The Pennsylvania State University

The Graduate School

College of the Liberal Arts

BUNDLING AND NONLINEAR PRICING IN TELECOMMUNICATIONS

A Dissertation in

Economics

by

Yao Luo

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2013

The dissertation of Yao Luo was reviewed and approved* by the following:

Isabelle Perrigne
Professor of Economics
Dissertation Adviser
Co-Chair of Committee

Paul Grieco
Assistant Professor of Economics
Co-Chair of Committee

Jonathan Eaton
Professor of Economics

Quang Vuong
Professor of Economics
Special Member

Fenghua Song
Assistant Professor of Finance

Robert Marshall
Professor of Economics
Head of the Department of Economics

*Signatures are on file in the Graduate School.

Abstract

This dissertation consists of three chapters.

Chapter 1: "**Nonlinear Pricing with Product Customization**"

This paper proposes a method to incorporate product customization and unobserved heterogeneity in structural analysis of nonlinear pricing data. My model is adapted from Maskin and Riley (1984). In equilibrium, average price per unit decreases with an index of quality/quantity, which aggregates observed attributes and unobserved heterogeneity of the product. I study the identification of the aggregation parameters and tariff function from data on payments and consumption quantities. My novel aggregation method leads to a problem reminiscent of transformation models. I then exploit both the firm's and consumer's first-order conditions to identify the model primitives, which are the consumer's utility function, taste distribution, distribution of product unobserved heterogeneity and firm's cost function. I also develop a semiparametric estimation method to recover these primitives. The estimation method is applied to data from a major mobile service provider in Asia. My empirical results support the model. Counterfactual experiments show that, with the same level of complexity, incremental discounts seem to be more preferable as it better approximates the second-best, leading to higher firm profit and consumer surplus relative to all-units discounts and quantity forcing.

Chapter 2: "**Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS**," with I. Perrigne and Q. Vuong

This paper studies multiproduct nonlinear pricing in the cellular phone industry with voice and message services. The model derives from Armstrong (1996) in which the unknown types of consumers are aggregated while the firm designs an optimal cost-based tariff. As

usual in multidimensional screening problems, there is pooling at equilibrium. Moreover, given that the consumers add a large number of unobserved extra features, we introduce two terms of unobserved heterogeneity for voice and message add-ons. The model defines two one-to-one mappings between the unknown aggregate type to the cost and the cost to the payment. We then study the identification of the model primitives. Under some identifying assumptions such as a parameterization of the cost function, we show that the primitives of the model (the aggregate type density, the indirect utility and the joint density of unobserved heterogeneity) are identified from observables. The empirical results support the model and display an important heterogeneity in types and unobserved heterogeneity. The cost of asymmetric information for the firm is assessed.

Chapter 3: "**Bundling and Nonlinear Pricing in Telecommunications**"

This paper studies bundling and price discrimination by a multiproduct firm selling internet and phone services in an imperfect information setting. I derive the optimal selling mechanism, and provide primitive conditions under which different bundling strategies arise, such as component pricing, pure bundling, semi-mixed bundling and mixed bundling. I show that the model structure is nonparametrically identified. I then propose a three-step semiparametric estimation procedure involving a new regression spline estimator under both monotonicity and bound restrictions. An illustration on China Telecom data shows that mixed bundling is beneficial to both the firm and the consumer relative to component pricing.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express the deepest appreciation to my advisors Professors Isabelle Perrigne and Quang Vuong, who are always by my side to make sure I make it through the difficult. Without their guidance and persistent help this dissertation would not have been possible.

I am indebted to Dr. Paul Grieco for his insightful comments on the thesis and consistent encouragement, and to Dr. Neil Wallace for valuable suggestions. I am grateful to Dr. Jonathon Eaton and Dr. Fenghua Song for serving as my committee members and providing helpful advices. I would like to thank Dr. Herman Bierens, Dr. Russell Cooper, Dr. Edward Green, Dr. Barry Ickes, Dr. Sung Jae Jun, Dr. Kala Krishna, Dr. Vijay Krishna, Dr. Robert Marshall, Dr. Peter Newberry, Dr. Joris Pinkse, Dr. Mark Roberts and Dr. James Tybout for their help during various stages of my graduate study. I owe my gratitude to Dr. John Riew for his continued support. In addition, a thank you to Dr. Junji Xiao, who introduced me to empirical research.

I am grateful to my classmates and friends for helpful discussions and encouragement. A special note of thanks to Liang Chen and Ruichang Lu for the many ways I have learned from them. Frank Erickson and Felix Tintelnot have been great housemates for three years as well as colleagues. I would also like to thank Robert C. Daniels, Bates White and NET Institute for providing financial assistance.

Last but not least, I thank my fiancée for always enduring my stresses and struggles, yet continually offering support, wisdom, encouragement and great food.

Dedication

This thesis is dedicated to my parents, for love, encouragement, and support all these years.

Chapter 1

Nonlinear Pricing with Product Customization

A commodity is a good or a service completely specified physically, temporally, and spatially.

– Debreu (1959)

1.1 Introduction

In the early 1900s, Ford launched mass production with its revolutionary assembly line for the Model T car. Today modern technology is starting an era of mass customization. With flexible computer-aided systems, firms deliver precisely the good or service that a customer wants, as and when he wants it. Ford now allows its customers to build a vehicle from a palette of online options. Computers, watches and mobile phone services are other examples. Products are adapted to meet a customer's individual needs, so no two items are the same. This leads to a large number of varieties of which many may exhibit sparse or zero market share. Discrete choice models become intractable. How to analyze data with a large number of varieties of products? Moreover, the researcher may not observe all characteristics. How to incorporate product unobserved heterogeneity?

This paper proposes a method to incorporate product customization and unobserved heterogeneity in structural analysis of nonlinear pricing data. My model is adapted from Maskin and Riley (1984). In equilibrium, average price per unit decreases with an index of quality/quantity, which aggregates observed attributes and unobserved heterogeneity of the product. I study the identification of the aggregation parameters and tariff function from data on payments and consumption quantities. My novel aggregation method leads to a problem reminiscent of transformation models. I then exploit both the firm's and consumer's first-order conditions to identify the model primitives, which are the consumer's utility function, taste distribution, distribution of product unobserved heterogeneity and firm's cost function. I also develop a semiparametric estimation method to recover these primitives. The estimation method is applied to data from a major mobile service provider in Asia.

Most of the empirical literature on nonlinear pricing uses discrete choice models while considering prices exogenous. See, e.g. Leslie (2004), McManus (2007), Cohen (2008) and Economides, Seim, and Viard (2008). Several papers employ convenient parametric assumptions to endogenize the optimal tariff schedules and recover the demand and cost structure. See, e.g. Ivaldi and Martimort (1994), Miravete (2002), Miravete and Röller (2004) and Crawford and Shum (2007). While endogenizing the price, Perrigne and Vuong (2011a) study nonparametric identification and estimation of a nonlinear pricing model with a single attribute and known tariff schedule. My paper provides an alternative method when the tariff schedule is not observed and there are multiple attributes.

As discussed above, a mobile service provider has the technological infrastructure that allows consumers to customize their own services and use them in various quantities. The firm I obtained data from provides a full range of different kinds of mobile services that can be temporally and spatially categorized, such as local outgoing calls, local incoming calls, long distance calls and so on. I observe only several quantities of phone calls. The monthly bill, however, includes a large number of additional services such as roaming, phone rings, etc. Thus I introduce a term of product unobserved heterogeneity to explain the data. The model is written with the effective quantity aggregating observed quantities and product unobserved heterogeneity. In equilibrium, tariff is an increasing and concave function of this index of quantity.

I first study the identification of the aggregation parameters and tariff function, which is reminiscent of transformation models and single index models. Specifically, I have a model of the form $\alpha'\mathbf{q} = \log T^{-1}(t) - \log \epsilon$, where $\mathbf{q}$ is the natural logarithm of the vector of quantities I observe, t is the payment and ϵ is the term of product unobserved heterogeneity. I show that the aggregation parameters and the tariff function are identified up to scale and location normalization under a rank condition.

Next, I study the identification of model primitives using the tariff function and payment distribution. Following Perrigne and Vuong (2011a), I rely on the first-order conditions of both the firm and the consumer. Under a parameterization of the cost function and a multiplicative separability of the utility function in the willingness-to-pay, I show that the primitives are identified. In particular, I rewrite the first-order conditions in order to express the one-to-one mapping from payment to consumer taste in terms of the tariff function and payment distribution. With this in hand, I translate the tariff function into the utility function, and the payment distribution into the taste distribution.

The estimation procedure proposed in this paper follows the steps of the identification arguments. Using data from a major mobile service provider in Asia, I estimate the consumer's utility function and the distributions of private information and product unobserved heterogeneity as well as firm's cost parameters. My empirical results support the model. Due to asymmetric information, the average informational rent left to consumers is \$10.37, which is approximately 38% of the average payment. With the estimated model primitives, I conduct various experiments to investigate the implications of alternative pricing policies: incremental discounts, all-units discounts and quantity forcing. It appears that the firm does not lose much relative to nonlinear pricing by using simpler price schemes, each of which can be characterized by four parameters. The three alternative schemes reduces firm profit by 3.17%, 6.53% and 16.12%, respectively. Also, I show that, with the same level of complexity, incremental discounts seem to be more preferable as it better approximates the second-best, leading to higher firm profit and consumer surplus relative to the other two.

The rest of the paper is organized as follows. Section 1.2 describes the data with a particular attention to aspects that are incorporated in the model. Section 1.3 presents the model. Identification and estimation are discussed in Section 1.4. Section 1.5 presents the estimation results and some counterfactuals. Section 1.6 concludes with future lines of research.

1.2 Mobile Phone Service Data

I collected data from a major mobile service provider in a major metropolitan area in Asia. For the billing period of May 2009, the data contain mobile voice services and the amount paid by each subscriber. Among the three mobile service providers allowed to operate in this area, the company from which I obtained the data has 72% of the mobile subscribers. This company proposes three different tariffs to three different types of consumers: students, rural residents and urban residents. My data concern urban residents only. The company's market share in this market segment is larger than 80%. Thus it is reasonable to assume that this company acts as a monopolist in this market segment.

The company proposes a new tariff every year. Subscribers can switch to the new tariff at no extra cost. I consider customers who are under the 2009 tariff in May 2009. I obtain a random sample of 2000 observations. The company distinguishes outgoing and incoming calls according to the features of the calls, such as the locations of the initiator and receiver, the length of the call, the time of the day, the day of the week and so on. The tariff changes with these features.

My data provide information on voice consumption measured in minutes and distinguishes three different types of phone calls: q_L is the number of minutes when both parties are in the same city, while q_D is when they are in different cities. q_R is the number of minutes when the consumer is outside his home city. Table A.1 provides summary statistics on the payment t measured in U.S. dollars and the consumptions of the three types of calls measured in minutes. The average payment is 27 dollars. As subscribers mainly stay in the home city, they consume more than seven times more local minutes than roaming minutes.

The firm implements nonlinear pricing, i.e., deeper discounts are offered to the customers who consume more. According to my conversations with several employees in this firm, there are literally hundreds of different discounts offered to consumers. Thus, the firm offers a large number of plans to approximate a continuous nonlinear price schedule, which leaves almost no room for pooling. For this reason, I use a continuous framework. Specifically, I consider that the tariff schedule offered by the monopolist is continuous and estimate it from data on payments and consumption quantities.

Product Customization and Unobserved Heterogeneity

As discussed above, the tariff changes with the temporal and spacial features of the calls. However, I do not observe all the temporal and spacial features of the calls except the number of minutes for the three types of phone calls. To see this, I sample consumers whose payments fall into chosen bins. Table A.2 provides the summary statistics of the total phone minutes (i.e., $q_L + q_D + q_R$) and payments when controlling for several categories of the payment value. While the variation of payment is controlled to be small in each bin, there is still large variations in the total minutes. Moreover, I regress payment on q_L, q_D, q_R and their squares. The adjusted R^2 of this regression is 0.58 in comparison to the almost perfect fit using data from yellow page industry (see, e.g., Perrigne and Vuong (2011a) and Busse and Rysman (2005)). Therefore, it is important to take into account product unobserved heterogeneity in my structural analysis.

1.3 The Model

My model builds on Maskin and Riley (1984). A consumer is characterized by a scalar taste parameter θ distributed as $F(\cdot)$ with a continuous density $f(\cdot) > 0$ on $[\underline{\theta}, \overline{\theta}] \subset I\!R_+$.[1] A subscriber consuming quantity of mobile services q has a payoff

$$U(q, \epsilon; \theta) - t(q, \epsilon),$$

where $t(q, \epsilon)$ is the total payment for consuming q when the other unobserved (by the econometrician) features of consumption is ϵ. The scalar term q aggregates the observed

[1] I do not consider uncertainty on types, which leads to a two-stage model. See, e.g., Miravete (2002), Miravete (2005), Narayanan, Chintagunta, and Miravete (2007), Economides, Seim, and Viard (2008) and Grubb and Osborne (2012).

attributes of the product (i.e., quantities of phone calls measured in minutes), which I will specify at the end of this section. The term θ is his taste for mobile services, which is his private information. The term ϵ represents all the product unobserved heterogeneity which is common knowledge to the firm and the consumer. In particular, it takes into account location and time of phone calls as well as all the extra features associated with mobile services.[2] As each subscriber can buy mobile services matching his exact need, ϵ is potentially different across consumers. It plays a similar role as the term of unobserved product characteristics ξ in estimating discrete choice demand (see, e.g., Berry (1994), Berry, Levinsohn, and Pakes (1995), and Berry and Pakes (2007)). Without ϵ it is impossible to find parameter values that make the implications of the model consistent with the data.

The firm incurs a cost $c(q, \epsilon)$ in serving the consumer with q with additional features ϵ. I make the standard assumption that the firm's total cost function is additively separable across consumers. As information goods, mobile services involve very small variable production costs but substantial transaction costs per customer, such as usage recording, billing and customer service. As ϵ is both observed (by the firm) and contractible, the firm chooses the quantity schedule $q(\cdot;\epsilon)$ and the tariff $t(\cdot;\epsilon)$ to maximize its profit conditional on ϵ. Given ϵ, the firm's problem is

$$\max_{q(\cdot;\epsilon),t(\cdot;\epsilon)} \int_{\underline{\theta}}^{\overline{\theta}} \left[t\big(q(\theta;\epsilon);\epsilon\big) - c\big(q(\theta;\epsilon);\epsilon\big)\right]\phi(\theta|\epsilon)d\theta$$

where $\phi(\cdot|\epsilon)$ is the conditional density of θ given ϵ.

Further restriction on the primitives is necessary for identification. Considering a separable cost function, D'Haultfoeuille and Février (2007) show that at least one of their three primitives, namely the surplus function, the taste distribution or the cost function, needs to be known to achieve identification.[3] While different identifying assumptions can be entertained, the production of mobile services tends to involve high fixed costs and small marginal costs. A linear cost function seems to be a good approximation. Thus, I make the

[2] For example, if a mobile phone is turned on, its geographical location can be easily determined by calculating the differences in time for a signal to travel from the mobile phone to each of several cellular towers nearby.

[3] In contrast, Perrigne and Vuong (2011a) assume a cost function for the total production in the context of yellow pages. I note that their *total* cost function and my *separable* cost function are nonnested.

following assumptions on the model structure.

Assumption 1. *The utility, cost and density functions satisfy*

(i) $U(q;\theta,\epsilon)=\theta u_0(q\epsilon)$, *where* $u_0(0)=0$, $u_0'(\cdot)>0$, $u_0''(\cdot)<0$

(ii) $c(q;\epsilon)=K+\gamma q\epsilon$,

(iii) $\theta-\frac{1-F(\theta)}{f(\theta)}$ *is increasing in* θ,

(iv) $\theta\perp\epsilon$.

Following the literature (see, e.g., Ekeland, Heckman, and Nesheim (2004) and Perrigne and Vuong (2011a)), I assume multiplicative separability of the utility function in the type θ as stated in Assumption 1-(i). Thus, I interpret $u_0(\cdot)$ as the base utility function. It is a standard assumption in the nonlinear pricing literature. Moreover, I interpret the term ϵ as a quantity multiplier. It captures the subscriber's usage of extra features and add-ons such as roaming. Thus, it is a "vertical" characteristic in the sense that every consumer would prefer more of it. Consumers with larger values of ϵ enjoy a larger utility from their consumptions of mobile services than subscribers with lower values. For example, consumers who travel more would consume more roaming minutes, which are of greater convenience than the same amount of local minutes. Treating ϵ as a quantity multiplier plays a similar role as setting the coefficient of unobserved product characteristic ξ to be one for all consumers in discrete choice models (See, e.g., Berry and Pakes (2007)). Finally, the outside option (not buying) provides a zero utility and the marginal utility is positive and decreasing. It can be easily seen that the standard Spence-Mirrlees single-crossing condition is satisfied. Namely, a consumer with a higher taste θ enjoys a larger marginal payoff across every q.

Assumption 1-(ii) says that the cost function is linear. The term K captures the cost that is triggered by any positive usage. For example, infrastructure costs stem from keeping mobile phones connected and administration costs stem from delivering statements. The variable cost $\gamma q\epsilon$ arises from monitoring, recording and reporting mobile service usages. Product customization is costly. For example, roaming as captured by ϵ is costly to the firm because more cellular sites and mobile-services switching centers are involved and it requires more financial settlements between the two providers involved. Assumption 1-(iii) says that the hazard rate does not decline too rapidly as θ increases. It is a standard assumption in the nonlinear pricing literature. Most commonly used unimodal distributions satisfy the hazard rate assumption.

While the independence assumption 1-(iv) is strong, it greatly simplifies the optimal selling mechanism. Namely, I can solve this problem in terms of "effective quantity" $Q \equiv q\epsilon$.

Proposition 1. *Under Assumptions 1, the functions $(q(\cdot;\epsilon), t(\cdot;\epsilon))$ that solve the monopolist's optimization problem satisfy: there exists a cutoff taste $\theta_0 \in [\underline{\theta}, \overline{\theta}]$ such that consumers with $\theta < \theta_0$ are not served by the provider, and whenever $q(\cdot;\epsilon) > 0$,*

(i) There exists a pair of functions $(Q(\cdot), T(\cdot))$ such that: $\forall \theta \in [\theta_0, \overline{\theta}]$

$$T'(Q(\theta)) = \theta u_0'(Q(\theta)), \tag{1}$$

$$\theta u_0'(Q(\theta)) = \gamma + \frac{1-F(\theta)}{f(\theta)} u_0'(Q(\theta)), \tag{2}$$

(ii) $\forall \theta \in [\theta_0, \overline{\theta}]$ and $\forall \epsilon \in [\underline{\epsilon}, \overline{\epsilon}]$, $q(\theta;\epsilon) = Q(\theta)/\epsilon$ and $t(q;\epsilon) = T(q\epsilon)$.

The cutoff taste is defined by

$$\theta_0 = \min \{\theta \in [\underline{\theta}, \overline{\theta}] : \theta u_0(Q(\theta)) - \gamma Q(\theta) - u_0(Q(\theta)) \frac{1-F(\theta)}{f(\theta)} \geq 0\}, \tag{3}$$

where $Q(\cdot)$ is defined by (2).

The proof of Proposition 1 follows Maskin and Riley (1984) and Sundararajan (2004). The optimal schedule and tariff $Q(\cdot)$ and $T(\cdot)$ are defined by Equations (1) and (2), along with a boundary condition

$$T(Q(\theta_0)) = \theta u_0(Q(\theta_0)) \tag{4}$$

because the informational rent left to the threshold consumer is zero. In particular, (1) says that the marginal utility equals the marginal price at the designated consumption of each subscriber. (2) says that the marginal utility equals the marginal cost plus a distortion term due to incomplete information. (3) balances the marginal again for expanding customer base and the marginal loss for reducing the tariff to every consumer above the cutoff.

Note that $Q(\cdot)$ is strictly increasing in θ. Moreover, Maskin and Riley (1984) indicates that $T'(\cdot) > 0$ and $T''(\cdot) < 0$. Therefore, there is a unique strictly increasing mapping between the unobserved taste θ and the observed bill t, which is the key of my second identification result. Finally, it is easy to see that if $K > 0$, $Q(\theta_0) > 0$. Thus the optimal tariff is a nonlinear two-part tariff. There is a minimum price $T(Q(\theta_0))$ for usage above 0 but lower than $Q(\theta_0)$, and a variable price beyond that.

Aggregating Quantities of Phone Calls

My data provide the quantities q_L, q_D, q_R of phone calls measured in minutes. I aggregate these quantities into a phone usage index $q = h(q_L, q_D, q_R)$. In particular, I make the following assumption on the aggregation function.

Assumption 2. *$h(\cdot,\cdot,\cdot)$ is of the form*

$$h(q_L, q_D, q_R) = q_L^{\alpha_L} q_D^{\alpha_D} q_R^{\alpha_R},$$

where $\alpha_L, \alpha_D, \alpha_R \geq 0$.

Although I could allow a more general function form, a Cobb-Douglas specification leads to an intuitive method to identify and estimate my model.[4] Under Assumption 2, the optimal tariff becomes

$$t = T(q_L^{\alpha_L} q_D^{\alpha_D} q_R^{\alpha_R} \times \epsilon),$$

where $T(\cdot)$ is strictly increasing and concave. Considering the inverse and taking the logarithm gives

$$\alpha_L \log q_L + \alpha_D \log q_D + \alpha_R \log q_R = \log T^{-1}(t) - \log \epsilon.$$

[4]Cobb-Douglas specifications have been widely used in many empirical studies to aggregate multiple quantities. For example, Murphy (2007) uses a Cobb-Douglas specification to aggregate characteristics of houses into an one-dimensional quality index. Moreover, first introduced by Solow (1957), the Cobb-Douglas specification is also extensively used for aggregating production function. In Consumption-Based CAPM models, the Cobb-Douglas specification is used to construct a consumption index for the representative agent. See, e.g., Dunn and Singleton (1986).

My model can be understood in the following way. The firm has a technology infrastructure that allows subscribers to customize their own products and use them in various quantities. I define a subscriber's exogenously determined variant of the product as a bundle of characteristics $v = (v_L, v_D, v_R, v_\epsilon)$, namely, the proportions of different kinds of phone calls and additional features. His consumption possible set is give by the ray in $I\!R^4_+$, $\mathcal{Q} \equiv \{X \in I\!R^4_+ : (xv_L, xv_D, xv_R, xv_\epsilon),\text{ where } x \in I\!R_+\}$. I obtain his effective quantity $Q = x^{\alpha_L+\alpha_D+\alpha_R+1} v_L^{\alpha_L} v_D^{\alpha_D} v_R^{\alpha_R} v_\epsilon$. While I take a subscriber's variant as given, my model endogenizes the quantity level and his payment.

Under Assumption 2, I obtain $t = T(q_L^{\alpha_L} q_D^{\alpha_D} q_R^{\alpha_R} \times \epsilon)$, where $T(\cdot)$ and α are unknown. This is reminiscent of single-index and transformation models. The novel element is that the error term enters in the unknown function as a quantity multiplier for its argument. While the error term is set at zero when calculating a hedonic price, the tariff function here incorporates unobserved heterogeneity. An alternative is to use a linear specification, namely $Q = (\alpha_L q_L + \alpha_D q_D + \alpha_R q_R) \times \epsilon$. The optimal tariff becomes $t = T\big((\alpha_L q_L + \alpha_D q_D + \alpha_R q_R)\epsilon\big)$. Considering the inverse and taking the logarithm gives $\alpha_L q_L + \alpha_D q_D + \alpha_R q_R = \log T^{-1}(t) - \log \epsilon$. My analysis still applies.

Finally, the effective quantity of mobile services is the index $Q = q_L^{\alpha_L} q_D^{\alpha_D} q_R^{\alpha_R} \times \epsilon$. All consumers agree on this quantity ranking. The reason consumers differ in their choices of Q is that they have different marginal utilities of income θ. In a random coefficient model such as Berry, Levinsohn, and Pakes (1995), consumers differ in their tastes for different product attributes. However, allowing heterogeneous tastes for different product attributes is out of the scope of this paper. It leads to a difficult multidimensional screening problem in which the firm's optimal selling mechanism is hard to derive (See, e.g., Armstrong (1996), Rochet and Chone (1998), Luo, Perrigne, and Vuong (2012) and Luo (2012)).

1.4 Identification and Estimation

In this section, I study identification of the model primitives. I then propose a multistep estimation procedure in view of my identification results.

In view of Section 1.3, the model primitives are $[u_0(\cdot), f(\cdot), g(\cdot), \alpha, K, \gamma]$, namely the base utility function, the taste density function, the density function of unobserved heterogeneity, the weights used to aggregate quantities of phone calls, fixed cost and variable cost parameters. Denote $\alpha \equiv (\alpha_L, \alpha_D, \alpha_R)'$ and $\mathbf{q} \equiv (\log q_L, \log q_D, \log q_R)'$. The tariff function gives

$$\alpha' \mathbf{q} = \log T^{-1}(t) - \log \epsilon, \tag{5}$$

where $Q(\cdot)$ and $T(\cdot)$ are defined by Equations (1) and (2), along with two boundary conditions (3) and (4).

In view of Section 1.2, the observables are the subscribers' consumed quantities of various phone calls and their bills, which include a large set of add-ons and additional features. The observables are denoted by q_L, q_D, q_R and t. The vector (q_L, q_D, q_R, t) is distributed as $G(\cdot,\cdot,\cdot,\cdot)$.

1.4.1 Identification

I proceed in several steps. First, I study the identification of the weights to aggregate quantities of phone calls, the tariff function $T(\cdot)$ and the distribution of product unobserved heterogeneity $G(\cdot)$ using Equation (5). Second, I exploit the one-to-one mapping between observed payment and unobserved taste to identify the taste distribution and base utility function using Equations (1) and (2).

Identification of α, $T(\cdot)$ and $G(\cdot)$

(5) is a special case of the model of Ai and Chen (2003). Since the finite-dimensional parameter α and the infinite-dimensional parameter $\Lambda(\cdot) \equiv \log T^{-1}(\cdot)$ are additively separable, it is also reminiscent of transformation models and single index models.[5]

Equation (5) continues to hold if α, Λ, and $-\log\epsilon$ are replaced by $k\alpha$, $k\Lambda$, and $-k\log\epsilon$ for any $k > 0$. It also holds if Λ and $-\log\epsilon$ are replaced by $\Lambda + k$ and $-\log\epsilon - k$ for

[5]The transformation model is $\Lambda(Y) = X'\beta + e$, where Y is a scalar dependent variable, $\Lambda(\cdot)$ is a strictly increasing function, X is a vector of explanatory variables, β is the vector of corresponding coefficients, and e is an unobserved error term independent of X. The semiparametric single index regression model is $Y = \Lambda(X'\beta) + e$, where Y is a scalar dependent variable, $\Lambda(\cdot)$ is an unknown link function, X is a vector of explanatory variables, β is the vector of corresponding coefficients, and e is an unobserved error term independent of X.

any $k \in I\!R$. Therefore, location and scale normalizations are needed for identification. The independence between ϵ and θ implies that t is independent with ϵ. Let α_0 be the true aggregation parameters. Given t, the variability in $\alpha_0'\mathbf{q}$ is equal to the variation in ϵ. This observation does not necessarily hold for $\alpha \neq \alpha_0$. Utilizing this homoskedasticity property, α are identified under a rank condition. Once α is identified, my model becomes reminiscent of transformation model and is identified under a mean independence assumption.

