ansari, S. Gupta. 1999. The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Sci.* **18**(2) 95–114.

Miravete, E. J. 2002. Estimating demand for local telephone service with asymmetric information and optional calling plans. *Rev. Econom. Stud.* **69** 943–971.

Miravete, E. J. 2003. Choosing the wrong calling plan? Ignorance and learning. *Amer. Econom. Rev*. **93**(1) 297–310.

Miravete, E., L. Röller. 2004. Estimating markups under nonlinear pricing competition. *J. Eur. Econom. Assoc.* **2** 526–535.

Moffitt, R. 1990. The econometrics of kinked budget constraints. *J. Econom. Perspectives* **4** 119–139.

Niraj, R., V. Padmanabhan, P. B. Seetharaman. 2008. A cross-category model of households' incidence and quantity decisions. *Marketing Sci.* **27**(2) 225–235.

Park, R. E., B. M. Wetzel, B. M. Mitchell. 1983. Price elasticities for local telephone calls. *Econometrica* **51** 1699–1730.

Rodini, M., M. Ward, G. Woroch. 2003. Going mobile: Substitutability between fixed and mobile access. *Telecomm. Policy* **27**(5–6) 457–476.

Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Seetharaman, P. B., S. Chib, A. Ainslie, P. Boatwright, T. Chan, S. Gupta, N. Mehta, V. Rao, A. Strijnev. 2005. Models of multi-category choice behavior. *Marketing Lett.* **16**(3–4) 239–254.

Song, I., P. K. Chintagunta. 2007. A discrete/continuous model for multicategory purchase behavior of households. *J. Marketing Res.* **44**(4) 595–612.

Sung, N., Y. Lee. 2002. Substitution between mobile and fixed telephones in Korea. *Rev. Indust. Organ.* **20** 367–374.

Tammo, B., H. Van Heerde, R. Pieters. 2005. New empirical generalizations on the determinants of price elasticity. *J. Marketing Res.* **42**(2) 141–156.

Taylor, L. D. 1994. *Telecommunication Demand in Theory and Practice*, 1st ed. Kluwer Academic, Dordrecht, The Netherlands.

Taylor, L. D., D. J. Kridel. 1990. Residential demand for access to the telephone network. A. de Fontenay, M. Shugard, D. S. Sibley, eds. *Telecommunications Demand Modeling: An Integrated View*. North-Holland Amsterdam, 105–117.