Assumption 3. *I assume*

(i) $\alpha_L + \alpha_D + \alpha_R = 1$,

(ii) for some $t_1, t_2, \ldots, t_J \in [\underline{t}, \overline{t}]$, *the following homogeneous quadratic system has an unique solution:*

$$\alpha' \left[\mathrm{Var}\left(\mathbf{q}|t_{\tilde{i}}\right) - \mathrm{Var}\left(\mathbf{q}|t_i\right) \right] \alpha = 0,$$

where $\tilde{i}, i = 1, 2, \ldots, J$.

(iii) $\mathrm{E}\left(\log \epsilon\right) = 0$.

Assumption 3-(i) normalizes the scale of α. It says that the aggregation function has constant return to scale. $\mathrm{Var}\left(\mathbf{q}|t_i\right)$ is the conditional variance covariance matrix of the random vector $\mathbf{q}$ given t_i. Assumption 3-(i) and (ii) lead to the identification of α. Assumption 3-(iii) says that $\log \epsilon$ has mean zero, which leads to the identification of $T(\cdot)$. With this centering assumption on $\log \epsilon$, there is no location assumption on $\Lambda(\cdot)$ as in Horowitz (1996). I interpret $T(\cdot)$ as the tariff function for the consumer using an average amount of additional features, namely $\epsilon = 1$.

Proposition 2. *Under Assumption 3, α is identified. In addition, the tariff function $T(\cdot)$ and the distribution of unobserved heterogeneity $G(\cdot)$ are identified on $[\underline{t}, \overline{t}]$ and $[\underline{\epsilon}, \overline{\epsilon}]$, respectively.*

The proof of Proposition 2 builds on the idea of Horowitz (1996). The basic idea is to exploit the separability of Equation (5). Let $\tilde{G}(\cdot|t)$ be the CDF of $Y \equiv \alpha'\mathbf{q}$ conditional on

t. From the independence between ϵ and t, one obtains $\Lambda'(t) = -\tilde{G}_t(y|t)/\tilde{G}_y(y|t)$. Then by imposing the normalization $\mathrm{E}(\log \epsilon) = 0$, one obtains $\Lambda(t) = \mathrm{E}(y|\underline{t}) - \int_{\underline{t}}^{t} \tilde{G}_t(y|x)/\tilde{G}_y(y|x)dx$. Once α and $T(\cdot)$ are identified, the unobserved heterogeneity of a consumer with payment t and consumption $\mathbf{q}$ can be identified as $\epsilon = T^{-1}(t)/\exp(\alpha'\mathbf{q})$. Thus, the distribution of unobserved heterogeneity $G(\cdot)$ is also identified.

To conclude this subsection, I compare my aggregation method with the quality-adjusted quantity method used in Perrigne and Vuong (2011a). They obtain the tariff function $T(\cdot)$ from an OLS regression using the multicolor price schedule for display advertisements. As they get an $R^2 = 0.999$, they treat it as known in the estimation. Quality-adjusted quantities can be constructed using $Q = T^{-1}(t)$ where t is observed.

In view of Section 1.2, I only observe the monthly paid price and quantities of phone calls. The tariff function $T(\cdot)$ needs to be estimated from data on payments and quantities. The monthly bill, however, includes a large number of add-ons and additional features such as roaming, phone rings, etc. Regressing payment on quantities does not give a good fit. Thus it is important to introduce a term of unobserved heterogeneity to explain the data. I identify and estimate $T^{-1}(\cdot)$ using (5). Effective quantities are constructed using $Q = T^{-1}(t)$.

Identification of K and γ

I now show that K and γ are identified at the minimum and maximum amount purchased, respectively.

Lemma 1. *The parameters K and γ are identified. In particular,*

$$\gamma = T'(T^{-1}(\bar{t})),$$
$$K = \gamma[\frac{\underline{t}}{T'(T^{-1}(\underline{t}))} - T^{-1}(\underline{t})],$$

where $\underline{t}$ and $\bar{t}$ are the minimum and maximum payments, respectively.

Identification of γ comes from the "no distortion at the top" property, i.e., the marginal willingness to pay equals the marginal cost for the highest taste consumer. Since there is exclusion in my data, K is identified using (3) with equality. The second equation implies that the cost for the cutoff consumer is smaller than his payment. If it were in a complete information scenario, the firm would expand his customer base because the marginal revenue is higher than marginal cost. In an incomplete information case, the firm optimally allows such a difference at the cutoff taste to ensure incentive compatibility constraints.

Identification of $u_0(\cdot)$ and $F(\cdot)$

I now turn to the identification of u_0 and $F(\cdot)$. Perrigne and Vuong (2011a) exploits the one-to-one mapping between the unobserved type and the observed consumption. Here, the first-order conditions (1) and (2) define a unique strictly increasing mapping from θ to Q. However, Q is not observed by the analyst. Instead I exploit the unique strictly increasing mapping between θ and t since the latter is observed. Specifically, I rewrite Equations (1) and (2) to express the marginal utility function $u_0(\cdot)$ and the unobserved type θ as functions of the tariff function $T(\cdot)$ and payment distribution $H(\cdot)$.

Lemma 2. $\forall Q \in [T^{-1}(\underline{t}), T^{-1}(\bar{t})]$, *the first-order conditions* (1) *and* (2) *can be rewritten as*

$$\theta(Q) = \theta_0 [1 - H(T(Q))]^{\frac{\gamma}{T'(Q)} - 1} \exp\left\{ \gamma \int_{T^{-1}(\underline{t})}^{Q} \frac{T''(x)}{T'(x)^2} \log\left[1 - H(T(x))\right] dx \right\}, \tag{6}$$

$$u_0'(Q) = T'(Q)/\theta(Q). \tag{7}$$

Notice that everything on the right-hand side of (6) and (7) are identified except θ_0. Lemma 1 suggests that a normalization of θ_0 leads to the identification of the consumers' marginal utility function and taste distribution. Moreover, the boundary condition $\underline{t} = \theta_0 u_0(Q(\theta_0)) = \theta_0 u_0(T^{-1}(\underline{t}))$ pins down the location of the base utility function. Thus the base utility function is identified on $[T^{-1}(\underline{t}), T^{-1}(\bar{t})]$.[6]

Assumption 4. $\theta_0 = 1$.

Under Assumption 4, $u_0(\cdot)$ can be interpreted as the utility function for the cutoff consumer using an average amount of additional features. By Proposition 2, for any $t \in [\underline{t}, \bar{t}]$,

[6]Perrigne and Vuong (2011a) identify the marginal utility function only. The reason is that the minimum consumption in their model (standard listing) is offered for free in yellow pages. Therefore, they could not relate the minimum consumption to a utility level.

Q is identified, and the corresponding taste can be obtained using (6). Thus the truncated type distribution $F^*(\cdot) \equiv [F(\cdot) - F(\theta_0)]/[1 - F(\theta_0)]$ can be uniquely recovered on $[\theta_0, \overline{\theta}]$.

The following proposition summarizes these results.

Assumption 5. *Under Assumptions 4, the utility function $u_0(\cdot)$ and the truncated taste distribution $F^*(\cdot)$ are identified on $[T^{-1}(\underline{t}), T^{-1}(\overline{t})]$ and $[\theta_0, \overline{\theta}]$, respectively.*

Since I do not observe the proportion of consumers who do not purchase from the firm, the distribution $F(\cdot)$ is identified up to a constant on $[\theta_0, \overline{\theta}]$. The data do not provide any variation to identify the utility function and the distribution function below on $[0, Q(\theta_0))$ and $[0, \theta_0)$, respectively.

1.4.2 Estimation

The data consist of $\{(q_{iL}, q_{iD}, q_{iR}, t_i)\}_{i=1}^N$. My semiparametric identification results in Section 1.4.1 lead naturally to a semiparametric procedure for estimation. I proceed in several steps. In a first step, I estimate α and $T(\cdot)$ using Equation (5). This allows me to calculate pseudo values of product unobserved heterogeneity $\{\widehat{\epsilon}_i\}_{i=1}^N$. Since each subscriber can customize mobile services matching his exact need, ϵ is potentially different across consumers. I then estimate γ and K using Lemma 1. In a second step, I use (6) and (7) to estimate the marginal utility function $u_0'(\cdot)$ and to construct a sample of pseudo tastes. Finally, I estimate the taste density and the density of unobserved heterogeneity by using a kernel estimator.

I provide below detailed information on every step.

Estimation of α, $T(\cdot)$, K and γ

As discussed above, (5) is a special case of the model of Ai and Chen (2003). They propose the sieve minimum distance estimator using sieves to approximate the unknown functions and estimating finite dimensional parameters and infinite dimensional unkown functionals jointly. Exploiting the separable structure of my model, I estimate α and $T^{-1}(\cdot)$ in two steps.

To estimate α, I partition the range of payments into $\mathcal{B}$ bins and define dummy variables $D_b(t) = 1$ if $t \in b$ and 0 otherwise. First, for any value b, multiplying both sides of (5) by

$D_b(t)$ and taking the variance conditional on t gives $\text{Var}\left[D_b(t)\alpha'\mathbf{q}|t\right] = D_b(t)\text{Var}\left[\log\epsilon|t\right]$. Second, taking the variance conditional on t on both sides of (5) gives $\text{Var}\left[\alpha'\mathbf{q}|t\right] = \text{Var}\left[\log\epsilon|t\right]$. Combining these two gives $\text{Var}\left[D_b(t)\alpha'\mathbf{q}|t\right] = D_b(t)\text{Var}\left[\alpha'\mathbf{q}|t\right]$. Taking expectation gives[7]

$$\alpha'\Big\{\mathrm{E}\left[D_b(t)\big(\mathbf{q}-\mathrm{E}\left(\mathbf{q}|t\right)\big)\big(\mathbf{q}-\mathrm{E}\left(\mathbf{q}|t\right)\big)'\right]-\mathrm{E}\left[D_b(t)\right]\mathrm{E}\left[(\mathbf{q}-\mathrm{E}\left(\mathbf{q}|t\right))(\mathbf{q}-\mathrm{E}\left(\mathbf{q}|t\right))'\right]\Big\}\alpha=0, \tag{8}$$

which says that the expected variance of $D_b(t)\alpha'\mathbf{q}$ equals $\mathrm{E}\left(D_b(t)\right)$ times the expected variance of $\alpha'\mathbf{q}$.

My estimator of α solves the following minimization problem:

$$\min_{\alpha} \quad \sum_{b=1}^{\mathcal{B}}\left\{\alpha'\left[\frac{\sum_{i:t_i\in b}(\mathbf{q}_i-\hat{\mathrm{E}}\left(\mathbf{q}_i|t_i\right))(\mathbf{q}_i-\hat{\mathrm{E}}\left(\mathbf{q}_i|t_i\right))'}{N_b}-\frac{\sum_{i=1}^{N}(\mathbf{q}_i-\hat{\mathrm{E}}\left(\mathbf{q}_i|t_i\right))(\mathbf{q}_i-\hat{\mathrm{E}}\left(\mathbf{q}_i|t_i\right))'}{N}\right]\alpha\right\}^2,$$

where $\mathbf{q}_i \equiv (\log q_{iL}, \log q_{iI}, \log q_{iO})'$, N_b is the number of observations in bin b and N is the total number of observations. $\mathrm{E}\left(\mathbf{q}_i|t_i\right)$ is estimated using a standard kernel estimator

$$\hat{\mathrm{E}}\left(\mathbf{q}_i|t_i\right)=\frac{\sum_{k=1}^{N}\mathbf{q}_k K(\frac{t_i-t_l}{h_t})}{\sum_{k=1}^{N}K(\frac{t_i-t_k}{h_t})},$$

where $K(\cdot)$ is a symmetric kernel function with compact support and h_t is some bandwidth.

To impose that $T^{-1}(\cdot)$ is increasing and convex, I use a constrained sieve estimator proposed by Dole (1999). The approximation splines are

$$\psi(\cdot;\beta,\delta)=\beta_0+\beta_1\cdot+\sum_{k=1}^{n}\delta_k s_k(\cdot),$$

where n is the number of interior knots, $\beta \equiv (\beta_0, \beta_1)'$ and $\delta \equiv (\delta_1, \ldots, \delta_n)'$. The range $[\underline{t}, \bar{t}]$ is partitioned into $n+1$ bins of the form $[\tau_{k-1}, \tau_k)$ for $k = 1, 2, \ldots, n+1$ with $\tau_0 = \underline{t}$ and

[7]See the Appendix for proof.

$\tau_{n+1} = \bar{t}$. The basis function $s_k(\cdot)$ is a cubic function defined as

$$s_k(t) = \begin{cases} 0 & \text{if } t \in [-\infty, \tau_{k-1}] \\ (t-\tau_{k-1})^3/[6(\tau_k - \tau_{k-1})] & \text{if } t \in [\tau_{k-1}, \tau_k] \\ ((t-\tau_{k+1})^3/[6(\tau_k - \tau_{k+1})]) + a_1 t + a_0 & \text{if } t \in [\tau_k, \tau_{k+1}] \\ a_1 t + a_0 & \text{if } t \in [\tau_{k+1}, +\infty] \end{cases}$$

where $a_1 = (\tau_{k+1} - \tau_{k-1})/2$ and $a_0 = ((\tau_k - \tau_{k-1})^2 - (\tau_k - \tau_{k+1})^2 + 3\tau_k(\tau_{k+1} - \tau_{k-1}))/6$.

My estimator of (β, δ) solves the following problem:

$$\min_{(\beta,\delta)} \quad \sum_{i=1}^{N} \left[\hat{\alpha}' \mathbf{q}_i - \log\left(\psi(t_i; \beta, \delta)\right)\right]^2,$$

where $\beta_1 \geq 0$, $\delta \geq 0$. I denote the estimate by $(\hat{\beta}, \hat{\delta})$.

Thus, $\hat{T}(\cdot) = \psi^{-1}(\cdot; \hat{\beta}, \hat{\delta})$ and $\hat{Q}(\cdot) = \hat{T}^{-1}(\cdot) = \psi(\cdot; \hat{\beta}, \hat{\delta})$. I estimate γ by $\hat{\gamma} = \hat{T}'(\hat{T}^{-1}(t_{\max}))$, and K by $\hat{K} = \hat{\gamma}\left[\frac{t_{\min}}{\hat{T}'(\hat{T}^{-1}(t_{\min}))} - \hat{T}^{-1}(t_{\min})\right]$, where $t_{\min} = \min_{i=1,2,\ldots,N} t_i$ and $t_{\max} = \max_{i=1,2,\ldots,N} t_i$.

Following the same lines as in Lavergne and Vuong (1996), I can show that the estimates for residual variance matrices are $\sqrt{N}$-consistent and asymptotically normally distributed. Thus, my estimate for α is also $\sqrt{N}$-consistent and asymptotically normally distributed. The estimation of $\log T^{-1}(\cdot)$ and $G(\cdot)$ becomes reminiscent of Horowitz (1996). Following the identification procedure, I can obtain "plug-in" estimators for $\log T^{-1}(\cdot)$ and $G(\cdot)$, which are $\sqrt{N}$-consistent and asymptotically normally distributed.

For simplicity, I use a constrained sieve estimator to impose that $T^{-1}(\cdot)$ is increasing and convex. Ai and Chen (2003) show that the sieve minimum distance estimator of the parametric component is $\sqrt{N}$-consistent and asymptotically normally distributed, and the estimator of infinite dimensional functions is consistent with a rate faster than $N^{-1/4}$ under certain metric. Meyer (2008) shows that the shape-restricted regression has smaller squared error loss than the unrestricted version, when the true regression function satisfies the shape assumption. Thus the rate is not slower than the unrestricted version.

Estimation of $u_0(\cdot)$, $F(\cdot)$ and $G(\cdot)$

I estimate $H(\cdot)$ as the empirical distribution of payment,

$$\hat{H}(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(t_i \leq t),$$

where $\mathbb{1}(\cdot)$ is an indicator function and $t \in [t_{\min}, t_{\max}]$.

For any $Q \in [\hat{T}^{-1}(t_{\min}), \hat{T}^{-1}(t_{\max})]$, the estimate for $u_0'(Q)$ is given by $\hat{u}_0'(Q) = \hat{T}'(Q)/\hat{\xi}(Q)$, and the estimate for $\theta(Q)$ by $\hat{\theta}(Q) = \hat{\xi}(Q)$, where

$$\hat{\xi}(Q) = [1 - \hat{H}(\hat{T}(Q))]^{\frac{\hat{\gamma}}{\hat{T}'(q)}-1} \exp\left\{\hat{\gamma}\int_{T^{-1}(\underline{t})}^{Q} \frac{\hat{T}''(x)}{\hat{T}'(x)^2}\log[1-\hat{H}(\hat{T}(x))]dx\right\}.$$

Thus, for any $Q \in [\hat{T}^{-1}(t_{\min}), \hat{T}^{-1}(t_{\max})]$, I can estimate the taste. A pseudo sample of taste can be constructed as $\left\{\hat{\theta}_i\right\}_{i=1}^{N}$, where $\hat{\theta}_i = \hat{\xi}(\hat{T}^{-1}(t_i))$. Moreover, I can estimate ϵ by $\hat{\epsilon}_i = \hat{T}^{-1}(t_i)/\exp(\hat{\alpha}'\mathbf{q}_i)$. Hence, a pseudo sample of unobserved heterogeneity can be constructed $\{\hat{\epsilon}_i\}_{i=1}^{N}$. Finally, using these two pseudo samples, I estimate the truncated density of taste and the density of unobserved heterogeneity by

$$\hat{f}^*(\theta) = \frac{1}{Nh_\theta}\sum_{i=1}^{N} K(\frac{\theta - \hat{\theta}_i}{h_\theta}),$$
$$\hat{g}(\epsilon) = \frac{1}{Nh_\epsilon}\sum_{i=1}^{N} K(\frac{\epsilon - \hat{\epsilon}_i}{h_\epsilon}),$$

for $(\theta, \epsilon) \in [1, \overline{\theta}] \times [\underline{\epsilon}, \overline{\epsilon}]$, where $K(\cdot)$ is a symmetric kernel function with compact support, h_θ and h_ϵ are some bandwidths.

1.5 Empirical Analysis

In this section, I present the estimated model primitives and examine several counterfactual experiments.

Estimation Results

First, I estimate the aggregation parameters $(\alpha_L, \alpha_D, \alpha_R)$ and the tariff function $T(\cdot)$. The estimated aggregation parameters are $\alpha_L = 0.4471$, $\alpha_D = 0.3053$ and $\alpha_R = 0.2476$, which are quite close to the estimates in Luo, Perrigne, and Vuong (2012). Figure A.1 displays the concave tariff $\hat{T}(\cdot)$. It implies an estimate of the fixed cost K is \$1.78. This is approximately 6.54% of the average bill. The estimate of the marginal cost parameter γ is \$0.1784. It is approximately a half of the average price charged per minute \$0.3389. These cost parameters imply that the firm has a profit margin of 37.48%.

Second, I obtain estimates of the one-to-one mapping $\theta(\cdot)$ and the base utility function $u_0(\cdot)$. The first is displayed in Figure A.2 while the latter in Figure A.3. The estimated mapping is increasing, thereby satisfying the prediction of my model. The estimated base utility function is increasing and concave, thereby satisfying Assumption 1-(i).

Finally, I estimate the truncated taste density function $f^*(\cdot)$ and the density function of unobserved heterogeneity $g(\cdot)$. The first is displayed in Figure A.4 while the latter in Figure A.5. Both densities are unimodal and displays an important skewness. The skewness of the taste density function will have important implications on the analysis of counterfactual experiments as discussed later. As displayed in Table A.1, unobserved heterogeneity shows an important variability with a standard deviation of 1.32. On the other hand, the correlation between the pseudo taste and unobserved heterogeneity is 0.0027. I use a nonparametric test of bivariate independence due to Blum, Kiefer, and Rosenblatt (1961) and obtain a p-value of 0.0009. Thus, the independence assumption seems to be satisfied.

With the estimated model primitives, I calculate the informational rent left to the consumer. As predicted by theory, this rent is increasing in taste θ and payment t. As shown in Table A.1, the ratio of the informational rent by payment is on average 29.39% while the ratio of the total informational rent by the total payment is 38.03%. These measures the cost of asymmetric information for the firm.

Counterfactual Experiments

I now describe several counterfactual experiments based on the estimated model primitives. In particular, I consider three alternative forms of pricing schemes: incremental discounts (ID), all-units discounts (AUD), and quantity forcing (QF).

An incremental discounts (ID) scheme is a menu of two-part tariffs, which are of the form $t = \kappa_j + p_j q$, where κ_j and p_j are the fixed fee and the marginal price. Thus, marginal prices of successive units decline in steps. An all-units discounts (AUD) scheme is a menu of minimum purchase tariffs, which are of the form $t = p_j q$ if $q \geq m_j$, $= \infty$ otherwise, where m_j and p_j are the minimum purchase and marginal price. Thus, the per-unit price progressively drops when the order size exceeds certain threshold. Under QF, only several quantities m_j, each associated with a gross price t_j, are offered for consumers to choose. To make a fair comparison, I allow the same level of complexity, namely, each of them has two options and can be characterized by four parameters. For completeness, I also simulate linear pricing scheme of the form $t = pq$, where p is the price.

Table A.3 shows the predicted revenue, consumer surplus, firm profit and purchase for different schemes. According to Wong (2012), ID has the largest "approximation power" among these three schemes. This is confirmed in Figure A.6, where the tariff function for ID is closer to the second-best than those for AUD and QF. As shown in Table A.3, the optimal firm profit under ID is \$10.09. Relative to nonlinear pricing, it drops by 3.17% while consumer surplus drops by 9.07%. Apparently, consumers are losing more than the firm. In addition to these effects there is heterogeneity in terms of who wins and who loses. Figure A.7 shows the predicted consumer surplus for different types of consumers under ID. Again, ID is closer to the second-best. An interesting result is that high-end consumers benefit, albeit by a small amount. More low-end consumers are excluded from purchasing because the new cutoff taste is 1.16. The reason is that ID requires a higher minimum payment of \$11.63, leading to a higher minimum purchase of 82.33 minutes.

Relative to ID, AUD seems to approximate the second-best better at the lower end. This is shown in Figures A.6 and A.7. Since the tariff becomes much more expensive for high-end consumers, their surplus drops a lot. On average, consumer surplus drops by 12.99% while firm profit drops by 8.97% relative to NLP. As AUD tends to focus on the lower end of the market, the expected purchase drops by 23.62%.

While ID and AUD leave some flexibility to the consumer, QF is most restrictive as it allows only certain quantities of purchase. As a result, its effects are close to the linear pricing case in terms of expected revenue, consumer surplus, firm profit and purchase.

At the optimal quantities $q_1 = 51.77$ and $q_2 = 66.65$, consumers pay \$20.31 and \$25.17, respectively. These are approximately two dollars higher than the corresponding prices under NLP, namely \$18.74 and \$23.06. As a result, Figure A.7 shows that every consumer is worse off relative to NLP or AUD.

From these simulations, it appears that the firm does not lose much relative to nonlinear pricing by using simpler price schemes. In the "worse" case, quantity forcing only reduces its profit by 16.12%. This is consistent with the results reported in Miravete (2007) using data from the early U.S. cellular telephone industry. The reason is that the estimated taste density is quite skewed. In this case, the firm can still reach most of the market even with limited pricing flexibility. Moreover, with the same level of complexity, incremental discounts seem to be more preferable as it does a better job approximating the second-best, leading to higher firm profit and consumer surplus relatives to AUD and QF.

1.6 Conclusion

This paper proposes a method to incorporate product customization and unobserved heterogeneity in structural analysis of nonlinear pricing data. The first step involves identifying and estimating the tariff from data on payments and consumption quantities. I aggregate observed multiple quantities and unobserved heterogeneity of the product, rendering recovering the tariff a problem reminiscent of transformation models and single index models. I then identify the model primitives by exploiting the one-to-one mapping between the consumer's taste and his payment. I propose a computationally convenient semiparametric estimation procedure. An application using data from a major mobile service provider in Asia illustrates my method. I conduct various experiments to investigate the implications of alternative simpler pricing policies.

While I consider a single product with multiple attributes (different kinds of phone minutes), my method can extend to multiproduct problems. For example, it is interesting to consider both voice and short message service (SMS) consumption. Relying on Armstrong (1996), Luo, Perrigne, and Vuong (2012) generalize the method developed in this paper to the multiproduct case and apply them to the empirical analysis of voice and SMS in the mobile phone industry. Luo (2012) studies bundling and nonlinear pricing by a multiproduct firm selling internet and phone services. It also uses the method developed in this paper to estimate tariffs as a first step.

Chapter 2

Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS

2.1 Introduction

This paper studies multiproduct nonlinear pricing in the cellular phone industry with voice and message services. The model derives from Armstrong (1996) in which the unknown types of consumers are aggregated while the firm designs an optimal cost-based tariff. As usual in multidimensional screening problems, there is pooling at equilibrium. Moreover, given that the consumers add a large number of unobserved extra features, we introduce two terms of unobserved heterogeneity for voice and message add-ons. The model defines two one-to-one mappings between the unknown aggregate type to the cost and the cost to the payment. We then study the identification of the model primitives. Under some identifying assumptions such as a parameterization of the cost function, we show that the primitives of the model (the aggregate type density, the indirect utility and the joint density of unobserved heterogeneity) are identified from observables. The empirical results support the model and display an important heterogeneity in types and unobserved heterogeneity. The cost of asymmetric information for the firm is assessed.

The paper is organized as follows. Section 2 presents the data. Section 3 introduces the model, while Section 4 establishes its nonparametric identification and develops a nonparametric estimation procedure. Section 5 is devoted to our estimation results and counterfactuals. Section 6 concludes with some future lines of research. An appendix collects the proofs.

2.2 Cellular Phone Service Data

We collected data from a major telecommunication company in a metropolitan area in Asia.[1] For the period of May 2009, the data contain voice service and short message service (SMS) consumptions as well as the amount paid by each subscriber. The company which gave us the data has a national market share of 73%. This company proposes three different tariffs to three different types of consumers: Students, rural residents and urban residents. Our data concern urban residents only. The cellular provider's market share in this market segment is larger than 80%. The cellular phone provider proposes every year a new tariff. Subscribers can switch to the new tariff at no extra cost. To avoid mixing of subscribers under different tariffs, we consider subscribers who are under the 2009 tariff in May 2009. This gives a sample of 4,844 observations. Eliminating coding errors and cleaning the data reduces the sample to 4,601 observations.

The bill paid by a consumer combines several types of phone calls, roaming charges, extra fees for peak hours usage, SMS and several add-ons such as voice mail service, music on hold, ring tones, multimedia message service (MMS), news, etc. Unlike the tarification by U.S. companies, which propose a few packages for voice and SMS with add-ons and minutes, a consumer pays here for all the services he consumes with the exception of incoming calls and SMS which are free of charge as long as the subscriber is located in the city. In particular, the company distinguishes outgoing and incoming calls according to the location of the subscriber when giving phone calls. Our data provide information on voice consumption measured in minutes and distinguishes three different types of phone calls. Local calls are when both parties are in the same city. Distance calls are when the two parties are in different locations, while roaming calls are when the subscriber is outside his home city. The

[1] Following an agreement with the company, we cannot reveal the name of the city or the country. For this reason, we cannot name the currency either and all values have been converted in U.S. dollars.

latter is introduced because of large fees associated with roaming. [2] We also observe the number of SMS sent. The data do not contain information on add-ons and other features. On the other hand, we observe the total bill paid by every subscriber which includes all these extra features. These extra features constitute an important source of variation across consumers.

Table 1 provides summary statistics on the bill measured in U.S. dollars, the consumptions of the three types of calls measured in minutes and the number of SMS. All consumptions are quite variable and very skewed, especially roaming calls. We aggregate the three quantities of phone calls into a single index denoted q_v. The aggregation method follows Luo (2011) and is explained in Appendix B. Regarding SMS, we keep the number of SMS and denote it q_m. Table 1 also includes summary statistics on q_v. Table 2 provides the average consumptions of q_v and q_m, the average bill as well as the correlation between these three variables for the full sample and when controlling for seven categories of the bill value. When controlling for the bill value, the correlation between the bill and q_v is quite small, though larger than the one between the bill and q_m. This confirms the importance of the unobserved extra features. In other words, the voice and SMS consumptions explain only part of the consumers' bills variability. These two correlations tend to decrease with the bill amount. The correlation between q_v and q_m is negative when controlling for the bill value suggesting that consumers tend to subsitute voice for SMS. It varies, however, across bill values indicating a complex pattern of substituability across consumption levels. On the other hand, these correlations are quite different when considering the full sample. Specifically, the two correlations with bills become larger, while the correlation between voice an message services becomes slightly positive. The latter arises because both consumptions tend to increase with the bill. This illustrates the importance of controlling for the bill value.

These data patterns are confirmed in Figures 1–3. Figures 1 and 2 display the scatter plots of the voice consumptions q_v and bills and message consumptions q_m and bills, respectively. These scatter plots show again an important variability, i.e. voice and message

[2] International calls are very rare and eliminated from the sample.

consumptions explain only a small proportion of the bill variability. Figure 3 also displays the pairs (q_v, q_m) showing no pattern between these two. Lastly, we show evidence of the tariff curvature. We regress the bill q_v, q_m and their squares. The R^2 of such a regression is 0.47. Figure 4 displays the fitted tariff in a three dimensional space. We observe a tariff increasing in both quantities and more importantly a concave tariff, i.e. subscribers consuming larger quantities of voice and message services enjoy a discount, i.e. they pay a smaller price per unit when they consume more.

In summary, because this company has a large market share, we can reasonably assume that it is a monopoly. We also observe an important variability of bills paid by subscribers, which is partly explained by the variability of their consumption of voice and message services. This fact is due to the various add-ons and extra features bought by the consumers. Subscribers pay for all the services they consume, which is in the spirit of product customization. Moreover, the amount they pay appears to be concave in the quantities of voice and message services.

2.3 The Model

We rely on Armstrong (1996) model of multiproduct nonlinear pricing. In particular, we consider two products: Voice and message services. A subscriber is characterized by a pair of types $(\theta_v, \theta_m) \in \Theta \equiv [\underline{\theta}_v, \overline{\theta}_v] \times [\underline{\theta}_m, \overline{\theta}_m] \subset I\!R^2_+$. This pair of types or tastes is consumer's private information. Moreover, because of the multiple features of voice and message services, which are not observed but are included in the subscribers' bills, we follow Luo (2011) model of product customization by introducing another pair of random variables $(\epsilon_v, \epsilon_m) \in \mathcal{E} \equiv [\underline{\epsilon}_v, \overline{\epsilon}_v] \times [\underline{\epsilon}_m, \overline{\epsilon}_m] \subset I\!R^2_+$ to capture the subscriber's additional unobserved heterogeneity. The latter pair takes into account location and time of the subscriber's phone calls as well as all the extra features associated with his phone and message services. The pair (ϵ_v, ϵ_m) is common knowledge to the subscriber and the firm but unobserved by the analyst. The vector $(\theta_v, \theta_m, \epsilon_v, \epsilon_m)$ is distributed as $F(\cdot, \cdot, \cdot, \cdot)$ with a continuous density $f(\cdot, \cdot, \cdot, \cdot) > 0$ on $\Theta \times \mathcal{E}$.

A subscriber consuming quantities of voice and message (q_v, q_m) has a payoff

$$U(q_v, q_m; \theta_v, \theta_m, \epsilon_v, \epsilon_m) - T(q_v, q_m; \epsilon_v, \epsilon_m),$$

where $T(q_v, q_m; \epsilon_v, \epsilon_m)$ is the total payment for consuming (q_v, q_m) when the other features of consumption are (ϵ_v, ϵ_m). We make the following standard assumptions on $U(\cdot, \cdot; \cdot, \cdot, \cdot, \cdot)$.

Assumption A1: *The function* $U(\cdot, \cdot; \cdot, \cdot, \cdot, \cdot)$ *satisfies*

(i) $U(0, 0; \theta_v, \theta_m, \epsilon_m, \epsilon_m) = 0 \ \forall (\theta_v, \theta_m, \epsilon_v, \epsilon_m) \in \Theta \times \mathcal{E}$,

(ii) $U(\cdot, \cdot; \cdot, \cdot, \cdot)$ *is strictly increasing in all its arguments,*

(iii) $U(\cdot, \cdot; \cdot, \cdot, \cdot)$ *is continuous, convex and homogenous of degree one in* (θ_v, θ_m).

These assumptions follow Armstrong (1996). The homogeneity of degree one may seem restrictive but this assumption is the simplest one that allows for nontrivial multidimensional types.[3]

The firm incurs a cost $C(q_v, q_m; \epsilon_v, \epsilon_m)$ in serving the consumer with the bundle (q_v, q_m) with additional features (ϵ_v, ϵ_m). We assume that the cost function is continuously differentiable. We consider the case when the firm's total cost is additively separable across subscribers. We allow, however, for scope economies in serving a subscriber with two products. We remark that the separability of the cost function across subscribers is a standard assumption in the nonlinear pricing literature.[4]

Hereafter to simplify the notations, we drop the index v and m and let $\theta = (\theta_v, \theta_m)$, $\epsilon = (\epsilon_v, \epsilon_m)$ and $q = (q_v, q_m)$. Since the firm observes the additional features ϵ, it chooses the quantity schedules $q_v(\cdot; \cdot)$ and $q_m(\cdot; \cdot)$ as well as the tariff $T(\cdot; \cdot)$ to maximize its profit conditional on ϵ. This gives

$$\max_{q(\cdot;\epsilon), T(\cdot;\epsilon)} \int_{\Theta} \left[T(q(\theta; \epsilon); \epsilon) - C(q(\theta; \epsilon); \epsilon) \right] f(\theta|\epsilon) d\theta, \tag{1}$$

where $f(\theta|\epsilon)$ is the conditional density of θ given ϵ. As usual in problems with adverse

[3] The function $U(q_v, q_m; \theta_v, \theta_m, \epsilon_v, \epsilon_m) = \theta_v U_v(q_v; \epsilon_v, \epsilon_m) + \theta_m U_m(q_m; \epsilon_v, \epsilon_m)$ satisfies A1 but do not allow cross-price effects in demand. See also Adams and Yellen (1976). Assumption A1 is more general.

[4] Perrigne and Vuong (2011a) consider instead the cost for the total production across subscribers. Luo (2011) shows that the two models are nonnested except if one assumes a constant marginal cost.

selection, we eliminate the tariff $T(\cdot;\cdot)$ by considering the consumer surplus $S(\theta;\epsilon) = \max_{q\geq 0} U(q;\theta,\epsilon) - T(q;\epsilon)$. Following A1 and $T(0;\epsilon) \leq 0$ for all $\epsilon \in \mathcal{E}$ because consumers cannot be forced to consume, $S(\cdot;\epsilon)$ is nonnegative, continuous, increasing and convex in θ. See Armstrong (1996). We are looking at demand functions $q(\cdot;\epsilon)$ that are implementable, i.e. that satisfies the maximization problem of the consumer facing some tariff $T(\cdot;\epsilon)$. In the single product case where the single crossing property holds, a necessary and sufficient condition for implementability is the monotonicity of $q(\cdot;\epsilon)$. Because we consider multidimensional screening, the single crossing property does not hold and we will ensure that the demand $q(\cdot;\epsilon)$ verifies implementability by constructing a corresponding tariff.

To solve the problem of multidimensional screening, we follow the reduction technique. In particular, the problem is rewritten in terms of the cost incurred to serve the customer.[5] Let $V(c;\theta,\epsilon)$ be the cost-based indirect utility function

$$V(c;\theta,\epsilon) \equiv \max_{q:C(q;\epsilon)\leq c} U(q;\theta,\epsilon) \tag{2}$$

associated with the utility $U(q;\theta,\epsilon)$. To insure implementability, we make separability assumptions on the indirect utility function and the joint density $f(\theta,\epsilon)$.

Assumption A2:

(i) The indirect utility function is multiplicatively separable in θ as $V(c;\theta,\epsilon) = h(\theta)V_0(c;\epsilon)$ with $h(\cdot) \geq 0$,

(ii) The joint density $f(\theta,\epsilon)$ is multiplicatively separable as $f(\theta,\epsilon) = f_h(h(\theta))f_0(\theta,\epsilon)$, where $f_0(\cdot,\cdot)$ is homogenous of degree zero in θ.

We remark that $h(\cdot)$ is increasing and homogenous of degree one following A1-(ii,iii).[6] Intuitively, h can be interpreted as an aggregation of the tastes (θ_v,θ_m) as (say) an average taste.

[5] The reduction technique is often applied with particular parametric specifications of the primitives as in Ivaldi and Martimort (1994). See also Rochet and Stole (2003). The idea of cost-based tariff developed by Armstrong (1996) avoids these restrictive specifications. It requires, however, some assumptions on the primitives as discussed in the text. See also Aryal and Perrigne (2011) in the context of insurance where certainty equivalence is used to reduce the two dimensions of adverse selection (risk and risk aversion) into a single one.

[6] We note that A2-(i) is more general than assuming $U(q;\theta,\epsilon) = h(\theta)U_0(q;\epsilon)$ as the latter implies A2-(i). On the other hand, the reverse is not true.

We denote by $\phi(\cdot|\epsilon)$ and $\Phi(\cdot|\epsilon)$ the density and distribution of the random variable $h = h(\theta)$ conditional on ϵ induced by the underlying distribution $F(\theta, \epsilon)$. The conditional density $\phi(\cdot|\epsilon)$ satisfies

$$\phi(h|\epsilon) = k_\epsilon h f_h(h), \tag{3}$$

where $k_\epsilon = 1/\int_{\mathcal{S}_{h|\epsilon}} h f_h(h) dh$ from equation (24) in Armstrong (1996), where $\mathcal{S}_{h|\epsilon}$ is the support of h given ϵ. Given that $f(\cdot, \cdot, \cdot, \cdot) > 0$ everywhere on its support $\Theta \times \mathcal{E}$ and $h(\cdot)$ is increasing, $\mathcal{S}_{h|\epsilon}$ is independent of ϵ thereby implying that k_ϵ is independent of ϵ and hence $h \equiv h(\theta)$ is independent of ϵ. Such an independence is resulting from A2-(ii).[7] On the other hand, we still allow for some dependence between (θ_v, θ_m) and (ϵ_v, ϵ_m).

We need to make an assumption involving the hazard rate of the density $\phi(\cdot)$ of h.

Assumption A3: *The density $\phi(\cdot)$ is such that $1 - \frac{1-\Phi(h)}{h\phi(h)}$ is nondecreasing in $h \in \mathcal{H} \equiv [\underline{h}, \overline{h}] = [h(\underline{\theta}_v, \underline{\theta}_m), [h(\overline{\theta}_v, \overline{\theta}_m)]$.*

A sufficient condition to obtain both a nondecreasing demand and a concave tariff is given by A3. We remark that A3 is stronger than the standard hazard rate condition, i.e. $h - [(1 - \Phi(h))/\phi(h)]$ is nondecreasing in h.

We are now in position to solve the firm's problem leading to a cost-based tariff $T(C(q; \epsilon); \epsilon)$ for any given value of ϵ.

Proposition 1: *Under Assumptions A1, A2 and A3, the optimal tariff is cost-based. In particular, there exists a function $T(\cdot; \epsilon)$ such that the payment of a (θ, ϵ) subscriber is $t = T(C(q; \epsilon); \epsilon)$. Moreover, the (θ, ϵ) subscriber chooses a quantity q solving*

$$\max_{\{q: c(q;\epsilon) \leq c\}} U(q; \theta, \epsilon), \tag{4}$$

where c is solution of

$$\max_{c \geq 0} \left\{ \left[h - \frac{1 - \Phi(h)}{\phi(h)} \right] V_0(c; \epsilon) - c \right\} \tag{5}$$

[7] Alternative assumptions such as $f(\theta, \epsilon) = f_h(h(\theta, \epsilon), \epsilon) f_0(\theta, \epsilon)$ could relax the independence of h and ϵ. Identification would become even more difficult. See also footnote 10.

with $h = h(\theta)$. Moreover, the resulting cost-based tariff $T(c;\epsilon)$ is increasing and concave in c.

The proof of Proposition 1 follows Armstrong (1996) (Propositions 2 and 3). Given ϵ, (5) implicitly defines a one-to-one increasing mapping between c and h as $c = C(h;\epsilon)$. The firm would serve all subscribers if

$$\left[\underline{h} - \frac{1-\Phi(\underline{h})}{\phi(\underline{h})}\right] V_0(C(\underline{h};\epsilon);\epsilon) - C(\underline{h};\epsilon) \geq 0.$$

In general, there is optimal exclusion of some consumers, i.e. there exists a value $h_0 \in (\underline{h}, \overline{h})$ defined as

$$\left[h_0 - \frac{1-\Phi(h_0)}{\phi(h_0)}\right] V_0(C(h_0;\epsilon);\epsilon) - C(h_0;\epsilon) = 0, \tag{6}$$

below which consumers will not buy cellular phone service. The term $[h_0 V_0(C(h_0;\epsilon);\epsilon) - C(h_0;\epsilon)]\phi(h_0)$ is the marginal gain for expanding the customer base by lowering h_0, while $[1-\Phi(h_0)]V_0(C(h_0;\epsilon);\epsilon)$ is the corresponding marginal loss for reducing the tariff to every customer above h_0. Equation (6) balances these two effects. In addition, there is a boundary condition given by

$$T(C(h_0;\epsilon);\epsilon) = h_0 V_0(C(h_0;\epsilon);\epsilon) \tag{7}$$

because the informational rent left to the threshold customer is zero. The optimal exclusion is proved in the appendix.

The next corollary provides the solution of the firm's optimization problem, i.e. the one-to-one mapping between the unobserved subscriber's aggregate taste h and the firm's cost c as well as the one-to-one mapping bewtween this cost and the tariff or payment paid by the subscriber.

Corollary 1: *Under Assumptions A1, A2 and A3, for any given value of ϵ, the tariff $T(\cdot;\epsilon)$*

satisfies

$$T'(c;\epsilon) \quad = \quad hV_0'(c;\epsilon) \tag{8}$$

subject to the boundary condition (7), where $c = C(h;\epsilon)$ *solves*

$$hV_0'(c;\epsilon) \quad = \quad 1 + \frac{1-\Phi(h)}{\phi(h)} V_0'(c;\epsilon). \tag{9}$$

Equation (9) is the first-order condition for (5). We turn to (8). The subscriber's optimization problem is to maximize $U(q;\theta,\epsilon) - T(q;\epsilon)$ with respect to q. However, as soon as the optimal tariff is cost-based, this problem is solved in two steps. In a first step, he maximizes his payoff subject to $C(q;\epsilon) \leq c$. This is equivalent to (4). In a second step, he maximizes $V(c;\theta,\epsilon) - T(c;\epsilon)$ with respect to c. Equation (8) is the first-order condition of this second step problem. The firm designs a payment $T(\cdot;\epsilon)$ such that the subscriber's optimal c will be given by (9). In particular, (8) together with the boundary condition (7), where h_0 satisfies (6), characterize the optimal tariff $T(\cdot;\epsilon)$. Equation (8) says that the marginal payment equals the marginal indirect utility, while (9) says that the marginal payoff equals the marginal cost adjusted by a positive distortion due to incomplete information. The optimal quantity $q(\cdot;\epsilon)$ is then obtained from (4).

2.4 Identification and Estimation

In view of Section 3, the model primitives are $[U(\cdot,\cdot;\cdot,\cdot,\cdot), f(\cdot,\cdot,\cdot,\cdot), h(\cdot,\cdot), C(\cdot,\cdot;\cdot,\cdot)]$, namely the subscribers' utility function, the type density with $f(\cdot,\cdot,\cdot,\cdot) = f_h(\cdot) f_0(\cdot,\cdot,\cdot,\cdot)$, the type aggregation function and the firm's cost. In view of Section 2, the observables are the subscribers' consumed quantities of various phone calls and SMS as well as their monthly bills, which include a large set of add-ons and other features. The observables are denoted by q_v, q_m and t. The vector (q_v, q_m, t) is distributed as $G(\cdot,\cdot,\cdot)$.

We study first the identification of the model primitives. Once identification is established, we develop a nonparametric estimation method for the model primitives.

2.4.1 Nonparametric Identification

We proceed in several steps. Given the observables, it is obvious that the model is not identified without further assumptions on the model primitives. First, we provide some identifying restrictions and we derive the resulting equations that will be used in identification. Second, we study the identification of the indirect utility and cost functions as well as the density of the aggregate type. Third, we investigate the identification of the utility function and the joint density of types.

Identifying Assumptions

The terms ϵ_v and ϵ_s capture the subscriber's tastes for extra features and add-ons to voice and SMS consumptions such as roaming charges and MMS. Consumers with larger values of ϵ_v and ϵ_m enjoy a larger utility from their consumptions of voice and message services than a subscriber with lower values. Following Luo (2011), we assume that these terms of unobserved heterogeneity can be interpreted as quantity multipliers.

Assumption A4: *The utility and cost functions satisfy*

(i) $U(q_v, q_m; \theta_v, \theta_m, \epsilon_v, \epsilon_m) = h(\theta_v, \theta_m)U_0(q_v\epsilon_v, q_m\epsilon_m)$,

(ii) $C(q_v, q_m; \epsilon_v, \epsilon_m) = \kappa(q_v\epsilon_v)^\gamma(q_m\epsilon_m)^{1-\gamma}$ *with* $\gamma \in (0, 1)$.

The multiplicative separability of the utility function satisfies A2-(i). A more general specification for the multiproduct cost function is $\kappa_0 + \kappa(q_v\epsilon_v)^{\gamma_v}(q_m\epsilon_m)^{\gamma_m}$. The term κ_0 captures a fixed cost associated to any subscriber such as billing costs and incoming calls. Moreover, the cellular phone technology requires phones to be connected to a tower even if no phone call is made. The variable cost $\kappa(q_v\epsilon_v)^{\gamma_v}(q_m\epsilon_m)^{\gamma_m}$ arises from the delivery and recording of voice and message services. For instance, roaming as captured by ϵ_v is costly to the firm because other cellular sites and mobile-services switching centers are involved thereby requiring financial settlements with other cellular phone providers. Similarly, delivering an MMS requires more data flow than a simple SMS, which is costly to the firm and captured by ϵ_m.

Assumption A4-(ii) imposes $\kappa_0 = 0$ and $\gamma_v + \gamma_m = 1$. These restrictions arise from identification. Specifically, as in any firm's optimization problem, the optimal solution results from the profit at the margin, which involves the marginal variable cost only. Thus this solution does not depend on the fixed cost κ_0, which is therefore not identified. On the other hand, because the firm is operating, the tariff covers the total cost thereby providing an upper bound for the fixed cost. When there is a single product with a single parameter of adverse selection and a known tariff $T(\cdot)$, Perrigne and Vuong (2011a) (Lemma 3) show that only the marginal cost evaluated at the largest quantity is identified at most. More generally, a one parameter specification of the variable cost function is identified in that case. Here, we have two products with two parameters of adverse selection while $T(\cdot)$ is unknown. On the other hand, optimal exclusion provides an additional equation. Thus we may expect to identify more than one parameter of the cost function. We then make the assumption $\gamma_v + \gamma_m = 1$ to identify κ. This corresponds to a situation where there is neither scale economies nor scale diseconomies, which seems reasonable for the cellular phone technology.

Under A4, a change of variables leads to the following indirect utility function given (ϵ_v, ϵ_m)

$$V_0(c; \epsilon_v, \epsilon_m) = \max_{(q_v\epsilon_v)^\gamma (q_m\epsilon_m)^{1-\gamma} \leq c} U_0(q_v\epsilon_v, q_m\epsilon_m) = \max_{\kappa Q_v^\gamma Q_m^{1-\gamma} \leq c} U_0(Q_v, Q_m),$$

where $Q_v \equiv q_v\epsilon_v$ and $Q_m \equiv q_m\epsilon_m$. Thus $V_0(\cdot; \cdot, \cdot)$ no longer depends on (ϵ_v, ϵ_m). We then drop these arguments as $V_0(c; \epsilon_v, \epsilon_m) = V_0(c)$. Similarly, $C(q_v, q_m; \epsilon_v, \epsilon_m) = C(Q_v, Q_m)$. Thus A4 has greatly simplified the functions as the terms (ϵ_v, ϵ_m) have disappeared.[8] In particular, using A4, (5) becomes

$$\max_{c \geq 0} \left\{ \left[h - \frac{1 - \Phi(h)}{\phi(h)} \right] V_0(c) - c \right\},$$

whose solution provides a one-to-one mapping between the optimal cost c and the aggregate type $h = h(\theta_v, \theta_m)$. We denote this mapping $C(\cdot)$ so that $c = C(h)$. Moreover, making a

[8] If (ϵ_v, ϵ_m) were observed, then we could drop A2-(ii) and A4 and conduct our analysis conditional on (ϵ_v, ϵ_m). See also footnote 9.

change of variables as above and using A4, (4) becomes

$$\max_{\kappa Q_v^{\gamma} Q_m^{1-\gamma} \leq C(h)} U_0(Q_v, Q_s). \tag{10}$$

The solution of (10) is the pair $(Q_v(h), Q_m(h))$, wich provides a one-to-one mapping between each effective quantity Q_v and Q_m with the aggregate type h. Thus, the voice and message quantities consumed by a $(\theta_v, \theta_m, \epsilon_v, \epsilon_m)$ consumer are

$$q_v(\theta_v, \theta_m; \epsilon_v, \epsilon_s) = \frac{Q_v(h(\theta_v, \theta_m))}{\epsilon_v}, \quad q_m(\theta_v, \theta_m; \epsilon_v, \epsilon_s) = \frac{Q_m(h(\theta_v, \theta_m))}{\epsilon_m}.$$

Proposition 1 states that the optimal multiproduct tariff for any given value of (ϵ_v, ϵ_m) is a cost-based tariff. Under A2 and A4, the functions $V_0(\cdot), Q_v(\cdot), Q_m(\cdot)$ do not depend on (ϵ_v, ϵ_m). Thus, the tariff does not depend on (ϵ_v, ϵ_m) either. In other words, there is a single cost-based tariff $T(\cdot)$ such as a $(\theta_v, \theta_m, \epsilon_v, \epsilon_m)$ subscriber pays $t = T[C[Q_v(h(\theta_v, \theta_m)), Q_m(h(\theta_v, \theta_m))]]$ with $T(\cdot)$ strictly increasing and concave. Moreover, from Corollary 1, the tariff $T(\cdot)$ and the cost c satisfy

$$\begin{aligned} T'(c) &= hV_0'(c), && (11) \\ hV_0'(c) &= 1 + \frac{1-\Phi(h)}{\phi(h)} V_0'(c) && (12) \end{aligned}$$

with boundary condition $T(C(h_0)) = h_0 V_0(C(h_0))$ from (7), where $h \in [h_0, \overline{h}] \subset [\underline{h}, \overline{h}]$ and $c \in [\underline{c}, \overline{c}]$, where $\underline{c} = C(h_0)$, $\overline{c} = C(\overline{h})$ and h_0 is the threshold type above which the firm serves consumers. From (6), the latter is defined as

$$\left[h_0 - \frac{1-\Phi(h_0)}{\phi(h_0)}\right] V_0(C(h_0)) - C(h_0) = 0. \tag{13}$$

To summarize, the model primitives are $[U_0(\cdot,\cdot), f_h(\cdot), f_0(\cdot,\cdot,\cdot,\cdot), h(\cdot,\cdot), \kappa, \gamma]$. The observations are (t, q_v, q_m). The econometric model is

$$\begin{aligned} t &= T[\kappa Q_v(h)^{\gamma} Q_m(h)^{1-\gamma}] \\ q_v &= \frac{Q_v(h)}{\epsilon_v}, \quad q_m = \frac{Q_m(h)}{\epsilon_m} \end{aligned}$$

where $h = h(\theta_v, \theta_m) \in [h_0, \overline{h}]$ is distributed as $\Phi(\cdot)$ with density $\phi(\cdot)$ given by (3), h_0 satisfies (13), $Q_v(\cdot)$ and $Q_m(\cdot)$ solve (10) with $C(\cdot)$ solving (12), and $T(\cdot)$ satisfies (11).

Figure 12 illustrates the cost-based tariff. The surface displays the cost-based tariff as a function of Q_v and Q_m. Some contours are drawn in the (Q_v, Q_m) plane. Because the tariff is cost-based, these contours can be interpreted as isocost curves. The dashed curve going through $Q = (Q_v, Q_m)$ is an indifference curve. The consumer chooses the combination (Q_v, Q_m) where his indifference curve is tangent to the isocost curve, i.e (Q_v, Q_m) satisfies

$$\frac{C_1(Q_v, Q_m)}{C_2(Q_v, Q_m)} = \frac{h(\theta_v, \theta_m) U_{01}(Q_v, Q_m)}{h(\theta_v, \theta_m) U_{02}(Q_v, Q_m)} = \frac{U_{01}(Q_v, Q_m)}{U_{02}(Q_v, Q_m)},$$

where the index 1 or 2 refers to the first or second argument for the partial derivative, respectively. When the payment and hence the cost increase, the bundle (Q_v, Q_m) varies along the curve displayed in Figure 12. The actual consumption pair (q_v, q_m) is represented in the figure as q, which deviates from Q due to (ϵ_v, ϵ_m) .

Because $T(\cdot)$ is strictly increasing, the payment t and the cost c are related through a one-to-one mapping. Since there is a one-to-one mapping between the cost c and the aggregate type h, we also have a one-to-one mapping between t and h. The latter property is crucial to our identification results. Moreover, for any $t \in [\underline{t}, \overline{t}]$, where $\underline{t} = t(h_0)$ and $\overline{t} = t(\overline{h})$, there exists a unique pair of quantities Q_v and Q_m such that $t = T(C(Q_v, Q_m))$. We also denote these two mappings by $Q_v(\cdot)$ and $Q_m(\cdot)$ whether the argument is t or h. Assuming that we can identify $(Q_v(\cdot), Q_m(\cdot))$ as shown below, several difficulties remain. In the previous literature on the identification of models with incomplete information such as auctions and nonlinear pricing, we exploit a one-to-one mapping between an unobserved type and an observed choice. For instance, in auctions, the one-to-one mapping between the unobserved bidder's private value and his observed bid is the key to identify the distribution of private values. See Guerre, Perrigne, and Vuong (2000) and Athey and Haile (2007) for a survey.[9] In the case of nonlinear pricing with a single product, the unobserved type is the consumer's willingness to pay and the observed choice is a quantity. This mapping combined

[9] When bidders are risk averse, the bidder's utility function is an additional primitive to identify. Because no other observation can be exploited in auctions, identification is achieved by exploiting some exogenous variation in the number of bidders as in Guerre, Perrigne, and Vuong (2009).

with the tariff as a function of quantity is used to identify the distribution of types and the consumer's utility. See Perrigne and Vuong (2011a). Because of the multidimensional screening problem, we have a cost-based tariff and thus an additional difficulty due to the unobserved cost. This problem is solved with the identification of the cost parameters κ and γ.

Identification of $V_0'(\cdot)$, κ, γ and $\phi(\cdot)$

We first address the identification of $Q_v(\cdot)$ and $Q_m(\cdot)$ as functions of the payment t. From the definition of Q_v and Q_m, we have

$$\log Q_v = \log q_v + \log \epsilon_v, \quad \log Q_m = \log q_m + \log \epsilon_m$$

Since (ϵ_v, ϵ_m) are unobserved, we make a standard location normalization on (ϵ_v, ϵ_m).

Assumption A5: *We have* $\mathrm{E}[\log \epsilon_v] = 0$ *and* $\mathrm{E}[\log \epsilon_m] = 0$*.*

Using the independence of h with (ϵ_v, ϵ_t) from A2-(ii) combined with the one-to-one mapping between t and h gives $\mathrm{E}[\log \epsilon_v | T = t] = 0$ and $\mathrm{E}[\log \epsilon_m | T = t] = 0$. Thus,

$$\begin{aligned} \log Q_v(t) &= \mathrm{E}[\log q_v + \log \epsilon_v | T = t] = \mathrm{E}[\log q_v | T = t], \\ \log Q_m(t) &= \mathrm{E}[\log q_m + \log \epsilon_m | T = t] = \mathrm{E}[\log q_m | T = t]. \end{aligned}$$

Thus $Q_v(\cdot)$ and $Q_m(\cdot)$ are identified as the regressions of $\log q_v$ and $\log q_m$ on t, respectively. Formally, we have

Proposition 2: *The functions $Q_v(\cdot)$ and $Q_m(\cdot)$ are identified on $[\underline{t}, \overline{t}]$.*

To clarify ideas, we first suppose that the cost c is observed. Let $G_c(\cdot)$ be its distribution on $[\underline{c}, \overline{c}]$. Thus, the tariff function $T(\cdot)$ is known as well from the observations on t and c. We can then rely on Perrigne and Vuong (2011a). Namely, we express the cost-based indirect utility $V_0(\cdot)$ and the unobserved aggregate taste as functions of the cost c of producing the bundle (Q_v, Q_m). These results are stated in the next lemma.

Lemma 1: *For any $h_\dagger \in [h_0, \overline{h}]$, the first-order conditions (11) and (12) are equivalent to*

$$V_0'(c) = \frac{T'(c)}{h_\dagger \xi(c)}, \quad h(c) = h_\dagger \xi(c), \tag{14}$$

for all $c \in [\underline{c}, \overline{c}]$, where

$$\xi(c) = \left(\frac{1 - G_c(c)}{1 - G_c(C(h_\dagger))} \right)^{\frac{1}{T'(c)} - 1} \exp \left\{ \int_{C(h_\dagger)}^{c} \frac{T''(x)}{T'(x)^2} \log(1 - G_c(x)) dx \right\}, \tag{15}$$

with $\xi(C(h_\dagger)) = 1$ and $\xi(\overline{c}) = \lim_{c \uparrow \overline{c}} \xi(c) = \overline{h}/h_\dagger$.

The proof of Lemma 1 follows Perrigne and Vuong (2011a) (Lemma 5) by noting that their marginal cost is equal to one here. Lemma 1 exploits the one-to-one mapping between the cost c of producing the bundle (Q_v, Q_m) and the consumer's aggregate type h. It is important to note that Lemma 1 applies only to individuals who are not excluded and hence who buy cellular phone services, i.e. whose aggregate type is larger than h_0.

Lemma 1 shows that the marginal indirect utility $V_0'(\cdot)$ and the unobserved aggregate type h are identified up to $h_\dagger$ from the observed cost and tariff suggesting a normalization on $h_\dagger$. Moreover, (15) involves the cost evaluated at $h_\dagger$, namely $C(h_\dagger) = \kappa Q_v(t_\dagger)^\gamma Q_m(t_\dagger)^{1-\gamma}$, where $h_\dagger = h(t_\dagger)$. Because we estimate nonparametrically $Q_v(\cdot)$ and $Q_m(\cdot)$, we choose $h_\dagger$ to be the median of the aggregate type distribution to minimize boundary effects and set it at 1.

Assumption A6: *The median of the subscriber's type is equal to 1.*

This identifies $V_0'(\cdot)$ and $h(\cdot)$ on $[\underline{c}, \overline{c}]$ when c is observed. Moreover, h_0 is identified from (14) and (15) by $h_0 = h(\underline{c}) = 2^{(1/T'(\underline{c}))-1}$.

In practice, however, the cost c is not observed. But we know that $t = T(q_v, q_m; \epsilon_v, \epsilon_m) = T(C(Q_v, Q_m)) = T(\kappa Q_v^\gamma Q_m^{1-\gamma})$, where $T(\cdot)$, κ and γ are unknown. The quantities (Q_v, Q_m) are not observed but can be recovered from the payment t using Proposition 2. The problem is reminiscent of single-index and transformation models. See e.g. Horowitz (1996) and Powell, Stock, and Stoker (1989). The next lemma shows that κ, γ and $T(\cdot)$ are identified.

Proposition 3: *The cost parameters γ and κ as well as the tariff $T(\cdot)$ are identified if and* *only if*

$$\left(Q_m'(\underline{t})-\frac{\log 2}{g_t(\underline{t})}Q_m''(\underline{t})-\frac{Q_m(\underline{t})}{\underline{t}}\right)\left(Q_v'(\underline{t})-\frac{\log 2}{g_t(\underline{t})}Q_v''(\underline{t})-\frac{Q_v(\underline{t})}{\underline{t}}\right)<0.$$

The proof is given in Appendix A. In particular, γ is identified by the lower boundary defined by (13) leading to the quadratic equation

$$\begin{aligned}&\gamma\frac{d\log Q_v(\underline{t})}{dt}+(1-\gamma)\frac{d\log Q_m(\underline{t})}{dt}-\frac{1}{\underline{t}}=\\&\frac{\log 2}{g_t(\underline{t})}\left[\left(\gamma\frac{d\log Q_v(\underline{t})}{dt}+(1-\gamma)\frac{d\log Q_m(\underline{t})}{dt}\right)^2+\gamma\frac{d^2\log Q_v(\underline{t})}{dt^2}+(1-\gamma)\frac{d^2\log Q_m(\underline{t})}{dt^2}\right],\end{aligned}\quad (16)$$

which has a unique solution under the condition of Proposition 3.[10] Moreover, κ is identified by exploiting (11) and (12) at the upper boundary leading to

$$\frac{1}{\kappa}=Q_v(\bar{t})^{\gamma}Q_m(\bar{t})^{1-\gamma}\left\{\gamma\frac{d\log Q_v(\bar{t})}{dt}+(1-\gamma)\frac{d\log Q_m(\bar{t})}{dt}\right\}. \quad (17)$$

To identify $T(\cdot)$, we know that $t=T(\kappa Q_v(t)^{\gamma}Q_m(t)^{1-\gamma})$ or equivalently $\log T^{-1}(t)=\log\kappa+\gamma\log Q_v(t)+(1-\gamma)\log Q_m(t)$ showing that $T^{-1}(\cdot)$ is identified using Proposition 2, namely

$$\log T^{-1}(t)=\log\kappa+\gamma\mathrm{E}[\log q_v|T=t]+(1-\gamma)\mathrm{E}[\log q_m|T=t], \quad (18)$$

for any $t\in[\underline{t},\bar{t}]$.

Once κ, γ and $T(\cdot)$ are identified, we use Lemma 1 to identify $V_0'(\cdot)$ and $h(\cdot)$ with the normalization in A6. The latter is then used to identify the subscribers type distribution $\Phi^*(\cdot)=[\Phi(\cdot)-\Phi(h_0)]/[1-\Phi(h_0)]$. This result is stated in the next proposition. The proof follows from Perrigne and Vuong (2011a) (Proposition 2).

Proposition 4: *The marginal indirect utility function $V_0'(\cdot)$ and the distribution of the subscribers' aggregate type $\Phi^*(\cdot)$ are identified on $[\underline{c},\bar{c}]$ and $[h_0,\bar{h}]$, respectively.*

As a matter of fact, the cost-based indirect utility function $V_0(\cdot)$ is identified on $[\underline{c},\bar{c}]$ because

[10]This condition can be checked ex post from the estimates of $Q_v(\cdot)$ and $Q_m(\cdot)$.

$V_0(c) = \int_{\underline{c}}^{c} V_0'(x)dx + V_0(\underline{c})$, where $V_0(\underline{c}) = \underline{t}/h_0$ from the boundary condition for the tariff. Moreover, if one knew the proportion of consumers $\Phi(h_0)$ who do not purchase cellular phone services, we could identifiy $\Phi(\cdot)$ on $[h_0, \overline{h}]$. In addition, we also identify the cost as a function of the bill, i.e. $C(t) = \kappa Q_v(t)^\gamma Q_m(t)^{1-\gamma}$ for $t \in [\underline{t}, \overline{t}]$.

Identification of $U_0(\cdot,\cdot)$, $f_h(\cdot)$ and $f_0(\cdot,\cdot,\cdot,\cdot)$

It remains to discuss the identification of the utility function $U_0(\cdot.\cdot)$ and the joint density of types $f(\cdot,\cdot,\cdot,\cdot)$. We have

$$V_0(c) = \max_{\kappa Q_v^\gamma Q_m^{1-\gamma} \leq c} U_0(Q_v, Q_m).$$

Because the optimal solution $(Q_v(c), Q_m(c))$ of this problem is equal to $(Q_v(t), Q_m(t))$ when $c = C(t)$, then

$$U_0(Q_v(t), Q_m(t)) = V_0[\kappa Q_v(t)^\gamma Q_m(t)^{1-\gamma}].$$

Hence, the utility function $U_0(\cdot,\cdot)$ is identified on the curve $\{(Q_v(t), Q_m(t)); t \in [\underline{t}, \overline{t}]\}$ depicted in Figure 5. The latter is similar to an income curve but with the difference that parallel budget lines are replaced by isocost curves. Without a parameterization of $U_0(\cdot,\cdot)$, we cannot infer further the complementarity and/or substitution between voice and message services because the domain where we identify the utility function is unlikely to allow us to derive the demand functions and estimate the cross-price elasticities.

Regarding the identification of $f(\cdot,\cdot,\cdot,\cdot)$, the model tells us that subscribers with the same aggregate type $h = h(\theta_v, \theta_m)$ have the same effective consumption levels Q_v and Q_m of voice and message services though consumers may differ through θ_v and θ_m. This arises from the pooling at equilibrium due to multi-dimensional screening. Thus, despite the identification of the aggregate type h, the function $h(\cdot,\cdot)$ is not identified. Moreover, even if $h(\cdot,\cdot)$ was known, we do not have any variation in the data that would allow us to identify θ_v and θ_m and hence their marginal densities. On the other hand, we have $\phi(h) = khf_h(h)$. Thus, $f_h(\cdot)$ is identified up to a multiplicative constant on $[h_0, \overline{h}]$ since the aggregate type h is identified on this interval following Lemma 1. Moreover, from Proposition 2, we identify

$\epsilon_v = Q_v(t)/q_v$ and $\epsilon_m = Q_m(t)/q_m$ since $Q_v(\cdot)$ and $Q_m(\cdot)$ are identified and (t, q_v, q_m) are observed. Thus the joint distribution of (ϵ_v, ϵ_m) conditional on $t \in [\underline{t}, \overline{t}]$ is identified. Since the distribution of payments is observed, it follows that the joint distribution of (ϵ_v, ϵ_m) is identified. The above results are summarized in the next proposition.

Proposition 5: *The utility function $U_0(\cdot,\cdot)$ and the type density $f(\cdot,\cdot,\cdot,\cdot)$ are partially identified. In particular, $U_0(\cdot,\cdot)$ is identified at values $\{(Q_v(t), Q_m(t)); t \in [\underline{t}, \overline{t}]\}$, while the joint density of (ϵ_v, ϵ_m) is identified.*

2.4.2 Nonparametric Estimation

The data consist of $(q_{vi}, q_{mi}, t_i), i = 1, \ldots, N$. Equations (14) and (15) provide the marginal cost-based indirect utility function $V_0'(\cdot)$ and the aggregated type $h(\cdot)$ as functionals of $T'(\cdot)$, $T''(\cdot)$ and $G_c(\cdot)$. We proceed in several steps. In a first step, we estimate the parameters κ and γ of the cost function and the tariff $T(\cdot)$ using Proposition 3. In particular, this step requires the estimation of the lower and upper boundaries of the tariff, its density at the lower boundary and the functions $Q_v(\cdot)$ and $Q_m(\cdot)$ as well as their first and second derivatives evaluated at these boundaries. In a second step, we use Lemma 1 to estimate $V_0'(\cdot)$ and $h(\cdot)$. We then construct a sample of pseudo aggregate types h and estimate its truncated density $\phi^*(\cdot)$. In a third step, we estimate $U_0(\cdot,\cdot)$ and the joint density of (ϵ_v, ϵ_m) using Proposition 5.

Estimation of κ, γ and $T(\cdot)$

The first step consists in estimating κ, γ and $T^{-1}(\cdot)$. We first estimate κ and γ using (16) and (17). These two equations requires estimators for $\underline{t}$, $\overline{t}$, $g_t(\underline{t})$, $Q_v(\cdot)$, $Q_m(\cdot)$ and their first and second derivatives at these two boundaries. We use consistent estimators $\hat{\underline{t}}$ and $\hat{\underline{t}}$ such as the minimum and maximum of the observed bills t_i. We estimate the bill density $g_t(\cdot)$ by kernel estimator. Because of boundary effects, we use a one-sided kernel. Following Campo, Guerre, Perrigne, and Vuong (2011), we use $K(x) = (-6x+4)\mathbb{1}(0 \le x \le 1)$ giving

$$\hat{g}_t(\hat{\underline{t}}) = \frac{1}{Nb_t} \sum_{i=1}^{N} K\left(\frac{\hat{\underline{t}} - t_i}{b_t}\right), \tag{19}$$

where b_t is a bandwidth.

To estimate $Q_v(\cdot)$ and $Q_m(\cdot)$, we use sieve estimators so as to impose easily some possible shape constraints. Specifically, we use the constrained regression splines proposed by Dole (1999) where

$$Q_j(t) \approx \beta_{j0} + \beta_{j1} t + \sum_{k=1}^{K_N} \delta_{jk} \psi_k(t)$$

where K_N is the number of interior knots, for $j = v, m$ and $t \in [\underline{t}, \overline{t}]$. The range $[\underline{t}, \overline{t}]$ is partioned into $K_N + 1$ bins of the form $[\tau_{k-1}, \tau_k)$ for $k = 1, \ldots, K_N + 1$ with $\tau_0 = \underline{t}$ and $\tau_{K_N+1} = \overline{t}$. The basis function $\psi_k(\cdot)$ is

$$\psi_k(t) = \begin{cases} 0 & \text{if } t \in [-\infty, \tau_{k-1}] \\ (t - \tau_{k-1})^3/[6(\tau_k - \tau_{k-1})] & \text{if } t \in [\tau_{k-1}, \tau_k] \\ ((t - \tau_{k+1})^3/[6(\tau_k - \tau_{k+1})]) + a_1 t + a_0 & \text{if } t \in [\tau_k, \tau_{k+1}] \\ a_1 t + a_0 & \text{if } t \in [\tau_{k+1}, +\infty] \end{cases}$$

where $a_1 = (\tau_{k+1} - \tau_{k-1})/2$ and $a_0 = ((\tau_k - \tau_{k-1})^2 - (\tau_k - \tau_{k+1})^2 + 3\tau_k(\tau_{k+1} - \tau_{k-1}))/6$. The estimator $\hat{Q}_j(\cdot)$ is obtained from $\log Q_j(t) = \mathrm{E}[\log q_j | T = t]$ leading to the least square minimization

$$\min_{\delta_{jN}} \frac{1}{N} \sum_{i=1}^{N} \left(\log q_{ji} - \log \left(\beta_{j0} + \beta_{j1} t_i + \sum_{k=1}^{K_N} \delta_{jk} \psi_k(t_i) \right) \right)^2, \tag{20}$$

where $\delta_{jN} = (\beta_{j0}, \beta_{j1}, \delta_{j1}, \ldots, \delta_{jK_N})$. Thus, $\hat{Q}_j(t) = \hat{\beta}_{j0} + \hat{\beta}_{j1} t_i + \sum_{k=1}^{K_N} \hat{\delta}_{jk} \psi_k(t_i)$. Thus, its first and second derivatives are $\hat{Q}'_j(t) = \hat{\beta}_{j1} + \sum_{k=1}^{K_N} \hat{\delta}_{jk} \psi'_k(t)$ and $\hat{Q}''_j(t) = \sum_{k=1}^{K_N} \hat{\delta}_{jk} \psi''_k(t)$, respectively. These expressions are used to estimate $d \log \hat{Q}_j(t)/dt = \hat{Q}'_j(t)/\hat{Q}_j(t)$ and $d^2 \log \hat{Q}_j(t)/dt^2 = (\hat{Q}''_j(t)/\hat{Q}_j(t)) - (\hat{Q}'_j(t)/\hat{Q}_j(t))^2$. These are then evaluated at $\hat{\underline{t}}$ and $\hat{\overline{t}}$ for $j = v, m$. Inserting these estimates in (16) and solving for γ gives $\hat{\gamma}$. Using $\hat{\gamma}$ in (17) gives $\hat{\kappa}$.

To estimate $T^{-1}(\cdot)$, we use (18). To impose that $T^{-1}(\cdot)$ is increasing and convex, we estimate $T^{-1}(\cdot)$ by a constrained sieve estimator. Specifically,

$$T^{-1}(t) \approx \beta_{t0} + \beta_{t1} t + \sum_{k=1}^{K_N} \delta_{tk} \psi_k(t)$$

where K_N is the number of interior knots and the $\psi_k(\cdot)$s are the basis functions defined above. Let $\delta_{tN} = (\beta_{t0}, \beta_{t1}, \delta_{t1}, \dots, \delta_{tK_N})$ and $\Delta_{tN} = \{\delta_{tN} : \beta_1 \geq 0, \delta_1 \geq 0, \dots, \delta_{K_N} \geq 0\}$. The estimator $\hat{T}^{-1}(\cdot)$ is obtained from (18) replacing κ and γ by their estimates $\hat{\kappa}$ and $\hat{\gamma}$ leading to the least square minimization

$$\min_{\delta_{tN} \in \Delta_{tN}} \frac{1}{N} \sum_{i=1}^{N} \left(\log \hat{\kappa} + \hat{\gamma} \log q_{vi} + (1-\hat{\gamma}) \log q_{mi} - \log \left(\beta_{t0} + \beta_{t1} t_i + \sum_{k=1}^{K_N} \delta_{tk} \psi_k(t_i) \right) \right)^2, \tag{21}$$

giving $\hat{T}^{-1}(t) = \hat{\beta}_{t0} + \hat{\beta}_{t1} t + \sum_{k=1}^{K_N} \hat{\delta}_{tk} \psi_k(t)$ for $t \in [\underline{t}, \overline{t}]$.[11]

Estimation of $V_0'(\cdot)$, $h(\cdot)$ and $\phi^*(\cdot)$

One could use (14) and (15) to estimate $V_0'(\cdot)$ and $h(\cdot)$. These are functions of the cost c that can be estimated from $c = \kappa Q_v^{\gamma}(t) Q_m^{1-\gamma}(t)$. To minimize estimation errors, we prefer instead to rewrite (14) and (15) directly from the observed bills t as there is a one-to-one mapping between c and t. In particular, for any $c \in [\underline{c}, \overline{c}]$, we have $G_c(c) = \Pr(\tilde{c} \leq c) = \Pr[T(\tilde{c}) \leq T(c)] \equiv G_t(T(c))$, where $G_t(\cdot)$ denotes the distribution of consumers' bills. Thus making the change of variable $c = T^{-1}(t)$, we can rewrite $\xi(\cdot)$ as

$$\begin{aligned} \xi(T^{-1}(t)) &= [2(1-G_t(t))]^{T^{-1\prime}(t)-1} \exp\left\{ \int_{t_\dagger}^{t} \frac{T''[T^{-1}(x)]}{T'[T^{-1}(x)]^2} \log[1-G_t(x)] T^{-1\prime}(x) dx \right\} \\ &= [2(1-G_t(t))]^{T^{-1\prime}(t)-1} \exp\left\{ -\int_{t_\dagger}^{t} T^{-1\prime\prime}(x) \log[1-G_t(x)] dx \right\}, \end{aligned} \tag{22}$$

for $t \in [\underline{t}, \overline{t}]$, where we have used $G_t(t_\dagger) = 1/2$ as $t_\dagger$ is the median, $T^{-1\prime}(t) = 1/T'[T^{-1}(t)]$

[11] An alternative estimation method for the first step would be to estimate simultaneously κ, γ and $T^{-1}(\cdot)$. In this case, δ_{tN} would also include the parameters (κ, γ) while Δ_{tN} would need to include the additional identifying conditions for κ and γ. For instance, (17) is equivalent to $T^{-1\prime}(\overline{t}) = 1$ or $\beta_1 + \sum_{k=1}^{K_N} \delta_k \psi_k'(\overline{t}) = 1$. Thus the estimator would rely on a minimum distance estimator based on the moment $\mathrm{E}\left[\mathrm{E}^2[\log\kappa + \gamma \log q_v + (1-\gamma)\log q_m - \log T^{-1}(t_i) | T = t_i]\right] = 0$ as proposed by Ai and Chen (2003). Alternatively, we could use Lavergne and Patilea (2010) estimator based on the moment $\mathrm{E}\big[(\log\kappa + \gamma \log q_v + (1-\gamma)\log q_m - \log T^{-1}(t_i)) \mathrm{E}[\gamma \log q_v + (1-\gamma)\log q_m - \log T^{-1}(t_i) | T = t_i]\big] = 0$

and $T^{-1\prime\prime}(t) = -(T''[T^{-1}(t)]/T'[T^{-1}(t)]^2) \times T^{-1\prime}(t)$.

For every $t \in [\underline{t}, \bar{t})$, the estimator of $\xi(\cdot)$ is obtained by replacing $T^{-1}(\cdot)$ by its estimator obtained from the first step and $G_t(\cdot)$ by its empirical distribution

$$\hat{G}_t(t) = \frac{1}{N}\sum_{i=1}^{N} 1\!\!I(t_i \leq t).$$

Because $\hat{G}_t(\cdot)$ is a step function with steps at the ordered observations $t^1, \ldots, t^N$, the integral in (22) can be written as a finite sum of integrals. On each of these integrals, $\log[1 - \hat{G}_t(\cdot)]$ is a constant, while the primitive of $T^{-1\prime\prime}(\cdot)$ is $T^{-1\prime}(\cdot)$. For t values above the median, i.e. $t \in [t_\dagger, \bar{t})$ with $t_\dagger = t^{[N/2]}$, we obtain

$$\begin{aligned} \hat{\xi}(\hat{T}^{-1}(t)) \quad = \quad & [2(1-\hat{G}_t(t))]^{\hat{T}^{-1\prime}(t)-1} \exp\left\{ - \sum_{k=[N/2]}^{j} \left(\hat{T}^{-1\prime}(t^{k+1}) - \hat{T}^{-1\prime}(t^k)\right) \log(1 - \hat{G}_t(t^k)) \right. \\ & \left. - \left(\hat{T}^{-1\prime}(t) - \hat{T}^{-1\prime}(t^j)\right) \log(1 - \hat{G}_t(t^j)) \right\} \end{aligned}$$

for $t \in [t^j, t^{j+1})$ and $j \geq [N/2]$. Similarly, for values below the median, we obtain

$$\begin{aligned} \hat{\xi}(\hat{T}^{-1}(t)) \quad = \quad & [2(1-\hat{G}_t(t))]^{\hat{T}^{-1\prime}(t)-1} \exp\left\{ \sum_{k=j+1}^{[N/2]} \left(\hat{T}^{-1\prime}(t^{k+1}) - \hat{T}^{-1\prime}(t^k)\right) \log(1 - \hat{G}_t(t^k)) \right. \\ & \left. + \left(\hat{T}^{-1\prime}(t^{j+1}) - \hat{T}^{-1\prime}(t)\right) \log(1 - \hat{G}_t(t^j)) \right\} \end{aligned}$$

for $t \in [t^j, t^{j+1})$ and $j < [N/2]$.

From (14) we then estimate $V_0'(\cdot)$ and $h(\cdot)$ by

$$\hat{V}_0'(\hat{T}^{-1}(t)) = \frac{1}{\hat{T}^{-1\prime}(t)\hat{\xi}(\hat{T}^{-1}(t))}, \quad \hat{h}(\hat{T}^{-1}(t)) = \hat{\xi}(\hat{T}^{-1}(t)) \tag{23}$$

where $t = \hat{T}(c)$, for $c \in [\underline{c}, \bar{c}) = [T^{-1}(\underline{t}), T^{-1}(\bar{t}))$. We can then construct a sample of pseudo values $\hat{h}_i = \hat{\xi}(\hat{T}^{-1}(t_i)), i = 1, \ldots, N$. We use these pseudo sample to estimate nonparametrically the density of aggregated type $\hat{\phi}^*(\cdot)$ on a subset of its support $[h_0, \overline{h}]$.

Namely,

$$\hat{\phi}^*(h) = \frac{1}{Nh_h} \sum_{i=1}^{N} K\left(\frac{h - \hat{h}_i}{h_h}\right) \tag{24}$$

for $h \in (h_0, \overline{h})$, h_h is a bandwidth and $K(\cdot)$ is a triweight kernel.

Estimation of $U_0(\cdot,\cdot)$ and $f_{\epsilon_v \epsilon_m}(\cdot,\cdot)$

Following Proposition 5, the utility function $U_0(\cdot,\cdot)$ can be estimated at $(Q_v(t), Q_m(t))$, which can be replaced by their estimates $\hat{Q}_v(t)$ and $\hat{Q}_m(t)$ obtained from the first step. Specifically, we have $\hat{U}_0(\hat{Q}_v(t), \hat{Q}_s(t)) = \hat{V}_0(\hat{\kappa}\hat{Q}_v(t)^{\hat{\gamma}}\hat{Q}_m(t)^{1-\hat{\gamma}}) = \hat{V}_0(\hat{C}(t)) \equiv \hat{V}_0(t)$ for $t \in [\underline{t}, \overline{t}]$, with

$$\begin{aligned} \hat{V}_0(\hat{C}(t)) &\equiv \int_{\hat{\underline{c}}}^{\hat{C}(t)} \hat{V}_0'(x)dx + \hat{V}_0(\hat{\underline{c}}) \\ &= \int_{\hat{\underline{t}}}^{t} \hat{V}_0'[\hat{T}^{-1}(y)]\hat{T}^{-1\prime}(y)dy + \hat{V}_0(\hat{\underline{t}}) \\ &= \int_{\hat{\underline{t}}}^{t} \frac{1}{\hat{\xi}[\hat{T}^{-1}(y)]}dy + \hat{V}_0(\hat{\underline{t}}), \end{aligned} \tag{25}$$

where the second equality uses the change of variable $x = \hat{T}^{-1}(y)$ and the third equality follows from (22). Moreover, $\hat{V}_0(\hat{\underline{t}}) = \hat{\underline{t}}/\hat{h}_0 = \hat{\underline{t}}2^{1-\hat{T}^{-1\prime}(\hat{\underline{t}})}$. The latter comes from the boundary condition combined with $h_0 = 2^{(1/T'(\underline{c}))-1}$ and $\underline{c} = T^{-1}(\underline{t})$.

Lastly, regarding the density of (ϵ_v, ϵ_m), from Proposition 2 we can estimate ϵ_{vi} and ϵ_{mi} by $\hat{\epsilon}_{vi} = \hat{Q}_v(t_i)/q_{vi}$ and $\hat{\epsilon}_{mi} = \hat{Q}_m(t_i)/q_{mi}$, respectively. Using this pseudo sample, we estimate nonparametrically $f_{\epsilon_v\epsilon_m}(\cdot,\cdot)$ by

$$\hat{f}_{\epsilon_v\epsilon_m}(e_v, e_m) = \frac{1}{Nh_e^2} \sum_{i=1}^{N} K\left(\frac{e_v - \hat{\epsilon}_{vi}}{h_e}\right) K\left(\frac{e_m - \hat{\epsilon}_{mi}}{h_e}\right) \tag{26}$$

for $(e_v, e_m) \in (\underline{\epsilon}_v, \overline{\epsilon}_v) \times (\underline{\epsilon}_m, \overline{\epsilon}_m)$, where $K(\cdot)$ is the triweight kernel and h_e is a bandwidth.

2.5 Empirical Results

The first step consists in estimating κ, γ and $T^{-1}(\cdot)$. The estimates of κ and γ are 0.2232 and 0.7544, respectively. This means that increasing by 1% the quantity of voice Q_v increases the production cost by 0.75%, while increasing the quantity of message Q_m by 1% increases the cost by 0.25%, respectively. The quantity Q_m includes MMS and other data features which are quite costly to produce. Figure 6 displays the concave cost-based tariff $\hat{T}(c)$.

The second step consists in estimating the marginal indirect utility $V_0'(\cdot)$, the aggregate type function $h(\cdot)$ and its truncated density $\phi^*(\cdot)$. Figure 5 displays the former two. Both satisfy the assumptions of the theoretical model, i.e. $\hat{V}_0'(\cdot)$ is decreasing in cost, while $\hat{h}(\cdot)$ is increasing in cost. In Figure 7, the density of aggregate types truncated at $\hat{h}_0$ is unimodal and displays an important skewness. Figure 8 displays $1 - \Phi^*(h)/[h\phi^*(h)]$, which appears to be increasing thereby satisfying A3. Since $t = T(c)$, we can also represent the payoff as a function of t. Figure 9 displays the estimated indirect base utility $\hat{V}_0(\cdot)$ and the payoff $\hat{h}(t)\hat{V}_0(t)$. Both are increasing in t as assumed by the theoretical model. The difference between $\hat{h}(t)\hat{V}_0(t)$ and t (the 45 degrees line) measures the rent left to the consumers due to asymmetric information. This rent is increasing in t, which is predicted. As displayed by Figure 9, this rent increase is mainly due to the dramatic increase in the aggregate type and modestly to the increase in the base indirect utility. This suggests that some subscribers with a high willingness to pay for cellular phone services enjoy a much larger payoff relative to subscribers with a lower taste for cellular phone services. As displayed in Table 3, the informational rents show an important variability as shown by a coefficient of variation larger than one and a high skewness. We remark the important difference between the median at \$6.55 and the mean at \$15.17. When computing the informational rent as a fraction of the payment, this proportion is on average 0.31 with a median value at 0.22. Figure 10 displays its estimated density. For the company, the ratio of the total informational rent left to consumers by the sum of payments and informational rent gives 30.45%. This measures the cost of asymmetric information for the firm.

Regarding unobserved heterogeneity, Table 4 provides some summary statistics on $\hat{\epsilon_v}$ and $\hat{\epsilon_m}$. In the model, they capture the extra features that subscribers add to customize their voice and message services. Though the median is smaller for message than for voice, the average value is three times larger for the former than for the latter. Figure 11 displays the two marginal densities estimated nonparametrically and confirms these results. The density of ϵ_v is more peaked than that of ϵ_m. Here again, a few percent of subscribers are heavy consumers of extra features even if their voice and message consumptions q_v and q_m are relatively small. Although we do not have detailed data on these extra features, they are likely to be related to MMS and other data features, which are becoming more popular amoung cellular phone users. On the other hand, the correlation between the two unobserved heterogeneity terms is quite small of the order of 0.03. Additional data on extra features would help to refine the model and the results.

2.6 Conclusion

This paper studies multiproduct nonlinear pricing in the cellular phone industry with voice and message services. We develop a model based on Armstrong (1996) in which the unknown types of consumers are aggregated while the firm designs an optimal cost-based tariff. Moreover, given that the consumers add a large number of unobserved extra features, we introduce two terms of unobserved heterogeneity for voice and message add-ons. Our model defines two one-to-one mappings between the unknown aggregate type to the cost and the cost to the payment. Under some identifying assumptions such as a parameterization of the cost function, we show that the primitives of the model are identified from observables. We develop a semiparametric estimation method to recover the model primitives and apply it to data from a major mobile service provider in Asia. The empirical results support the model and display an important heterogeneity in types and unobserved heterogeneity. The cost of asymmetric information for the firm is assessed.

Chapter 3

Bundling and Nonlinear Pricing in Telecommunications

Ah, the Internet! The source of so much goodness. The font of e-mail, news, chat, TV, blogs, books and Facebook. What would we do without it?

– David Pogue (The New York Times. March 21, 2012)

3.1 Introduction

Bundling and nonlinear pricing have become increasingly prevalent practices in sectors such as travel services, retail and telecommunications. Many telecommunication service providers are now offering bundled service packages with internet, phone and cable TV. Moreover, they often offer nonlinear tariffs by providing additional discounts when subscribers buy larger quantities. In this paper, I study bundling and nonlinear pricing by a multiproduct firm that sells its products to customers with privately known multidimensional types. I develop a new model that endogenizes both the firm's bundling and pricing decisions. It explains which bundling strategy should be adopted by the firm: component pricing, pure bundling, semi-mixed bundling or mixed bundling. In particular, the complementarity utility effect, the cost saving effect and the dependence between the multidimensional types will play crucial roles. The existing theoretical literature, by contrast, has mainly focused on the benefits of mixed/pure bundling over component pricing. See, e.g., Adams and Yellen (1976), McAfee, McMillan, and Whinston (1989) and Salinger (1995). In addition, while the existing literature mostly addresses bundling and nonlinear pricing separately, my model incorporates both simultaneously as recently done by Armstrong and Vickers (2010). An illustration on China Telecom data with internet and phone services empirically assesses the benefits to the firm of adopting such a pricing strategy as well as the consumers' gain.

There is an extensive theoretical and empirical literature on bundling and nonlinear pricing within an incomplete information framework. Regarding nonlinear pricing, Armstrong (2006) and Stole (2007) provide recent theoretical surveys. In particular, Armstrong (1996) and Rochet and Chone (1998) show that nonlinear pricing for a multiproduct firm is a complex problem because of multidimensional screening. On the empirical side, Leslie (2004), McManus (2007) and Crawford and Shum (2007) render this problem one-dimensional by, e.g., considering unit-demand consumers or homogenizing the products. Using convenient parameterization of model primitives, Ivaldi and Martimort (1994) and Miravete and Röller (2004) endogenize the firm's pricing decision. Recently, Luo, Perrigne, and Vuong (2012) rely on Armstrong (1996) model to analyze nonlinear pricing of mobile voice and messaging services.[1]

Regarding bundling, the early theoretical literature considered a benchmark case with two single unit products and additive separable utility. See, e.g., Adams and Yellen (1976), McAfee, McMillan, and Whinston (1989) and Salinger (1995). The main condition under which pure or mixed bundling is preferred by the firm over component pricing relates to the correlation of types for the two products. Considering a large number of products, Armstrong (1999) and Bakos and Brynjolfsson (1999) show that the firm can approximate the first-best by bundling. Chu, Leslie, and Sorensen (2011) provide a simulation of bundle-size pricing. On the empirical side, Crawford and Yurukoglu (2012) and Ho, Ho, and Mortimer (2012) consider the firm's bundling decision as exogenous and estimate the welfare

[1]Nonlinear pricing for a single product firm leads to a closed-form solution. See, e.g., Maskin and Riley (1984). Most of the empirical literature, including Leslie (2004) and McManus (2007), uses discrete choice models while considering prices exogenous. While endogenizing the price, Perrigne and Vuong (2011a) show that the model primitives are identified and develop a nonparametric estimation method for nonlinear pricing models.

effects of bundling between upstream and downstream firms.

I develop a model combining bundling and nonlinear pricing to analyze data from a major telecommunication company in China selling internet and phone services.[2] Assuming a separable utility into the benefit of consuming phone service only and the complementary benefit of using both internet and phone services, as well as positively dependent types for internet and phone services, I exploit the discrete level of internet service to solve the multidimensional screening problem. In particular, the number of effective incentive compatibility constraints is significantly reduced. This allows me to characterize the optimal exclusion conditions, the assignment schedules and the tariff functions in an equivalent one-dimensional formulation. Specifically, the provider offers usage-based nonlinear tariffs for phone service and a fixed-fee for internet-only users. The curvature of the tariffs for phone service will differ according to the internet service level. In other words, bundling enables the provider to further discriminate consumers choosing different levels of internet. In addition, I provide the conditions on the primitives under which the firm optimally chooses his bundling strategy, i.e., component pricing, pure bundling, semi-mixed bundling or mixed bundling. Bundling is more likely to dominate component pricing when the cost to provide phone service is lower or consumers value phone service more highly.

I study the identification of the model structure from observables: the price paid and the level of consumption. Under a parameterization of the cost function and a multiplicative separability of the utility function in the willingness-to-pay for phone service, I show that the primitives are identified. In particular, the complementary utility function is identified by exploiting the phone usage and tariff variation across consumers adopting different internet levels. While cost saving effects are identified by exploiting the firm's optimal exclusion conditions, the dependence between the two types is identified using the one-to-one mapping between phone usage and the corresponding consumer type.

In view of these identification results, I propose a three-step semiparametric estimation method involving a new regression spline estimator under both monotonicity and bound

[2]While considering a duopoly and multi-unit demand, Armstrong and Vickers (2010) establish the conditions on model primitives under which nonlinear pricing and mixed bundling generate more profit than linear pricing. They also analyze the harming effects on consumer surplus.

restrictions. This estimator is used to estimate the distribution of the consumers' tastes. It has several advantages over classical kernel estimators also used in auctions to estimate the bidders' private value distribution. See, e.g., Guerre, Perrigne, and Vuong (2000). Specifically, the difference between the willingness-to-pay and the inverse hazard rate function, which is usually assumed to be increasing, contains information on the underlying screening mechanism. Also, this difference is bounded by the identity function. A key advantage of this regression spline estimator is that it easily incorporates monotonicity and bound restrictions suggested by theory. See also Chen (2007). Moreover, since the support of consumer taste is a compact interval and exclusion is likely in multiproduct nonlinear pricing models as shown by Armstrong (1996), a kernel estimator can be biased at the boundary. Sieve estimators are known not to suffer from boundary effects.

The analysis of China Telecom data shows that (i) internet and phone services tend to be substitutes, (ii) the additional fixed cost to bundle internet with phone service is small, and (iii) a higher speed internet adopter tends to have a higher willingness-to-pay for phone service. Counterfactuals assess the gain for both the firm and consumers from bundling internet and phone services. My simulation results show that unbundling would lead to a 10.14% decrease in firm's profit and a 17.18% decrease in consumer surplus.

The paper is organized as follows. Section 2 presents the data. Section 3 introduces the model, while Section 4 studies the identification of the model primitives and develops a semiparametric estimation procedure. Section 5 presents estimation results and counterfactuals. Section 6 concludes. An appendix collects the proofs.

3.2 Data

The Chinese telecommunications industry is dominated by three state-owned firms: China Telecom, China Unicom and China Mobile. Table C.1 gives the nationwide market structure by the number of subscribers. While China Mobile has dominated mobile services since its inception, China Telecom and China Unicom roughly divide the territory in half for internet and land line services: China Telecom in southern China and China Unicom in northern China.

I collected data from China Telecom in a major metropolitan area, where it enjoys a market share of 85% for both internet and land line subscriptions. The sample is composed of new residential subscribers in August 2009, who receive internet and phone services (including land line and mobile) through the One Home label. For the month of September 2009, the data contain for each subscriber: his/her choice of internet service, the total number of minutes used and the amount paid. There are two internet levels, 1 Mbps or 2 Mbps, resulting in three possible bundles with phone service: phone service only (no internet), a bundle with 1 Mbps internet and a bundle with 2 Mbps internet. Table C.2 provides summary statistics on the bill measured in RMB and the number of phone minutes by internet choice with the corresponding number of subscribers. The bill paid by a consumer combines internet, different types of phone calls, roaming charges, extra fees for peak hours usage and several add-ons such as voice mail service, music on hold, ring tones, news, etc. The data do not provide detailed information on these extra features and the corresponding prices. All these extra features explain the important variability of the per minute rate. The consumption of phone calls tends to increase with the level of internet. I remark that the per minute rate tends also to increase with the level of internet.

China Telecom implements mixed bundling, i.e., internet and phone services are offered either separately or in a bundle. The firm charges a fixed fee to internet-only users and a usage-based tariff to bundle users. Specifically, the monthly fixed fee is 78 and 88 RMB for 1 and 2 Mbps, respectively. The usage-based tariff differs with the level of internet and is nonlinear. Table C.3 provides the regression of the bill on the number of total minutes and its square for each bundle. The three tariffs are increasing and concave, i.e., consumers pay a lower price per minute of call time when they consume more. Moreover, subscribers tend to pay more for the same amount of phone calls when they choose a higher internet speed. To see this, I calculate the mean of the per minute rate for the three bundles: 21.29 cents with no internet, 24.96 cents with 1 Mbps internet, and 27.27 cents with 2 Mbps internet. Table C.3 suggests that bundling enables the provider to discriminate further phone service users depending on their choice of internet.

Given that I need the tariff functions in view of Sections 3 and 4, I follow Luo (2011) method to estimate the tariff function for each bundle while taking into account unobserved add-ons and features. Details can be found in Appendix B. Figure C.3 displays the resulting three tariff functions denoted by $T_0(\cdot), T_1(\cdot), T_2(\cdot)$ for the three bundles. I then construct a quantity of phone usage $q \equiv T_j^{-1}(t)$, where t is his/her payment and j is his/her internet choice. The quantity q aggregates all observed quantity of minutes as well as the unobserved phone services chosen by the consumer. In Section 4 (on identification and estimation), I consider that $(t, q, j, T_j(\cdot))$ are the observables.

3.3 The Model

3.3.1 Assumptions and Notations

In view of the discussion in Section 2, I consider a monopoly provider selling internet and phone services as separate products or in a bundle. Internet is offered at several speed levels, denoted by $j \in \mathcal{J}$, where $\mathcal{J} \equiv \{0, 1, 2, \ldots, J\}$ is the possible choice set of internet levels with 0 denoting no internet.[3] Phone service is measured by $q \in I\!R^+$. Due to implementation difficulties of random contracts, the provider offers non-random nonlinear pricing schedules of the general form $T(q, j)$, with $q \in I\!R^+$ and $j \in \mathcal{J}$.

A consumer is characterized by a vector of types $(\theta, \beta) \in \Theta \equiv [\underline{\theta}, \overline{\theta}] \times [\underline{\beta}, \overline{\beta}]$, where $0 < \underline{\theta} < \overline{\theta} < \infty$ and $\underline{\beta} \leq 0 < \overline{\beta} < \infty$. The type θ represents his taste or willingness-to-pay for phone service and β defines the minimum internet need above which he will consider buying internet. The latter can be nonpositive because some consumers may have negative perspectives on internet. The vector (θ, β) is private information. That is, the provider does not know the consumer' types but knows the joint distribution $F(\cdot, \cdot)$ on $[\underline{\theta}, \overline{\theta}] \times [\underline{\beta}, \overline{\beta}]$.[4]

The consumer chooses internet speed and phone usage. Following Dubin and McFadden (1984) and Hanemann (1984), I assume that there is no uncertainty in the decision on the

[3]This choice set is exogenous given by technological constraint. See Mazzeo (2002) and Seim (2006) for endogenizing the product decisions of the firm.

[4]Because modeling competition is out of the scope of this paper, we can view θ as a sufficient statistic that summarizes preferences of the consumer for phone service by China Telecom and its competitors. Moreover, I do not consider uncertainty on types, which leads to a two-stage model. See, e.g., Miravete (2002), Miravete (2005), Narayanan, Chintagunta, and Miravete (2007), Economides, Seim, and Viard (2008) and Grubb and Osborne (2012).

continuous variable (phone usage) at the time of the choice for internet. In other words, both choices are made simultaneously. I make the following assumption on the consumer's utility function.

Assumption 1. *A (θ, β) agent consuming (q, j) gets utility*

$$U(q, j; \theta, \beta) = \begin{cases} U(q, j; \theta) & \text{for } j \geq \beta, \\ -\infty & \text{for } j < \beta. \end{cases}$$

Assumption 1 prevents the consumer to choose a bundle with an internet speed that falls below his minimum need β. It is analogous to the absolute spending limit assumption in Che and Gale (2000). The value β results from the consumer's internet use such as sending emails, shopping online, playing games or streaming movies. Internet connection speed determines whether these applications will run effectively. Thus I assume failing to meet the minimum internet need leads to a very large disutility. Once the minimum internet speed is satisfied, the underlying family of indifference curves of (q, j) bundles can be described by the variation of a single parameter, i.e., the taste for phone service θ.[5]

Assumption 1 has several advantages. First, it makes the optimal selling mechanism with multidimensional types tractable. See e.g. Armstrong (1996), Rochet and Chone (1998) and Rochet and Stole (2003) for mechanism design with multidimensional types. Second, while maintaining multidimensional types, this general utility function allows complementarity between the two products as well as dependence between the types θ and β. Consequently, all the possible scenarios of bundling may arise at the equilibrium such as component pricing, pure bundling, semi-mixed bundling and mixed bundling. See footnote 7 for a discussion on the generalization of my results with a multi-dimensional type θ.

A (θ, β) consumer chooses a quantity of phone service and internet level (q, j) to maxi-

[5] An alternative assumption would be to consider the utility as $U(q, j; \theta) - \delta(\beta)$ for $j < \beta$ where $\delta(\beta)$ captures the disutility for not getting the desired amount of internet speed. If $\delta(\beta)$ is large enough, Proposition 1 extends resulting in the same optimal selling mechanism. However, the proofs of Lemmas 1 and 3 wound be significantly longer.

mize his payoff

$$\max_{q\in I\!R^{+},j\in\mathcal{J}} \quad U(q,j;\theta)-T(q,j)$$
$$s.t. \quad j\geq\beta.$$

The firm needs to design a price schedule $T(\cdot,\cdot)$ that maximizes his expected profit. Without loss of generality, I apply the Revelation Principle. In particular, any implementable allocation achieved with a price schedule $T(\cdot,\cdot)$ can also be achieved with a truthful direct mechanism of the form $(t(\cdot,\cdot),q(\cdot,\cdot),j(\cdot,\cdot))$. This mechanism specifies the payment made $t(\theta,\beta)$, the quantity of phone service $q(\theta,\beta)$ and internet speed $j(\theta,\beta)$ for a (θ,β) consumer. As information goods, internet and phone services involve very small variable production costs but substantial transaction costs per customer, such as usage recording, billing and customer service. Thus I assume that the provider's total cost is additively separable across consumers. The cost to serve a consumer with a bundle (q,j) is denoted as $c(q,j)$.[6]

The optimal selling mechanism solves

$$\begin{aligned}\max_{t(\cdot,\cdot),q(\cdot,\cdot),j(\cdot,\cdot)} \quad & \int_{\Theta}\Big[t(\theta,\beta)-c\big(q(\theta,\beta),j(\theta,\beta)\big)\Big]f(\theta,\beta)d\theta d\beta \\ s.t. \quad & U(q(\theta,\beta),j(\theta,\beta);\theta)-t(\theta,\beta)\geq U(q(\tilde{\theta},\tilde{\beta}),j(\tilde{\theta},\tilde{\beta});\theta)-t(\tilde{\theta},\tilde{\beta}), \\ & U(q(\theta,\beta),j(\theta,\beta);\theta)-t(\theta,\beta)\geq 0, \\ & j(\theta,\beta)\geq\beta,\end{aligned}$$

for all $(\theta,\beta)\in\Theta$ and $(\tilde{\theta},\tilde{\beta})\in\Theta$ such that $j(\tilde{\theta},\tilde{\beta})\geq\beta$. The first inequality is the incentive compatibility (IC) constraint, which requires that the consumer truthfully reports his types. The second inequality is the individual rationality (IR) constraint, which requires that the consumer has the option of not buying anything from the provider. The outside option is normalized to 0. The third inequality is the minimum need (MN) constraint, which requires

[6]Perrigne and Vuong (2011a) consider instead the cost for the total amount produced across consumers. Luo (2011) shows that these two models are nonnested except if one assumes a constant marginal cost.

that the consumer can use an internet level above his minimum need for internet.

Hereafter, the subscript q (θ) denotes the partial derivative with respect to q (θ) respectively. I make the following assumptions on the model structure.

Assumption 2. *For all* $(\theta, \beta) \in \Theta$, $q \in I\!R^+$, *and* $j \in \mathcal{J}$, $U(\cdot, \cdot; \cdot)$, $c(\cdot, \cdot)$ *and* $F(\cdot, \cdot)$ *satisfy*

(i) $U(0, 0; \theta) = 0$, $U_q(q, j; \theta) \geq 0$, $U_{qq}(q, j; \theta) \leq 0$, $U_\theta(q, j; \theta) > 0$, $U_{\theta\theta}(q, j; \theta) \leq 0$,

(ii) $U_{q\theta}(q, j; \theta) > 0$,

(iii) $\frac{\partial}{\partial\theta} \frac{-U_{qq}(q,j;\theta)}{U_q(q,j;\theta)} \leq 0$,

(iv) $\frac{c_{qq}(q,j)}{c_q(q,j)} > \frac{U_{qq}(q,j;\theta)}{U_q(q,j;\theta)}$,

(v) $H(\theta|j) \equiv \theta - \frac{1-F(\theta|D(\beta)=j)}{f(\theta|D(\beta)=j)}$ *is increasing in* θ, *where* $D(\beta) \equiv \min\{j \in \mathcal{J} : j \geq \beta\}$,

(vi) $U(0, j; \theta) = v(0, j) \geq c(0, j)$, *for some function* $v(\cdot, \cdot) : \Theta \to I\!R^+$.

All these assumptions with the exception of (vi) are standard in the nonlinear pricing literature. See e.g. Maskin and Riley (1984). Assumption 2-(i) says that the outside option (not buying) provides a zero utility and the marginal utility from phone service is nonnegative and decreasing. Moreover, a consumer with a higher taste θ gets a larger utility and this increase is diminishing as θ increases. Assumption 2-(ii) is the standard Spence-Mirrlees single-crossing condition, which says that a consumer with a higher taste θ enjoys a larger marginal payoff for phone usage across every (q, j). Assumption 2-(iii) implies nonincreasing absolute risk aversion, while 2-(iv) requires that the cost function is not too concave in q. The latter is satisfied by any linear or convex cost function. Assumption 2-(v) says that the conditional hazard rate does not decline too rapidly as θ increases, where the term $D(\beta)$ represents a (θ, β) consumer's minimum acceptable internet speed. Most commonly used unimodal distributions satisfy the hazard rate assumption. Assumption 2-(vi) says that the willingness-to-pay for phone service θ does not matter to the consumer unless his phone usage is positive. Thus internet-only users have no parameter θ in their utility function. I call this assumption "weak complementarity". Consequently, the firm can charge a fixed fee $v(0, j)$ and leave no rent to internet-only users. Finally, Assumption 2-(vi) also implies that serving internet as a separate product is profitable for the firm.

3.3.2 Characterization of the Optimal Selling Mechanism

The approach I adopt is close to the separability case discussed in Rochet and Stole (2003). Specifically, I partition the set of types into one-dimensional subsets and reduce the multi-dimensional problem to a unidimensional problem. To clarify ideas, I first study the case when the provider observes β. Thus the problem becomes unidimensional and I explicitly characterize the optimal selling mechanism. I then show that this mechanism is still optimal under a standard affiliation assumption when β is not observed, thereby solving the multidimensional screening problem.

Solving the Model when β is Observed

When β is observed, the provider solves the following problem for each value of β

$$\begin{aligned}
\max_{t(\cdot,\beta),q(\cdot,\beta),j(\cdot,\beta)} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \Big[t(\theta,\beta) - c(q(\theta,\beta), j(\theta,\beta))\Big] f(\theta|\beta) d\theta \\
s.t. \quad & U(q(\theta,\beta), j(\theta,\beta);\theta) - t(\theta,\beta) \geq U(q(\tilde{\theta},\beta), j(\tilde{\theta},\beta);\theta) - t(\tilde{\theta},\beta), \\
& U(q(\theta,\beta), j(\theta,\beta);\theta) - t(\theta,\beta) \geq 0, \\
& j(\theta,\beta) \geq \beta,
\end{aligned}$$

for all (θ,β) and $(\tilde{\theta},\beta)$ such that $j(\tilde{\theta},\beta) \geq \beta$. Since the consumer cannot misreport β, the two-dimensional IC and IR constraints reduce to one-dimensional constraints. Moreover, the information structure reduces to the conditional density $f(\theta|\beta)$. The problem then becomes a multiproduct nonlinear pricing problem in which a consumer's private information is one-dimensional. [7] I denote the optimal selling mechanism as $(t^*(\cdot,\cdot), q^*(\cdot,\cdot), j^*(\cdot,\cdot))$.

I make the following assumption on the utility and the cost functions.

Assumption 3. *For all $\theta \in [\underline{\theta}, \overline{\theta}]$, $q \in I\!R^+$ and $j \in \mathcal{J}$, $U(\cdot,\cdot;\cdot)$ and $c(\cdot,\cdot)$ satisfy*

(i) The utility function is additively separable as follows

$$U(q, j; \theta) = u(q;\theta) + v(q, j),$$

[7] The model can be extended to entertain both q and θ multidimensional relying on Armstrong (1996). The basic idea would be to design a cost-based tariff. To do so, I would need to define the cost-based indirect utility function $V(c, j;\theta) \equiv \max_{c(q_1,q_2,j)\leq c} U(q_1, q_2, j; \theta_1, \theta_2)$, and $V(c, j;\theta) = h(\theta_1,\theta_2)u(c, j) + v(c, j)$. This model is left for future research. See also Luo, Perrigne, and Vuong (2012).

where $u(\cdot;\cdot)$ *satisfies* $u(0;\theta) = 0$.[8]

(ii) For all $\tilde{j} > j$*, where* $\tilde{j} \in \mathcal{J}$

$$U_q(q, \tilde{j}; \theta) - U_q(q, j; \theta) \leq c_q(q, \tilde{j}) - c_q(q, j),$$
$$v(0, \tilde{j}) - v(0, j) \leq c(0, \tilde{j}) - c(0, j).$$

Assumption 3-(i) borrows from Sundararajan (2003) and Chen and Luo (2012) in the context of nonlinear pricing with network effects where j is replaced by the total quantity of product used by all consumers in the market $Q = \int_{\underline{\theta}}^{\overline{\theta}} q(\theta) f(\theta) d\theta$. I remark that the cross derivative $U_{qj}(\cdot, \cdot; \cdot)$ becomes independent of θ. Although the interaction between internet and phone services is the same for consumers with different tastes, I allow this interaction to vary with the bundle. Hereafter, I call $u(\cdot;\cdot)$ the intrinsic utility function for phone service, and $v(\cdot,\cdot)$ the complementary utility function for the bundle.[9]

Assumption 3-(ii) says that the increment of marginal cost for phone service is larger than the marginal utility when one increases the internet level. Similarly, the cost increment for serving internet only is larger than the corresponding incremental utility. It implies $U(q, \tilde{j}; \theta) - c(q, \tilde{j}) \leq U(q, j; \theta) - c(q, j)$. Hence, when the provider observes β, he always prefers to assign the minimum internet speed the consumer can accept.[10]

I can now show that the problem reduces to several standard single product nonlinear pricing problems. I need first to show that the provider will always assign the minimum internet speed $D(\beta)$ that the consumer can accept when β is observed. This gives the following lemma.

Lemma 1. *Under Assumptions 1 and 3,* $j^*(\theta, \beta) = D(\beta)$*,* $q^*(\theta, \beta) = q^*\big(\theta, D(\beta)\big)$*, and*

[8]The utility $U(q, j; \theta) = u(q; \theta) + v(q, j) + \omega(\theta, j)$ is more general but does not satisfy the weak complementarity assumption. See Assumption 2-(vi).

[9]Liu, Chintagunta, and Zhu (2010) find evidence of strong complementarity between local phone consumption and DSL internet. In view of their results, I call $v(\cdot,\cdot)$ the complementary utility function. However, I do not impose any restriction on the cross derivative of $v(\cdot,\cdot)$.

[10]Assumption 3-(ii) can be weakened. I can assume instead that the two terms $\big[U_q(q, j; \theta) - c_q(q, j)\big] - \big[U_q(q, \tilde{j}; \theta) - c_q(q, \tilde{j})\big]$ and $\big[v(0, j) - c(0, j)\big] - \big[v(0, \tilde{j}) - c(0, \tilde{j})\big]$ have the same sign for all $\theta \in [\underline{\theta}, \overline{\theta}]$, $q \in \mathbb{R}^+$, j belonging to a larger choice set $\mathcal{J}^0$. Under complete information on β, the provider solves for each β $D^0(\beta) \equiv \arg\max_{j \geq \beta, j \in \mathcal{J}^0} \big\{v(0, j) - c(0, j)\big\}$. This weaker assumption ensures that the assignment does not depend on θ. With $\mathcal{J} \equiv \{j \in \mathcal{J}^0 : j = D^0(\beta)$ for some $\beta \in [\underline{\beta}, \overline{\beta}]\}$, I can show that Assumption 3-(ii) is satisfied, thereby endogenizing the choice of internet options.

$t^*(\theta,\beta) = t^*(\theta, D(\beta))$ *for all* $(\theta,\beta) \in \Theta$.

Lemma 1 says that β only affects the optimal selling mechanism through the step function $D(\beta)$.[11] Thus consumers having the same minimum acceptable internet speed $D(\beta)$ face the same phone service assignment $q^*(\cdot,\cdot)$, the same internet assignment $j^*(\cdot,\cdot)$ and price schedule $t^*(\cdot,\cdot)$. The following lemma characterizes the optimal mechanism.

Lemma 2. *Under Assumptions 1, 2 and 3, for any given value of* β*, the optimal phone service assignment* $q^*(\cdot,\beta)$ *and price schedule* $t^*(\cdot,\beta)$ *satisfy:*

(i) There exists a cutoff taste $\theta^c_\beta \in [\underline{\theta}, \overline{\theta}]$*, above which consumers are assigned a bundle with internet* $D(\beta)$ *and phone service, and below which consumers are assigned internet* $D(\beta)$ *only. The cutoff taste* θ^c_β *is defined as*

$$\theta^c_\beta \equiv \min\left\{\theta \in [\underline{\theta}, \overline{\theta}] : M(\theta, D(\beta)) \geq 0\right\}, \tag{1}$$

and

$$M(\theta,j) \equiv \Big[U(q^*(\theta,j),j;\theta) - v(0,j)\Big] - \Big[c(q^*(\theta,j),j) - c(0,j)\Big] - U_\theta(q^*(\theta,j),j;\theta)\frac{1-F(\theta|j)}{f(\theta|j)}.$$

(ii) If $\theta \in [\theta^c_\beta, \overline{\theta}]$*,* $q^*(\cdot,\beta)$ *and* $t^*(\cdot,\beta)$ *are solution of*

$$U_q[q^*(\theta,\beta), D(\beta);\theta] = c_q[q^*(\theta,\beta), D(\beta)] + U_{q\theta}[q^*(\theta,\beta), D(\beta);\theta]\frac{1-F[\theta|D(\beta)]}{f[\theta|D(\beta)]}, \tag{2}$$

$$t^*(\theta,\beta) = U[q^*(\theta,\beta), D(\beta);\theta] - \int_{\theta^c_\beta}^{\theta} U_\theta[q^*(x,\beta), D(\beta);x]\,dx. \tag{3}$$

(iii) If $\theta \in [\underline{\theta}, \theta^c_\beta)$*,* $q^*(\theta,\beta) = 0$ *and* $t^*(\theta,\beta) = v(0, D(\beta))$*.*

The proof is in two steps. In a first step, I derive the optimal selling mechanism conditionally on serving consumers with a willingness-to-pay equal or above θ^c, which is an arbitrary cutoff value. Thus the optimal selling mechanism is defined by (2) and (3) by replacing θ^c_β with θ^c. An important feature is that the phone service assignment

[11] By considering non-random nonlinear pricing schedules of the form $T(x,j)$, consumers' report of β is constrained to belong to $\mathcal{J}$. The problem of discrimination on any value of β is left for future research.

$q(\cdot,\cdot)$ does not depend on θ^c while the price schedule $t(\cdot,\cdot)$ is decreasing in θ^c. In a second step, I find the optimal cutoff value θ^c that maximizes the profit. It is defined by (1). Intuitively, when the firm slightly lowers θ^c, the term $[U(q^*(\theta^c, j), j; \theta^c) - v(0, j)]$ is the incremental utility for a (θ^c, j) consumer switching from internet only to a bundle, $[c(q^*(\theta^c, j), j) - c(0, j)]$ is the incremental cost for the firm, and $U_\theta(q^*(\theta^c, j), j; \theta^c)$ is the additional informational rent everyone in the customer base gets. Therefore, the term $\{[U(q^*(\theta^c, j), j; \theta^c) - v(0, j)] - [c(q^*(\theta^c, j), j) - c(0, j)]\} f(\theta^c|j)$ is the marginal gain for expanding the customer base by lowering θ^c, while $U_\theta(q^*(\theta^c, j), j; \theta^c)[1 - F(\theta^c|j)]$ is the corresponding marginal loss for reducing the tariff to every consumer above θ^c. Equation (1) balances these two effects. In addition, it is easy to see that $\theta^c_\beta = \theta^c_j$, where $j = D(\beta)$. The term β only affects the cutoff taste through the step function $D(\beta)$. My results are in the spirits of Armstrong and Rochet (1999) recommendation where they advise to discretize the type space to simplify the multidimensional screening problem.

Equations (2) and (3) define the phone service assignment $q^*(\cdot,\cdot)$ and price schedule $t^*(\cdot,\cdot)$ for the bundle users. In particular, (2) says that the marginal payoff for each type equals the marginal cost plus a nonnegative distortion term due to incomplete information. Intuitively, $\{U_q[q^*(\theta, \beta), D(\beta); \theta] - c_q[q^*(\theta, \beta), D(\beta)]\} f(\theta|D(\beta))$ represents the firm's desire to implement an efficient allocation weighted by the density while $U_{q\theta}(q^*(\theta, \beta), D(\beta); \theta)[1 - F(\theta|D(\beta))]$ represents the informational rent the firm has to give up to the consumer for revealing their private information. Equation (3) says that the payment equals the consumer's utility minus some informational rent. Following Assumption 2, the resulting usage-based tariffs $T^*(\cdot, j)$ are increasing and concave for all j.

If a consumer is excluded from consuming phone service, then his utility function becomes $U(q, j; \theta) = v(0, j)$ by Assumption 2-(vi). Since there is no asymmetric information, the firm can take all the consumer surplus by charging $T^*(0, j) = v(0, j)$, leading to $t^*(\theta, \beta) = v(0, D(\beta))$, for any $(\theta, \beta) \in \Theta$ such that $\theta \in [\underline{\theta}, \theta^c_\beta)$.

Solving the Model when both θ and β are Unobserved

I now show that the previous mechanism is still optimal when both θ and β are not observed by the firm. To do so, I would need to make an additional assumption. This constitutes an interesting result given the complexity of multidimensional screening problems.

I first need to establish the desirability of the minimum acceptable internet speed. Let $(t^{sb}(\cdot,\cdot), q^{sb}(\cdot,\cdot), j^{sb}(\cdot,\cdot))$ be the optimal selling mechanism when both θ and β are private information. The following lemma shows that β affects this mechanism only through $D(\beta)$.

Lemma 3. *Under Assumptions 1 and 3, $j^{sb}(\theta,\beta) = D(\beta)$, $q^{sb}(\theta,\beta) = q^{sb}(\theta, D(\beta))$, and $t^{sb}(\theta,\beta) = t^{sb}(\theta, D(\beta))$ for all $(\theta,\beta) \in \Theta$.*

This result parallels Lemma 1. It remains to show that this second-best mechanism is the same as the one with asymmetric information on θ only. The basic idea is to reduce the number of binding IC constraints. Since the consumer can either overreport, underreport or report truthfully each parameter of his private information. The number of potential deviations increases to eight in a two-dimensional problem from two in a unidimensional problem. However, I show that only three deviations matter as stated in the next lemma.

Lemma 4. *Following Assumption 1, when both θ and β are private information, the IC constraints are satisfied if and only if the following two one-dimensional IC constraints are satisfied. Namely,*

$$U(q(\theta,\beta), D(\beta);\theta) - t(\theta,\beta) \geq U(q(\tilde{\theta},\beta), D(\beta);\theta) - t(\tilde{\theta},\beta) \quad \forall \theta, \beta, \tilde{\theta},$$

$$U(q(\theta,\beta), D(\beta);\theta) - t(\theta,\beta) \geq U(q(\theta,\tilde{\beta}), D(\tilde{\beta});\theta) - t(\theta,\tilde{\beta}) \quad \forall \theta, \beta, \tilde{\beta} \text{ such that } D(\tilde{\beta}) \geq \beta.$$

Following Assumption 1, the (θ,β) and $(\theta,\tilde{\beta})$ consumers have the same preferences over outcomes as long as their minimum needs for internet are satisfied. If the two one-dimensional IC constraints above are true, then the (θ,β) consumer does not want to pretend to be $(\theta,\tilde{\beta})$, and the $(\theta,\tilde{\beta})$ consumer does not want to pretend to be $(\tilde{\theta},\tilde{\beta})$. Therefore, by transitivity, the (θ,β) consumer has no incentive to pretend to be $(\tilde{\theta},\tilde{\beta})$. Thus consumers report truthfully. Intuitively, if one considers θ on the x-axis and β on the y-axis, the potential deviations can be horizontal for θ and vertical for β. Lemma 4 tells us that the only binding constraints are only upward for β and upward and downward for θ, while the other deviations are redundant, thereby reducing the number of binding constraints to three.

Before showing that the second-best mechanism is the one given in Lemma 2, I make an affiliation assumption on the joint distribution of θ and β.

Assumption 4. *$\forall \theta \in [\underline{\theta}, \overline{\theta}]$ and $\forall j \in \mathcal{J}$, $\frac{1-F(\theta|D(\beta)=j)}{f(\theta|D(\beta)=j)}$ is increasing in j.*

Assumption 4 follows Che and Gale (2000). It is equivalent to assume $H(\theta|j)$ be decreasing in j. Intuitively, a consumer is relatively more likely to have a higher willingness-to-pay for phone service if he needs a higher speed internet. According to Horrigan (2010), the Federal Communications Commission 2009 survey shows that higher speed internet adopters tend to be better educated with higher incomes. Since phone service is a normal good, I consider it as a reasonable assumption.

I can now show that the consumer has no incentive to misreport β given a willingness-to-pay θ. This is the purpose of the next lemma.

Lemma 5. *Under Assumptions 1, 2, 3 and 4, $q^*(\theta, \beta)$ is decreasing in β and θ^c_β is increasing in β. Therefore, a (θ, β) consumer's indirect utility is decreasing in β. Moreover, $T^*(q,j) - v(q,j)$ is increasing in j.*

The intuition is as follows. Following Assumption 4, the firm knows that a consumer is more likely to have a higher taste for phone service when he needs a higher speed internet. By exploiting this positive dependence, it can charge the consumer more for the same amount of q when the consumer chooses a higher j. This in turn implies that the subscriber would consume less phone service giving a decreasing assignment $q^*(\cdot,\cdot)$ in β. While the optimal exclusion is the result of a trade-off between the marginal gain and the loss of expanding the customer base, Assumption 4 favors the latter as one moves from a low speed internet to a high speed one. Therefore, it becomes more profitable to exclude a larger range of low taste consumers if they adopt a higher speed internet. This explains why the cutoff value θ^c_β is increasing in β. Under Assumption 3, a (θ, β) consumer solves

$$\max_{(q,j): j \geq \beta} \quad u(q;\theta) + v(q,j) - T^*(q,j).$$

Following Lemma 5, because $T^*(q,j) - v(q,j)$ is increasing in j, the additional payment for a higher internet level is larger than the additional utility it brings. Therefore, the consumer will choose the minimum internet speed meeting his needs.

We are now in a position to show that $(t^*(\cdot,\cdot), q^*(\cdot,\cdot), j^*(\cdot,\cdot))$ is the optimal mechanism when both θ and β are private information. When β is observed, the firm's profit is weakly higher than the one when β is not observed. Thus, I only need to show that the two-dimensional IC and IR constraints still hold if the provider uses $(t^*(\cdot,\cdot), q^*(\cdot,\cdot), j^*(\cdot,\cdot))$. Regarding the IR constraints, they are satisfied automatically because they are the same. Following Lemma 4, the IC constraints are satisfied as long as the two one-dimensional IC constraints are. On one hand, misreporting θ is not profitable because of $(t^*(\cdot,\cdot), q^*(\cdot,\cdot), j^*(\cdot,\cdot))$. On the other hand, misreporting β is not profitable either following Lemma 5. The following proposition summarizes these results.

Proposition 1. *Under Assumptions 1, 2, 3 and 4, we have* $t^*(\cdot,\cdot) = t^{sb}(\cdot,\cdot)$, $q^*(\cdot,\cdot) = q^{sb}(\cdot,\cdot)$ *and* $t^*(\cdot,\cdot) = j^{sb}(\cdot,\cdot)$.

In view of Lemma 2 and Proposition 1, consumers are segmented into several groups based on their internet needs and tastes for phone service. All consumers with β such that $D(\beta) = j$ uses the same internet level, j. I refer to them as group j. Consumers in group j are further divided into 2 parts according to their values of θ, namely, $[\underline{\theta}, \theta_j^c)$ and $[\theta_j^c, \overline{\theta}]$. The former or low taste subscribers will consume internet service j only, while the latter or high taste subscribers consume the bundle with $q > 0$ and internet service j. I call these two groups internet j users and bundle j users, respectively. In addition, the firm proposes J usage-based tariffs to bundle users ($q > 0$ and $j > 0$) and phone-only users ($q > 0$ and $j = 0$). It proposes $J - 1$ fixed fees to internet-only users ($q = 0$ and $j > 0$). These correspond to the data I will analyze in Section 5.

Armstrong and Rochet (1999) remark that phenomena such as bunching and exclusion that arise in multiproduct nonlinear pricing models create technical difficulties, making it hard to generate closed-form solutions. Following Proposition 1, the optimal selling mechanism reduces to a combination of optimal selling mechanisms for a series of one-dimensional problems. Therefore, I characterize explicitly the optimal exclusion, the assignment schedules and the tariff functions. In my model, bunching arises at equilibrium through $D(\beta)$ and because people with low taste for phone service will consume internet only.

Bundling Decisions

The results on optimal exclusion have important implications on bundling choices. For instance, the firm will sell internet j only if $\theta_j^c = \overline{\theta}$. Similarly, he will sell internet only in a bundle if $\theta_j^c = \underline{\theta}$. For any other value of $\theta_j^c \in (\underline{\theta}, \overline{\theta})$, the firm will propose both. Thus, my model is general as it allows the three possible incentives to bundle, namely utility complementarity, cost efficiency and correlation between θ and β. To some extent, my model confirms Schmalensee (1984) and Fang and Norman (2006) results, i.e. the higher the cost or the lower the valuation, the less likely that bundling dominates component pricing. Moreover, my model admits a variety of bundling outcomes, including component pricing, pure bundling, semi-mixed bundling and mixed bundling.

When it is too costly to provide phone service to internet users, i.e. $\underline{\beta} \leq 0$ and $\theta_j^c = \overline{\theta}$ for all $j > 0$, the firm can exclude all internet users from consuming phone service. In this case, the firm would sell internet and phone services separately, which is known as component pricing (CP). When it is optimal for the firm to exclude some but not all internet users from consuming phone service, the provider wound sell internet and phone services not only as separate products but also in a bundle. If all the possible combinations of internet and phone services are offered, i.e. $\underline{\beta} \leq 0$ and $\theta_j^c \in (\underline{\theta}, \overline{\theta})$ for all $j \in \mathcal{J}$, the firm implements mixed bundling (MB). If some combination is not offered, the firm implements semi-mixed bundling (SMB). Moreover, if we let $\underline{\beta} > 0$, pure bundling (PB) arises when it is optimal for the firm to serve phone service to everyone, i.e. $\underline{\beta} > 0$ and $\theta_j^c = \underline{\theta}$ for all $j > 0$.

Hereafter, I provide a numerical example to illustrate Proposition 1 and the various bundling possibilities.

Numerical Example

Let the quadratic utility function

$$U(q, j; \theta) = \begin{cases} (w + \theta)q - \frac{1}{2}q^2 + \nu(j) & \text{if } q \leq w + \theta, \\ \frac{(w+\theta)^2}{2} + \nu(j) & \text{if } q > w + \theta, \end{cases}$$

where w represents the average household income and determines the consumer's minimum willingness-to-pay for phone service. The cost function is $c(q, j) = \kappa_0 \mathbb{1}(q > 0) + \kappa_j \mathbb{1}(j >$

$0) + \gamma q$, where κ_0 is a fixed cost associated to any positive phone usage, κ_j is a fixed cost for internet, and γ is the marginal variable cost of phone service. In addition, consumers' tastes for phone service are distributed as $F(\theta|j) = 1 - (1-\theta)^{b_j}$, with support $\theta \in [0,1]$ and $b_j \geq 1$. Suppose b_j is decreasing in j.

After elementary algebra, Lemma 2 leads to the following optimal phone service assignment

$$q^*(\theta, j) = \begin{cases} (w - \gamma - \frac{1}{b_j}) + \theta(1 + \frac{1}{b_j}), & \text{if } \theta \in [\theta_j^c, 1] \\ 0, & \text{if } \theta \in [0, \theta_j^c) \end{cases}$$

where

$$\theta_j^c = \frac{\sqrt{2\kappa_0} + \gamma - w + \frac{1}{b_j}}{1 + \frac{1}{b_j}}.$$

Figure C.1 displays $H(\cdot,\cdot)$, $q^*(\cdot,\cdot)$ and $T^*(\cdot,\cdot) - v(\cdot,\cdot)$ for the values $\mathscr{J} = \{0,1,2\}, w = 0.25, \kappa_0 = 0.1, \gamma = 0.05, b_0 = 2, b_1 = 1.5$ and $b_2 = 1$. Note that $H(\theta|j) = \theta - [1 - F(\theta|j)]/f(\theta|j) = [\theta(1+b_j) - 1]/b_j$ is decreasing in j and increasing in θ. I remark that these functional forms satisfy Assumptions 3 and 4. To see why this mechanism is still optimal when both θ and β are private information, I consider a consumer who overreports and switches to a higher speed internet. Figure C.1-(c) shows that the additional payment is larger than the marginal utility he can get. Therefore, no consumer would like to overreport his internet need.

The firm's optimal bundling strategy depends on $\Lambda \equiv \sqrt{2\kappa_0} + \gamma - w$. In particular, the firm will provide group j consumers with internet only if $\Lambda \geq 1$, the bundle of internet and phone services if $\Lambda \leq -1/b_j$, and both otherwise. When $j > 0$, some consumers will switch from bundles to internet only if either the cost parameters κ_0 or γ increase or their income level w decreases. Intuitively, the firm bundles internet and phone services when the latter is cheap to produce and unbundles them otherwise. Figure C.2 displays how the population of subscribers is segmented under different bundling scenarios. The gray color is used for consumers who are excluded from consumption, while the green color is used for phone-only users. The two red colors are used for internet-only users: the light one for $j = 1$ and the dark one for $j = 2$. The two blue colors are used for bundle users: the light one for $j = 1$ and the dark one for $j = 2$. The bold lines give the cutoff values of θ. In Figure C.2-(a), the cutoff values are at $\overline{\theta}$ because $\Lambda > 1$. The firm will not propose either bundling or phone service because they are too costly to produce. Despite the fact that there is no phone service, this case corresponds to component pricing. In Figure C.2-(e), the cutoff values are at $\underline{\theta}$ because $\Lambda < -1/b_2$. The firm proposes bundles and phone-only because either phone calls are very cheap to produce or consumers value them very highly. This is the case of semi-mixed bundling without exclusion. Figures C.2-(b) and C.2-(c) display mixed bundling with exclusion where the cutoff value increases with j. While in C.2-(b) there will be some consumers excluded from buying anything, the firm provides services to every consumer in C.2-(c). In Figure C.2-(d), only consumers with a high need for internet may not buy a bundle leading to semi-mixed bundling with exclusion. Lastly, in Figure C.2-(f) with $\underline{\beta} > 0$, the firm will propose bundling to everyone.

3.4 Identification and Estimation

In view of Section 3.3, the optimal mechanism is defined by (1), (2) and (3). Because the data display mixed bundling, I focus on this case. However, the results below can be readily adapted to the other cases of bundling. It is useful to recall the model structure and the observables. The model primitives are $\{u(\cdot;\cdot), v(\cdot,\cdot), F(\cdot,\cdot), c(\cdot,\cdot)\}$, which are the intrinsic utility function from consuming phone service, the complementary utility function from consuming internet and phone services, the joint distribution of consumers' types and the firm's cost function. Because j can take values in $\{0, 1, 2\}$, the model primitives become $\{u(\cdot;\cdot), v_j(\cdot), F_j(\cdot), c_j(\cdot)\}$, where $v_j(\cdot) \equiv v(\cdot, j)$, $F_j(\cdot) \equiv F(\cdot|D(\beta) = j)$ and $c_j(\cdot) \equiv c(\cdot, j)$ for $j = 0, 1, 2$. Regarding the observables, following Section 2, I observe the tariff $T_j(\cdot)$ for $j = 0, 1, 2$. Moreover, data on internet-only users provide information on $T_j(0)$ for $j = 1, 2$. Since I observe q, we have the distribution of consumption $G^*_j(\cdot)$ for $j = 0, 1, 2$ and $q > 0$ as well as $G_j(0)$ from internet-only users. To summarize, the observables are $\{T_j(\cdot), G^*_j(\cdot)\}$ for $q > 0$ and $j = 0, 1, 2$ and $\{T_j(0), G_j(0)\}$ for $j = 1, 2$.

3.4.1 Identification

From the auction literature, the one-to-one mapping at the equilibrium plays a crucial role to identify the model primitives. See Guerre, Perrigne, and Vuong (2000) and Athey and Haile (2007). In nonlinear pricing models, Perrigne and Vuong (2011a) show that the optimality of tariff in addition to the one-to-one mapping between the observed quantity and the unobserved taste are both needed to identify the model primitives. See also Luo, Perrigne, and Vuong (2012). The multidimensional screening problem adds additional difficulties. In the context of insurance, Aryal, Perrigne, and Vuong (2009) exploit a repeated outcome, i.e. the number of accidents, to identify the model structure.

Here, the idea is to exploit the variation offered by the data across the different groups of users, including those using either internet or phone only, in addition to the first-order conditions (1), (2) and (3). I proceed in several steps. First, I study which primitives the data on internet-only users and phone-only users will allow me to identify. While assuming multiplicative separability of the intrinsic utility function and linearity of the cost function, I will show that the intrinsic utility function, the marginal cost parameter as well as the conditional density of θ for $j = 0$ are identified. Second, I investigate the identification of the complementary utility function and the conditional densities of θ for $j = 1, 2$ from the bundle users data. Third, I show how to exploit the firm's optimal exclusion conditions to identify the fixed cost parameters. Therefore, the optimality of tariffs and bundling as well as one-to-one mapping at the equilibrium between the consumption of phone service and the unknown type θ will play crucial roles.

Identifying Assumptions

I make the following identifying assumptions on the model primitives. Hereafter, the prime denotes a derivative with respect to q.

Assumption 5. *For all* $\theta \in [\underline{\theta}, \overline{\theta}]$ *and* $q \in \mathbb{R}^+$,

(i) The intrinsic utility $u(q;\theta)$ *satisfies*

$$u(q;\theta) = \theta u_0(q),$$

with $u_0(0) = 0$, $u_0'(q) \geq 0$ *and* $u_0''(q) \leq 0$.

(ii) The cost function is of the form

$$c_j(q) = \kappa_0 \mathbb{1}(q > 0) + \kappa_j \mathbb{1}(j > 0) - \Delta_j \mathbb{1}(qj > 0) + \gamma q,$$

where $\gamma > 0$, $\kappa_0 > 0$, $\kappa_j > 0$, *and* $\Delta_j \in \big[0, \min\{\kappa_0, \kappa_j\}\big)$ *for* $j = 1, 2$.

(iii) $v_0(q) = 0$.

Following Perrigne and Vuong (2011a) and Perrigne and Vuong (2011b), I assume multiplicative separability of the intrinsic utility function in the type θ as stated in Assumption 5-(i). Thus, I interpret $u_0(\cdot)$ as the base intrinsic utility function. However, Assumption 5-(i) will not be sufficient to achieve identification. I provide an intuitive argument. Equations (2) and (3) provide $2J$ one-to-one mappings between θ and q and between t and q, respectively. On the other hand, I have to identify J cost functions $c_j(\cdot)$, J complementary utility functions $v_j(\cdot)$, and J conditional type distributions $F_j(\cdot)$ as well as the base intrinsic utility function $u_0(\cdot)$. It is clear that additional restrictions need to be imposed.

Several identifying assumptions can be entertained. Following Section 2, the production of telecommunication services tends to involve high fixed costs and small marginal costs. Therefore, I assume a linear cost function as stated in Assumption 5-(ii). The term κ_0 is a fixed cost associated to any positive production of phone service, while κ_j is a fixed cost to provide internet level j. The term Δ_j is the difference between the sum of these two fixed costs and the fixed cost to sell a bundle. It measures the cost saving effect of selling internet and phone services as a bundle. I also assume that the fixed cost of serving a bundle is higher than the separate fixed costs, i.e. $\kappa_0 + \kappa_j - \Delta_j \geq \max\{\kappa_0, \kappa_j\}$. Thus $\Delta_j \in [0, \min\{\kappa_0, \kappa_j\})$.

Lastly, Assumption 5-(iii) says that there is no complementary utility when there is no consumption of internet. It implies that $v_0'(q) = 0$ for all $q \in I\!R^+$.[12] I can then rewrite the

[12] An alternative normalization would be to assume a similar value for another internet level instead.

first-order conditions (2) and (3) defining the optimal selling mechanism. Namely,

$$\theta u_0'(q) + v_j'(q) = \gamma + u_0'(q)\frac{1-F_j(\theta)}{f_j(\theta)}, \qquad \forall q \in [\underline{q}_j, \overline{q}_j] \tag{4}$$

$$T_j'(q) = \theta u_0'(q) + v_j'(q), \qquad \forall q \in [\underline{q}_j, \overline{q}_j] \tag{5}$$

where $\underline{q}_j = q^*(\theta_j^c, j)$ and $\overline{q}_j = q^*(\overline{\theta}, j)$ for $j = 0, 1, 2$. Together with the boundary conditions $T_j(\underline{q}_j) = \theta_j^c u_0(\underline{q}_j) + v_j(\underline{q}_j)$ and the cutoff tastes in (1), (4) and (5) define the optimal mechanism.

Identification of γ, $v_j(0)$, $F_j(\theta_j^c)$, $F_0(\cdot)$ and $u_0(\cdot)$

Under Assumption 5-(ii), only the marginal variable cost enters in (4). Thus, γ is identified from (4) and (5) evaluated at the maximum phone usage. This gives $\gamma = T_j'(\overline{q}_j)$, $\forall j \in \{0, 1, 2\}$. The identification of the fixed cost parameters κ_0, κ_j and Δ_j will be shown later.

From the model of Section 3, internet-only users pay $T_j(0) = v_j(0)$ for any $j = 1, 2$, which renders the identification of $v_j(0)$ immediate. Moreover, the proportion of individuals using internet only among their group of users gives $F(\theta_j^c|j)$ for $j = 1, 2$. These results are summarized in the following proposition while the identification of θ_j^c will be addressed later.

Proposition 2. *The cost parameter γ is identified. $v_j(0)$ and $F_j(\theta_j^c)$ are identified for $j = 1, 2$.*

I now turn to data from phone-only users. The following lemma exploits the one-to-one mapping between the observed q and the unobserved type θ as well as the price schedule $T_0(\cdot)$. See Perrigne and Vuong (2011a) for a detailed proof.

Lemma 6. *Under Assumptions 1-5, the first-order conditions* (4) *and* (5) *for $j = 0$ are equivalent to*

$$u_0'(q) = \frac{T_0'(q)\xi(q)}{\overline{\theta}}, \tag{6}$$

$$\theta_0(q) = \frac{\overline{\theta}}{\xi(q)}, \tag{7}$$

for all $q \in [\underline{q}_0, \overline{q}_0]$, *where*

$$\xi(q) = \Big[1 - G_0^*(q)\Big]^{1-\frac{\gamma}{T_0'(q)}} \exp\Big\{\gamma \int_q^{\overline{q}_0} \frac{T_0''(x)}{T_0'(x)^2} \log\left[1 - G_0^*(x)\right] dx\Big\}, \tag{8}$$

with $\gamma = T_0'(\overline{q}_0)$.

Lemma 6 shows that the marginal intrinsic utility $u_0'(\cdot)$ and the unobserved taste for phone service θ for phone-only users are identified up to a constant. In view of Lemma 6, a natural normalization is $\overline{\theta}$=1.

Assumption 6. $\overline{\theta} = 1$.

Under such a normalization, $u_0(\cdot)$ can be interpreted as the intrinsic utility function for the highest taste. Since $\theta_0^c = \theta_0(\underline{q}_0)$, θ_0^c is identified. Moreover, I can further identify $u_0(\cdot)$ using the boundary condition $T_0(\underline{q}_0) = \theta_0^c u_0(\underline{q}_0)$.[13] Namely,

$$u_0(q) = \frac{T_0(\underline{q}_0)}{\theta_0^c} + \int_{\underline{q}_0}^{q} u_0'(x) dx, \tag{9}$$

for all $q \in [\underline{q}_0, \overline{q}_0]$. The next proposition summarizes these results.

Proposition 3. *Under Assumptions 1-6, the base intrinsic utility function* $u_0(\cdot)$ *and the type* $\theta_0(\cdot)$ *are identified on* $[\underline{q}_0, \overline{q}_0]$. *Moreover, the truncated conditional taste distribution* $F_0^*(\cdot)$ *is identified on* $[\theta_0^c, \overline{\theta}]$, *while the conditional taste distribution* $F_0(\cdot)$ *is identified up to a constant on* $[\theta_0^c, \overline{\theta}]$.

The distribution $F_0(\cdot)$ is identified up to a constant because I do not observe the proportion of consumers who do not buy internet and phone services. Moreover, usage and payment data do not provide any variation to identify $u(\cdot)$ and $F_0(\cdot)$ on $[0, \underline{q}_0)$ and $[\underline{\theta}, \theta_0^c)$, respectively.

Identification of $v_j(\cdot)$ and $F_j(\cdot)$

[13]Perrigne and Vuong (2011a) identify the marginal utility function only. The reason is that the minimum consumption in their model (standard listing) is offered for free in yellow pages. Therefore, they could not relate the minimum consumption to a utility level.

I now turn to the second step, which addresses the identification of $v_j(\cdot)$ and $F_j(\cdot)$ for $j = 1, 2$. I exploit the one-to-one mapping between phone usage q and taste θ for bundle users, which implies for each $q \in [\underline{q}_j, \overline{q}_j]$, $G_j^*(q) = [F_j(\theta) - F_j(\theta_j^c)]/[1 - F_j(\theta_j^c)]$. Taking the derivative gives $g_j^*(q) = \theta_j'(q) f_j(\theta)/[1 - F_j(\theta_j^c)]$. Thus the inverse hazard rate becomes $[1 - F_j(\theta)]/f_j(\theta) = \theta_j'(q)[1 - G_j^*(q)]/g_j^*(q)$.

From (4), replacing the left-hand side by $T_j'(q)$ and $[1 - F_j(\theta)]/f_j(\theta)$ by $\theta_j'(q)[1 - G_j^*(q)]/g_j^*(q)$ in the right-hand side gives

$$\theta_j'(q) = \frac{T_j'(q) - \gamma}{u_0'(q)} \frac{g_j^*(q)}{1 - G_j^*(q)}.$$

Integrating both sides from q to $\overline{q}_j$ leads to

$$\theta_j(q) = 1 - \int_q^{\overline{q}_j} \frac{T_j'(x) - \gamma}{u_0'(x)} \frac{g_j^*(x)}{1 - G_j^*(x)} dx, \tag{10}$$

where the normalization of Assumption 6 is used. Equation (10) shows that $\theta_j(\cdot)$ is identified wherever both $u_0'(\cdot)$ is identified and $g_j^*(\cdot)$ is observed.

Once $\theta_j(\cdot)$ is identified, $v_j'(\cdot)$ is identified by plugging (10) into (5). That is,

$$v_j'(q) = T_j'(q) - \theta_j(q) u_0'(q). \tag{11}$$

To understand better the intuition behind these results, I consider alternative expressions for (10) and (11). Using (5), (10) can be rewritten equivalently as

$$\theta_j(q) = 1 - \int_q^{\overline{q}_j} \frac{T_j'(x) - \gamma}{T_0'(x)} \frac{g_j^*(x)}{1 - G_j^*(x)} \theta_0(x) dx. \tag{12}$$

Intuitively, the difference between a consumer's taste and the highest taste is the weighted average of $\theta_0(\cdot)$ over $[q, \overline{q}_j]$, where the weight is determined by the slope of the tariff functions and the conditional distribution of phone usage.

Similarly, (11) can be rewritten equivalently as

$$v_j'(q) = T_j'(q) - \frac{\theta_j(q)}{\theta_0(q)} T_0'(q).$$

Here again, the weighted shape difference between the two tariff functions $T_j(\cdot)$ and $T_0(\cdot)$ is used to recover the complementary utility function $v_j(\cdot)$. The weight is determined by the ratio of corresponding tastes $\theta_j(q)/\theta_0(q)$.

The following proposition summarizes these results.

Proposition 4. *Under Assumptions 1-6, the marginal complementary utility function* $v_j'(\cdot)$ *and the type* $\theta_j(\cdot)$ *are identified on* $[\underline{\underline{q}}_j, \overline{q}_j]$*, where* $\underline{\underline{q}}_j \equiv \max\{\underline{q}_j, \underline{q}_0\}$ *and* $j = 1, 2$*. Moreover, the truncated conditional taste distribution* $F_j^*(\cdot)$ *is identified on* $[\theta_j(\underline{\underline{q}}_j), \overline{\theta}]$*, while the conditional taste distribution* $F_j(\cdot)$ *is identified on* $[\theta_j(\underline{\underline{q}}_j), \overline{\theta}]$*.*

I remark that $v_j'(\cdot)$ is identified only on the range where $u_0'(\cdot)$ is identified. By Lemma 5, $\overline{q}_j \leq \overline{q}_0$. However, it is not necessary that $\underline{q}_j \geq \underline{q}_0$. Therefore, $v_j'(\cdot)$ is not identified on $[\underline{q}_j, \underline{q}_0)$ if $\underline{q}_j < \underline{q}_0$. The type distribution $F_j(\cdot)$ can be recovered from $F_j^*(\cdot)$ on $[\theta_j(\underline{\underline{q}}_j), \overline{\theta}]$ because (i) I observe the proportion of consumers whose phone usage is less than $\underline{\underline{q}}_j$, $F_j(\theta_j(\underline{\underline{q}}_j))$, and (ii) $F_j(\cdot) = F_j(\theta_j(\underline{\underline{q}}_j)) + [1 - F_j(\theta_j(\underline{\underline{q}}_j))]F_j^*(\cdot)$. On the other hand, usage and payment data do not provide any variation to identify $v_j'(\cdot)$ and $F_j(\cdot)$ on $[0, \underline{\underline{q}}_j)$ and $[\underline{\theta}, \theta_j(\underline{\underline{q}}_j))$, respectively. Finally, if $\underline{q}_j \geq \underline{q}_0$, then θ_j^c is identified because $\theta_j^c = \theta_j(\underline{q}_j)$. As in Proposition 3, $v_j(\cdot)$ is identified using the boundary condition $T_j(\underline{q}_j) = \theta_j^c u_0(\underline{q}_j) + v_j(\underline{q}_j)$.

Corollary 1. *If* $\underline{q}_j \geq \underline{q}_0$*, the complementary utility function* $v_j(\cdot)$ *is identified as*

$$v_j(q) = T_j(\underline{q}_j) - \theta_j^c u_0(\underline{q}_j) + \int_{\underline{q}_j}^{q} v_j'(x)dx, \tag{13}$$

for all $q \in [\underline{q}_j, \overline{q}_j]$*.*

Identification of κ_0, κ_j and Δ_j

It remains to address the identification of the fixed cost parameters κ_0, κ_j and Δ_j. Only the marginal variable cost enters the first-order condition (4). Among the equilibrium conditions, only the cutoff tastes involve the fixed costs. First, I use information from phone-only user data to identify κ_0. Second, I exploit information from bundle users data to identify Δ_1 and Δ_2. However, κ_1 and κ_2 remain not identified. By Assumption 2-(vi), κ_1 and κ_2 are bounded by the monthly fees for internet-only users. The following proposition formalizes these results.

Proposition 5. *Under Assumptions 1-6, we have*

(i) The minimum phone usage $\underline{q}_j > 0$ *for* $j = 0, 1, 2$.

(ii) The parameters κ_1 *and* κ_2 *are bounded, i.e.* $\kappa_j \leq T_j(0)$ *for* $j = 1, 2$.

(iii) The parameter κ_0 *is identified as*

$$\kappa_0 = \frac{\gamma}{T_0'(\underline{q}_0)} T_0(\underline{q}_0) - \gamma \underline{q}_0.$$

(iii) If $\underline{q}_j \geq \underline{q}_0$, *the parameters* Δ_1 *and* Δ_2 *are identified.*

$$\Delta_j = [\frac{\gamma}{T_0'(\underline{q}_0)} T_0(\underline{q}_0) - \gamma \underline{q}_0] - [T_j(\underline{q}_j) - T_j(0) - \gamma \underline{q}_j] + \theta_0(\underline{q}_j) u_0(\underline{q}_j) \frac{T_j'(\underline{q}_j) - \gamma}{T_0'(\underline{q}_j)}, \quad \textit{for } j = 1, 2$$

Because $\underline{q}_j > 0$ for $j = 0, 1, 2$, the optimal usage-based tariffs are nonlinear two-part tariffs. There is a minimum price $T_j(\underline{q}_j)$ for usage above 0 but lower than $\underline{q}_j$, and a variable price beyond that. Regarding the identification of the fixed cost parameters, following (1), some consumers switch from bundle-j to internet-j as the firm lowers the cutoff taste. As a result, the difference in the fixed cost $(\kappa_0 + \kappa_j - \Delta_j) - \kappa_j$ affects the optimality of the cutoff tastes. Thus κ_0 and Δ_j relate to the utility, cost and inverse hazard rate at the cutoff values. If $\underline{q}_j \geq \underline{q}_0$, the latter are identified from following Propositions 3 and 4 as discussed above, thereby identifying κ_0 and Δ_j. On the other hand, if $\underline{q}_j < \underline{q}_0$, Δ_j is not identified although I can still derive a bound for it. Using $u_0''(\cdot) \leq 0$, we have

$$\Delta_j \leq [\frac{\gamma}{T_0'(\underline{q}_0)} T_0(\underline{q}_0) - \gamma \underline{q}_0] - [T_j(\underline{q}_j) - T_j(0) - \gamma \underline{q}_j] + \theta_0(\underline{q}_j)[u_0(\underline{q}_0) - (\underline{q}_0 - \underline{q}_j) u_0'(\underline{q}_0)] \frac{T_j'(\underline{q}_j) - \gamma}{T_0'(\underline{q}_j)}.$$

3.4.2 Estimation

Since the estimation of $v_j(0)$ and $F(\theta_j^c)$ can be calculated directly from internet-only users data, I focus on the estimation of all the other primitives. I then use data from phone-only and bundle users. For convenience, I order them lexicographically by their consumption bundles. For all the consumers having a positive quantity of phone service, I group those using the same internet level $j = 0, 1, 2$ and then order them according to their phone usage q. I denote N_j^* as the number of users for group j. This would give $\{(q_j^i, t_j^i)\}_{i=1,\ldots,N_j^*}$, where

$0 < q_j^1 \leq q_j^2 \leq \ldots \leq q_j^{N_j^*}$ and $0 < t_j^1 \leq t_j^2 \leq \ldots \leq t_j^{N_j^*}$.

I propose a three-step estimation procedure. First, I estimate γ and $\xi(\cdot)$ using $\gamma = T_0'(\overline{q}_0)$ and (8), respectively. An estimate for $\xi(\cdot)$ will allow me (i) to obtain an estimate of the marginal intrinsic utility function $u_0'(\cdot)$ using (6) and (ii) to construct a sample of pseudo tastes for phone-only users from (7). To complete the estimation of $u_0(\cdot)$, I will estimate θ_0^c and $u_0(\underline{q}_0)$ using $\theta_0^c = \theta_0(\underline{q}_0)$ and $T_0(\underline{q}_0) = \theta_0^c u_0(\underline{q}_0)$, respectively. Second, the estimated marginal intrinsic utility function is used to (i) estimate the marginal complementary utility functions $v_j'(\cdot)$ using (11) and (ii) to construct a sample of pseudo tastes for bundle-j users from (12). To complete the estimation of $v_j(\cdot)$, I will estimate θ_j^c, $u_0(\underline{q}_j)$ and $v_j(\underline{q}_j)$ using $\theta_j^c = \theta_j(\underline{q}_j)$ and $T_j(\underline{q}_j) = \theta_j^c u_0(\underline{q}_j) + v_j(\underline{q}_j)$, respectively. Third, I use the estimated pseudo tastes to estimate the conditional taste densities.

Estimation of γ, $u_0(\cdot)$ and $\theta_0(\cdot)$

In this subsection, I use phone-only users data, i.e. $\{(q_0^i, t_0^i)\}_{i=1,\ldots,N_0^*}$. To obtain an estimate of γ, I need to estimate $\overline{q}_0$. A convenient estimator that converges very fast is to take the maximum value, i.e. $\hat{\underline{q}}_0 = q_0^{N_0^*}$, leading to an estimator of γ, i.e. $\hat{\gamma} = T_0'(q_0^{N_0^*})$.

Following (8), I need to estimate $G_0^*(\cdot)$. I use the following empirical distribution estimator leading to

$$\hat{G}_0^*(q) = \frac{1}{N_0^*} \sum_{i=1}^{N_0^*} \mathbb{1}(q_0^i \leq q), \tag{14}$$

for an arbitrary value of $q \in [\underline{q}_0, \overline{q}_0]$. An estimator of $\xi(\cdot)$ is obtained by replacing $G_0^*(\cdot)$ by its empirical distribution $\hat{G}_0^*(\cdot)$ and γ by $\hat{\gamma}$. Since $\hat{G}_0^*(\cdot)$ is a step function, the integral in (8) can be rewritten as a finite sum of integrals. Since in each of these integrals, $\log(1 - G_0^*(\cdot))$ is a constant and the primitive of $T_0''(\cdot)/T_0'(\cdot)$ is $-1/T_0'(\cdot)$, $\hat{\xi}(q)$ is equivalent to

$$\begin{aligned}\hat{\xi}(q) = & [1 - \hat{G}_0^*(q)]^{1 - \frac{\hat{\gamma}}{T_0'(q)}} \\ & \times \exp\Big\{\hat{\gamma}\Big[\frac{1}{T_0'(q)} - \frac{1}{T_0'(q_0^{l+1})}\Big] \log(1 - \hat{G}_0^*(q_0^l)) + \hat{\gamma} \sum_{k=l+1}^{N_0^*-1} \Big[\frac{1}{T_0'(q_0^k)} - \frac{1}{T_0'(q_0^{k+1})}\Big] \log(1 - \hat{G}_0^*(q_0^k))\Big\},\end{aligned}$$

for $q \in [q_0^l, q_0^{l+1})$, where $l = 0, 1, \ldots, N_0^* - 1$. For $q \in [q_0^{N_0^*}, \overline{q}_0]$, I have $\hat{\xi}(q) = 1$.

Using (6) and (7), I then estimate $u_0'(\cdot)$ and $\theta_0(\cdot)$ by

$$\hat{u}_0'(q) = T_0'(q)\hat{\xi}(q),$$
$$\hat{\theta}_0(q) = \frac{1}{\hat{\xi}(q)},$$

for an arbitrary value of $q \in [\underline{q}_0, \overline{q}_0]$. Lastly, I estimate θ_0^c by $\hat{\theta}_0^c = \hat{\theta}_0(q_0^1)$ and $u_0(\underline{q}_0)$ by $\widehat{u_0(\underline{q}_0)} = T_0(\underline{q}_0)/\hat{\theta}_0^c$. These estimates will allow me to obtain an estimate of $u_0(\cdot)$ following (9).

Estimation of $v_j(\cdot)$, $\theta_j(\cdot)$, κ_0, κ_1, κ_2, Δ_1 and Δ_2

In this subsection, I am now using the bundle users data $\{(q_1^i, t_1^i)\}_{i=1,\ldots,N_1^*}$ and $\{(q_2^i, t_2^i)\}_{i=1,\ldots,N_2^*}$. Following (10), since estimates for γ and $u_0'(\cdot)$ have been obtained previously, I need an estimate of $g_j^*(\cdot)$ and $G_j^*(\cdot)$. I use the empirical distribution for $G_1^*(\cdot)$ and $G_2^*(\cdot)$ from (14) by replacing 0 by 1 and 2, respectively. For the density, I use a kernel density estimator to estimate $g_1^*(\cdot)$ and $g_2^*(\cdot)$. Following (10) and (11), the pseudo type $\theta_j(\cdot)$ and the marginal complementary utility $v_j'(\cdot)$ can be estimated as

$$\hat{\theta}_j(q) = 1 - \int_q^{\overline{q}_j} \frac{T_j'(x) - \hat{\gamma}}{\hat{u}_0'(x)} \frac{\hat{g}_j^*(x)}{1 - \hat{G}_j^*(x)} dx,$$
$$\hat{v}_j'(q) = T_j'(q) - \hat{\theta}_j(q)\hat{u}_0'(q),$$

where $q \in [\underline{q}_j, \overline{q}_j]$ and $j = 1, 2$. Lastly, I estimate θ_j^c by $\hat{\theta}_j^c = \hat{\theta}_j(q_j^1)$ and $u_0(\underline{q}_j)$ by $\widehat{u}_0(q_j^1)$. These estimates will allow me to obtain an estimate of $v_j(\cdot)$ following (13).

Following Proposition 5, the fixed cost parameters κ_0, Δ_1 and Δ_2 are estimated by

$$\hat{\kappa}_0 = \frac{\hat{\gamma}}{T_0'(\underline{q}_0)} T_0(\underline{q}_0) - \hat{\gamma}\underline{q}_0,$$
$$\hat{\Delta}_j = [\frac{\hat{\gamma}}{T_0'(\underline{q}_0)} T_0(\underline{q}_0) - \hat{\gamma}\underline{q}_0] - [T_j(\underline{q}_j) - T_j(0) - \gamma\underline{q}_j] + \hat{\theta}_0(\underline{q}_j)\hat{u}_0(\underline{q}_j)\frac{T_j'(\underline{q}_j) - \hat{\gamma}}{T_0'(\underline{q}_j)}, \quad \text{for } j = 1, 2,$$

where $\underline{q}_0$ and $\underline{q}_j$ can be replaced by their estimated counterparts, which are q_0^1 and q_j^1. The bounds for κ_1 and κ_2 are directly obtained from the data as the monthly fees for the internet-only users.

Estimation of $f_0(\cdot)$, $f_1(\cdot)$ and $f_2(\cdot)$

The previous two steps provide estimates of the pseudo types $\{\hat{\theta}_0^1, \hat{\theta}_0^2, \ldots, \hat{\theta}_0^{N_0^*}, \hat{\theta}_1^1, \hat{\theta}_1^2, \ldots, \hat{\theta}_1^{N_1^*}, \hat{\theta}_2^1, \hat{\theta}_2^2, \ldots, \hat{\theta}_2^{N_2^*}\}$, where $\hat{\theta}_j^i = \hat{\theta}_j(q_j^i)$ for $i = 1, 2, \ldots, N_j^*$ and $j = 0, 1, 2$. I could use standard kernel estimators to estimate $f_0^*(\cdot)$, $f_1^*(\cdot)$ and $f_2^*(\cdot)$ using these pseudo values. From the model of Section 3, the conditional density of types should satisfy the hazard rate property given by Assumption 2-(v). I then propose a new regression spline estimator that allows me to impose the monotonicity restriction on $H(\cdot|j)$ for $j = 0, 1, 2$. In addition, I remark that $H(\theta|j)$ is bounded by θ since $[1 - F_j(\theta)]/f_j(\theta) \geq 0$. This represents a bound restriction that will also be imposed in the estimator. Specifically, I estimate $f_j^*(\cdot)$ under the restrictions that $H_j(\cdot) \equiv \cdot - \frac{1-F_j(\cdot)}{f_j(\cdot)}$ is increasing on $[\theta_j^c, 1]$, and $H_j(\theta) \leq \theta$ for all $\theta \in [\theta_j^c, 1]$. This new estimator is studied in Luo (in progress). For the sake of completeness, I briefly describe the main idea here.

Let the hazard function be $h_j(\theta) = f_j^*(\theta)/[1 - F_j^*(\theta)] = f_j(\theta)/[1 - F_j(\theta)] = 1/[\theta - H_j(\theta)]$ for $j = 0, 1, 2$. Using splines to approximate $H_j(\cdot)$, I can use the well-known expression

$$f_j^*(\theta) = h_j(\theta) \exp\big[- \int_{\theta_j^c}^{\theta} h_j(x)dx\big],$$

to obtain a maximum likelihood estimate for $f_j^*(\cdot)$. However, the direct implementation of the usual quadratic splines can ensure the monotonicity restriction but may violate the bound restriction. Instead, I define quadratic splines imposing the former, and then transform the coordinates to ensure the latter. In particular, I define the knots $\theta_j^c = \vartheta_j^0 < \vartheta_j^1 < \cdots < \vartheta_j^{k_j} < \vartheta_j^{k_j+1} = 1$. For any $\theta \in [\theta_j^c, 1]$, let $\psi(\theta; \delta_j) \equiv \sum_{l=1}^{k_j+3} \delta_j^l s_j^l(\theta)$, where the $s_j^l(\cdot)$s are quadratic basis functions satisfying $\psi(\cdot; \delta_j)$ positive and increasing if and only if the coefficients δ_j are positive. I then define

$$H_j(\theta; \delta_j) = \theta \frac{\psi(\theta; \delta_j)}{1 + \psi(\theta; \delta_j)}.$$

One can show that $H_j(\cdot; \delta_j)$ is positive, increasing and bounded by the 45 degree line if the coefficients δ_j^l are non-negative. Therefore, the hazard function can be expressed as

$h_j(\theta;\delta_j) = [1+\psi(\theta;\delta_j)]/\theta$, from which I can construct the log-likelihood of the pseudo sample as

$$l_j(\delta_j) = \sum_{i=1}^{N_j^*} \Big\{ \log\big[\frac{1+\psi(\hat{\theta}_j^i;\delta_j)}{\hat{\theta}_j^i}\big] - \int_{\hat{\theta}_j^c}^{\hat{\theta}_j^i} \frac{1+\psi(x;\delta_j)}{x} dx \Big\}.$$

Finally, I estimate $f_j^*(\cdot)$ by

$$\hat{f}_j^*(\theta) = \hat{h}_j(\theta)\exp[-\int_{\theta_j^c}^{\theta} \hat{h}_j(x)dx],$$

for $\theta \in [\theta_j^c, 1]$, where $\hat{h}_j(\theta) = [1+\psi(\theta;\hat{\delta}_j)]/\theta$ and $\hat{\delta}_j$ maximizes $l_j(\delta_j)$.[14]

3.5 Empirical Analysis of China Telecom Data

3.5.1 Estimation Results

As discussed in Section 2, I estimate the tariff functions $T_0(\cdot)$, $T_1(\cdot)$ and $T_2(\cdot)$ and the resulting phone usage q. See Appendix B and Figure 3 displaying $T_0(\cdot)$, $T_1(\cdot)$ and $T_2(\cdot)$.

Regarding the cost, I obtain an estimate for the marginal variable cost γ which is equal to 5.95 cents. This is approximately a fourth of the average price charged per minute. The fixed cost for phone service is 6.58 RMB, while the estimate of the bound for the fixed cost of 1 Mbps (2 Mbps) internet κ_1 (κ_2) is 78 RMB (88 RMB). The estimate of the cost saving parameter for 1 Mbps internet Δ_1 is 2.60 RMB while this value increases to 3.27 RMB for Δ_2. The fixed cost for phone service seems to be small. It is approximately a tenth of the average bill of phone-only subscribers. These cost parameters suggest that the firm has a comfortable profit margin as discussed later. Relative to providing internet only, providing a bundle does not impose much additional cost to the firm as suggested by the estimates of Δ_1 and Δ_2. China Telecom mainly uses Asymmetric Digital Subscriber Line (ADSL) to provide internet service. Thus internet is transmitted through telephone lines. Moreover, fixed transaction costs such as mailing statement do not increase much because bills are merged if the consumer uses a bundle.

[14] This estimator does require much computing time because of the explicit form of the log-likelihood function. It remains to study the asymptotic properties of this new estimator.

I then obtain estimates of the marginal intrinsic utility $u_0'(\cdot)$ and the marginal complementary utility functions $v_1'(\cdot)$ and $v_2'(\cdot)$. The first is displayed in Figure C.4 while the latter two are displayed in Figure C.5. The estimated marginal intrinsic utility $u_0'(\cdot)$ is positive and decreasing, thereby satisfying Assumption 5-(i). The estimated marginal complementary utility functions $\hat{v}_1'(\cdot)$ and $\hat{v}_2'(\cdot)$ are both negative and increasing with $\hat{v}_1'(\cdot)$ above $\hat{v}_2'(\cdot)$, thereby satisfying Assumption 3-(ii). Since both are negative, internet and phone services seem to be substitutes. Internet offers alternative communication tools such as email, skype and so on, which can explain the substitutability with phone service. Thus the utility of a bundle user is smaller than the sum of the utilities for a phone service user only and a internet user only. Moreover, this substitution effect is stronger with a higher level of internet because a faster internet service allows better alternative communication tools.

Figure C.6 displays the inverse of the estimated $\theta_0(\cdot)$, $\theta_1(\cdot)$ and $\theta_2(\cdot)$. They are increasing in the type and decreasing in the internet choice j, thereby satisfying Lemma 5. As internet speed increases, a larger range of low taste (for phone service) consumers are excluded from using phone service. These observations satisfy the predictions of my model in Section 3. Figure C.7 displays the estimated type densities $f_0^*(\cdot)$, $f_1^*(\cdot)$ and $f_2^*(\cdot)$. As the internet speed increases, the density function becomes less skewed to the left, thereby implying that consumers are more likely to have a higher taste for phone service. Here again, one sees the increase in the cutoff taste as the level of internet increases. Figure C.8 displays the hazard rate functions $H_0(\cdot)$, $H_1(\cdot)$ and $H_2(\cdot)$. They are increasing in the type and decreasing in the internet choice j, thereby satisfying Assumptions 2-(v) and 4.

Using these estimated values, I can access empirically the firm's profit as well as the consumers' informational rents. The informational rent is estimated by $\hat{\theta}_j^i \hat{u}_0(q_j^i) + \hat{v}_j(q_j^i) - T_j(q_j^i)$. The ratio of the total informational rent across all consumers by the total amount paid is 29.76%. This measures the cost of asymmetric information. When considering by group of users, I find 53.27% for phone-only users, 27.02% for bundle users with 1 Mbps of internet level and 27.66% for bundle users with 2 Mbps. These overall rents tend to decrease with the level of internet. I recall that internet-only users do not enjoy any rent since the firm can extract all their rents by charging them a fixed fee. Regarding the firm's profit, because I obtain only bounds for the cost parameters κ_1 and κ_2, namely 78 RMB for the former and 88 RMB for the latter, the firm's profit margin ranges from 39.46% (if κ_1=78 and κ_2=88) to 74.06% (if κ_1=0 and κ_2=0). The profit margins for different groups are of the following: group 0 at 54.36%, group 1 between 33.24% (if κ_1=78) and 76.43% (if κ_1=0) and group 2 between 40.96% (if κ_2=88) and 75.78% (if κ_2=0). Overall, China Telecom seems to be making a comfortable profit margin.

3.5.2 The Welfare Effects of Bundling and Nonlinear Pricing

With structural estimates at hand, I can perform a counterfactual to evaluate the effects of bundling on firm's profit, consumer surplus and social welfare relative to component pricing. In particular, I simulate the case where the firm offers instead two fixed-fee contracts for internet and one usage-based contract for phone service. I assume that the firm does not change the internet speeds it offers.

A Theoretical Discussion

In general, the effects of bundling on consumer surplus and social welfare are ambiguous. However, the literature offers a consensus that lower prices or higher output levels are necessary for welfare improvement. For instance, in a discrete choice framework, Salinger (1995) shows that bundling can increase consumer surplus when it results in lower prices. Schmalensee (1981) and Schwartz (1990) formally show that welfare must fall if output does not rise with third degree price discrimination. While the previous literature has mainly focused on the use of bundling as a price discrimination device (See Kobayashi (2005)), my model allows utility complementarity, cost saving effects and dependence between the two dimensions of asymmetric information. They may have different roles in determining the welfare effects of mixed bundling relative to unbundling. I will construct two examples below showing the ambiguity of the results in my model.

Under component pricing, the firm's problem is to maximize its profit by designing a tariff function while the consumers' taste are distributed according to a mixture of three conditional distributions. Since these three distributions differ in location and shape, their mixture is obtained by *shifting* and *reshaping* them. To isolate the role of these two procedures, I consider two examples. Let $U(q,j,\theta) = \theta q - \frac{1}{2}q^2$ if $q \leq \theta$ and $\theta^2/2$ otherwise, while the cost function is $c(q,j) = \kappa_0 + \gamma q$.

First, I consider two groups whose tastes for phone service are uniformly distributed on $[\mathsf{L}, \mathsf{M}]$ and $[\mathsf{M}, \mathsf{R}]$, respectively. If the firm can discriminate among the two groups, the optimal phone service assignments would be $q_1^*(\theta) = 2\theta - \mathsf{M} - \gamma$ and $q_2^*(\theta) = 2\theta - \mathsf{R} - \gamma$, while the cutoff tastes would be $\theta_1^c = (\sqrt{2\kappa_0} + \mathsf{M} + \gamma)/2$ and $\theta_2^c = (\sqrt{2\kappa_0} + \mathsf{R} + \gamma)/2$. If the firm cannot discriminate, it proposes a single assignment $q^*(\theta) = 2\theta - \mathsf{R} - \gamma$ with a cutoff taste $\theta^c = (\sqrt{2\kappa_0} + \mathsf{R} + \gamma)/2$. I remark that in this case, the firm does as it was facing only consumers with a higher need of internet. Thus bundling benefits to the consumers because consumers with a lower taste will not be excluded. This would results an increase in the consumer surplus. Similarly, since the firm would get the same profit from the consumers with a higher taste of internet, it will get a larger profit as bundling will allow it to get profit from the other group of consumers as well.

Second, I consider two groups whose tastes for phone service are distributed on the same interval $[0.25, 1.25]$ with densities $f_1(\theta) = 2.5 - 2\theta$ and $f_2(\theta) = 1$. Group 1 accounts for a proportion of λ. Figure C.9 displays the firm's profit (dash-dot lines) and consumer surplus (dashed lines) under mixed bundling in blue and component pricing in red. In this case, the firm benefits from bundling while the consumer are penalized. Thus bundling can reduce consumer surplus because it provides an additional instrument for the firm to discriminate across consumers.

Counterfactual Simulation

Solving the firm's problem under unbundling is a hard task because the model does not lead to a closed-form solution, I will propose instead a numerical approximation of the solution. In particular, I will search numerically an optimal usage-based tariff function that is approximated by quadratic splines $T(\cdot;\delta) = \sum_{k=1}^{K} \delta_k \psi_k(\cdot)$, where $\psi_k(\cdot)$ is a quadratic basis function. The tariff $T(\cdot;\delta)$ is non-negative and increasing if and only if the coefficients δ_k are non-negative. Moreover, as I do not identify the type densities $f_0(\cdot)$, $f_1(\cdot)$ and $f_2(\cdot)$

below the cutoff tastes, I assume

$$\hat{f}_j(\theta) = \begin{cases} \hat{f}_j^*(\hat{\theta}_j^c)[1 - F_j(\theta_j^c)](\frac{\theta}{\hat{\theta}_j^c})^k & \text{if } \theta < \hat{\theta}_j^c, \\ \hat{f}_j^*(\hat{\theta}_j^c)[1 - F_j(\theta_j^c)] & \text{if } \theta \geq \hat{\theta}_j^c, \end{cases}$$

where $k = [\hat{\theta}_j^c \hat{f}_j^*(\hat{\theta}_j^c)(1-F_j(\theta_j^c))/F_j(\theta_j^c)]-1$. This approximation satisfies all the assumptions of Section 3. It allows continuity at the cutoff point $(\hat{\theta}_j^c, \hat{f}_j^*(\hat{\theta}_j^c)[1-F_j(\theta_j^c)])$ and its integration from 0 to $\hat{\theta}_j^c$ equals $F_j(\theta_j^c)$.

The estimated optimal tariff function solves the following problem

$$\begin{aligned} \max_{\delta \geq 0} \quad & N_0 \int_{\underline{\theta}}^{\overline{\theta}} \left(T(q_0(\theta;\delta);\delta) - (\hat{\kappa}_0 + \hat{\gamma} q_0(\theta;\delta))\right) \hat{f}_0(\theta) d\theta \\ & + N_1 \int_{\underline{\theta}}^{\overline{\theta}} \left(T(q_1(\theta;\delta);\delta) - (\hat{\kappa}_0 + \kappa_1 - \hat{\Delta}_1 + \hat{\gamma} q_1(\theta;\delta))\right) \hat{f}_1(\theta) d\theta \\ & + N_2 \int_{\underline{\theta}}^{\overline{\theta}} \left(T(q_2(\theta;\delta);\delta) - (\hat{\kappa}_0 + \kappa_2 - \hat{\Delta}_2 + \hat{\gamma} q_2(\theta;\delta))\right) \hat{f}_2(\theta) d\theta, \end{aligned}$$

where $q_j(\theta;\delta) \equiv \arg\max_q \{\theta \hat{u}_0(q) + \hat{v}_j(q) - T(q;\delta)\}$ and N_j is the number of group-j consumers. Since κ_1 and κ_2 are not identified, I set them to zero. I use equally-spaced knots and increase number of parameters K until the marginal benefit of adding one more knot is less than 0.1%. The resulting estimated tariff function captures the optimal nonlinear price schedule under unbundling. See e.g. Wilson (1997) who shows that the loss of using n-part price schedules relative to the optimal nonlinear price schedule is of order $1/n^2$.

My simulation results show that unbundling would lead to a 10.14% decrease in firm's profit and a 17.18% decrease in consumer surplus, resulting in a 12.16% decrease in social welfare. These can be explained by the fact that unbundling would exclude too many consumers as discussed previously in my first numerical example. Figure C.10 compares the tariff functions under mixed bundling (dashed lines) and component pricing (solid lines). Relative to mixed bundling, groups 0 and 1 would face more expensive tariff functions under component pricing, while group 2 would face a less expensive one. As a result, group 0 would lose by 58.11% of consumer surplus, while group 1 would lose 39.57%. On the contrary, group 2 would see an increase in their surplus by 9.48%. Thus unbundling would only benefit to those who value highly internet.

Figure C.11 displays the breakdown of expected social welfare into consumer surplus and firm profit, while Figure C.12 displays the breakdown of expected bill into cost and firm profit. I treat mixed bundling as the benchmark and normalize its corresponding welfare and bill to 100. Since the cost function is linear, changes in expected cost reflect changes in expected phone usage. Figure C.11 confirms that groups 0 and 1 are losing the most in terms of consumer surplus under component pricing. The firm is losing profit as well. The loss is decreasing with internet speed. Figure C.12 provides a justification, namely the production cost much decreases under component pricing because of a dramatic decrease in consumption of phone service. For instance, group 0 users' expected phone usage would drop from 497.62 to 253.09 minutes, while their expected indirect utility would decrease from 34.54 to 14.47 RMB. On the contrary, group 2 users would use 7.09% more of phone calls and thus would see their consumer surplus increasing by 9.48%. Finally, the ratio of the total informational rent by the total of bill would be 27.71% in component pricing, which represents a decrease in the cost of asymmetric information relative to bundling. This arises from a larger proportion of consumers who would be excluded under component pricing.

3.6 Conclusion

This paper studies bundling and price discrimination by a multiproduct firm selling internet and phone services in an imperfect information setting. Consumers are characterized by a taste for phone service and a minimum need for internet, thereby leading to a multidimensional screening problem. I derive the optimal selling mechanism, as well as the conditions on the model primitives under which different bundling strategies arise. I show that the model primitives are identified under parameterization of the cost function and multiplicative separability of the utility function. I develop a semiparametric estimator involving kernel density estimation and sieve estimators. The empirical analysis of China Telecom data suggests that both the firm and consumers benefit from bundling internet and phone services.

A first extension would be to consider a multiproduct model that would distinguish land line and mobile services. With the methodology I develop in this paper, potential applications include insurance contracts in which insurees bundle automobile and home insurance, and also a large number of products from manufacturing industries such as automobiles or computers where each product can be viewed as a bundle of various customized attributes. Lastly, the results I developed in this paper can also be used to analyze products under nonlinear pricing with important network effects. See e.g. Chen and Luo (2012).

Appendix A

Nonlinear Pricing with Product Customization

Proofs

Proposition 1. Under Assumptions 1, given a cutoff taste $\theta_c \in [\underline{\theta}, \overline{\theta}]$ such that consumers with $\theta < \theta_c$ are excluded, the functions $(q(\cdot, \epsilon), \tau(\cdot; \epsilon))$ that solve the firm's problem satisfy:

$$\tau'\left(q(\theta, \epsilon)\, ; \epsilon\right) = \epsilon\theta u_0'(\epsilon q(\theta, \epsilon)),$$
$$\epsilon\theta u_0'(\epsilon q(\theta, \epsilon)) = \epsilon\gamma + \epsilon u_0'(\epsilon q(\theta, \epsilon))\frac{1 - F(\theta)}{f(\theta)},$$

where the prime denotes a derivative with respect to its first argument for $\tau(\cdot; \cdot)$. The first equation is the consumer's first-order condition while the second one is the firm's. Moreover, "no informational rent at the bottom" defines a boundary condition: $\underline{\tau}(\epsilon) = \tau(q(\theta_c, \epsilon); \epsilon) = \theta_c u_0(\epsilon q(\theta_c, \epsilon))$. Dividing both sides by ϵ, the firm's first-order condition implies that $\epsilon q\left(\theta, \epsilon\right)$ would not depend on ϵ. Thus, there exists a function $Q(\cdot)$ such that $Q(\theta) \equiv \epsilon q(\theta, \epsilon)$ and

$$\theta u_0'(Q(\theta)) = \gamma + \frac{1 - F(\theta)}{f(\theta)} u_0'(Q(\theta)).$$

Moreover, the boundary condition does not depend on ϵ either. Namely, $\underline{\tau} = \theta_c u_0(Q(\theta_c))$.

The consumer's first-order condition then can be rewritten as

$$\frac{\tau'(q(\theta,\epsilon);\epsilon)}{\epsilon} = \theta u_0'(Q(\theta)),$$

whose right-hand side depends only on θ. Denote the right-hand side as $r(\theta)$. Then this equation implies $\tau'(q;\epsilon) = \epsilon r(Q^{-1}(q\epsilon))$. Integrating from $Q(\theta_c)/\epsilon$ to q and making the change of variable $\tilde{x} = x\epsilon$ gives

$$\tau(q;\epsilon) = \underline{\tau} + \int_{Q(\theta_c)/\epsilon}^{q} \epsilon r(Q^{-1}(x\epsilon))dx = \underline{\tau} + \int_{Q(\theta_c)}^{q\epsilon} r(Q^{-1}(\tilde{x}))d\tilde{x},$$

which implies that there exists a function $T(\cdot)$ such that $\tau(q;\epsilon) = T(q\epsilon)$. It is easy to see that the consumer's first-order condition is equivalent to $T'(Q(\theta)) = \theta u_0'(Q(\theta))$.

Notice that the allocation $Q(\cdot)$ does not depend on the cutoff taste θ_c while the optimal tariff function $T(\cdot)$ does. I denote the latter as $T(\cdot;\theta_c)$. I then find an optimal θ_c to maximize the firm's profit:

$$\int_{\theta_c}^{\bar{\theta}} \left[T(Q(x);\theta_c) - \gamma Q(x)\right] f(x)dx$$

whose first-order condition with respect to θ_c leads to Equation 3. □

Proposition 2. The identification of α is immediate under Assumption 2. I now show that $T(\cdot)$ is identified. Define $Y \equiv \alpha'\mathbf{q}$ and $\lambda(\cdot) \equiv \Lambda'(\cdot)$. Let $\tilde{G}(\cdot|t)$ be the CDF of Y conditional on t and $\tilde{F}(\cdot)$ be the CDF of $-\log\epsilon$. Equation (5) implies that $\tilde{G}(y|t) = \tilde{F}(y - \Lambda(t))$. Therefore, $\tilde{G}_y(y|t) = \tilde{f}(y-\Lambda(t))$, $\tilde{G}_t(y|t) = -\lambda(t)\tilde{f}(y-\Lambda(t))$, and $\lambda(t) = -\tilde{G}_t(y|t)/\tilde{G}_y(y|t)$ for any (y,t) such that $\tilde{G}_y(y|t) \neq 0$. It follows that, given $t_\dagger \in [\underline{t},\bar{t}]$,

$$\Lambda(t) = \Lambda(t_\dagger) - \int_{t_\dagger}^{t} \frac{\tilde{G}_t(y|x)}{\tilde{G}_y(y|x)}dx = \mathrm{E}\left(y|t_\dagger\right) - \int_{t_\dagger}^{t} \frac{\tilde{G}_t(y|x)}{\tilde{G}_y(y|x)}dx,$$

where the second equation is obtained using Assumption 2-(iii). This equation implies that

$$\Lambda(t) = \int_{S_w} w(y)\left[\mathrm{E}\left(y|t_\dagger\right) - \int_{t_\dagger}^{t} \frac{\tilde{G}_t(y|x)}{\tilde{G}_y(y|x)}dx\right]dy,$$

where $w(\cdot)$ is a scalar-valued function on $I\!R$ with compact support S_w such that $\int_{S_w} w(y)dy =$

1. Therefore, $T(\cdot)$ is identified. $\square$

Lemma 1. Evaluating Equations (1) and (2) at $\overline{\theta}$ gives $\gamma = T'(\overline{Q}) = T'(T^{-1}(\bar{t}))$. Moreover, evaluating FOCs and boundary conditions at θ_0 gives

$$\begin{aligned}
K &= \theta_0 u_0(\underline{Q}) - \gamma \underline{Q} - \frac{1 - F(\theta_0)}{f(\theta_0)} u_0(\underline{Q}) \\
&= \theta_0 u_0(\underline{Q}) - \gamma \underline{Q} - \frac{T'(\underline{Q}) - \gamma}{u_0'(\underline{Q})} u_0(\underline{Q}) \\
&= \theta_0 u_0(\underline{Q}) - \gamma \underline{Q} - \frac{T'(\underline{Q}) - \gamma}{T'(\underline{Q})/\theta_0} u_0(\underline{Q}) \\
&= \underline{t} - \gamma \underline{Q} - \frac{T'(\underline{Q}) - \gamma}{T'(\underline{Q})} \underline{t} \\
&= \gamma \Big[\frac{\underline{t}}{T'(T^{-1}(\underline{t}))} - T^{-1}(\underline{t}) \Big].
\end{aligned}$$

The first equation follows from Equation (3), while the second and third equations follow from FOCs evaluated at θ_0. The second last equation follows from Equation (4), while the last equation follows definition. $\square$

Equation (8). I have the estimating equations

$$\mathrm{Var}[D_b(t)\alpha'\mathbf{q}|t] - D_b(t)\mathrm{Var}[\alpha'\mathbf{q}|t] = 0$$

for every $b \in \{1, 2, \ldots, \mathcal{B}\}$ and $t \in [\underline{t}, \bar{t}]$. Taking expectation of the left-hand side gives

$$\begin{aligned}
&\alpha' \Big[\mathrm{E}\left\{ \mathrm{Var}[D_b(t)\mathbf{q}|t] \right\} - \mathrm{E}[D_b(t)]\mathrm{E}(\mathrm{Var}[\mathbf{q}|t]) \Big] \alpha \\
=&\alpha' \Big[\mathrm{E}\Big\{ \mathrm{E}\big[D_b(t)(\mathbf{q} - \mathrm{E}(q|t))(\mathbf{q} - \mathrm{E}(\mathbf{q}|t))'|t\big] \Big\} - \mathrm{E}[D_b(t)]\mathrm{E}\Big\{ \mathrm{E}\big[(\mathbf{q} - \mathrm{E}(q|t))(\mathbf{q} - \mathrm{E}(\mathbf{q}|t))'\big]|t \Big\} \Big] \alpha \\
=&\alpha' \Big[\mathrm{E}\Big\{ D_b(t)(\mathbf{q} - \mathrm{E}(\mathbf{q}|t))(\mathbf{q} - \mathrm{E}(\mathbf{q}|t))' \Big\} - \mathrm{E}[D_b(t)]\mathrm{E}\Big\{ (\mathbf{q} - \mathrm{E}(\mathbf{q}|t))(\mathbf{q} - \mathrm{E}(\mathbf{q}|t))' \Big\} \Big] \alpha
\end{aligned}$$

for $b \in \{1, 2, \ldots, \mathcal{B}\}$. I use $D_b^2(t) = D_b(t)$ and $\mathrm{Var}[\alpha'\mathbf{q}|t]$ is a constant. $\square$

Figures and Tables

Figure A.1: Tariff $\hat{T}(\cdot)$

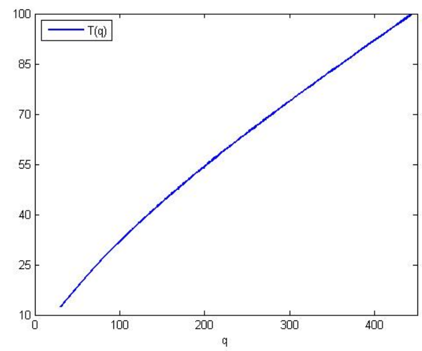


Figure A.2: One-to-one mapping $\hat{\theta}(\cdot)$

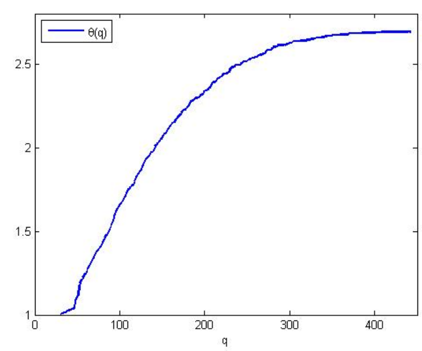


Figure A.3: Base utility $\hat{u}_0(\cdot)$

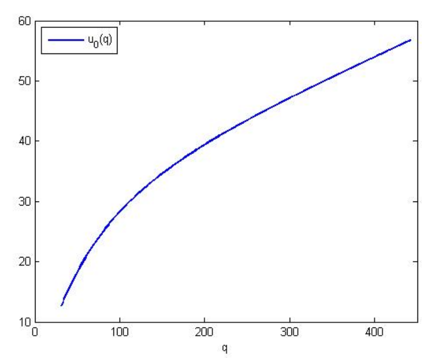


Figure A.4: Truncated taste density $\hat{f}^{*}(\cdot)$

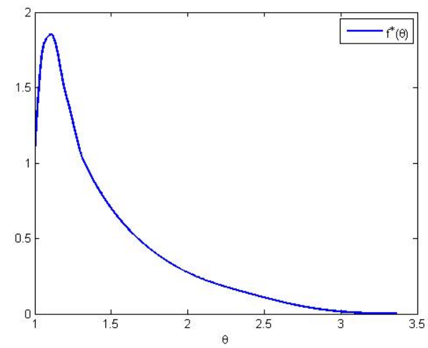


Figure A.5: Density of unobserved heterogeneity $\hat{g}(\cdot)$

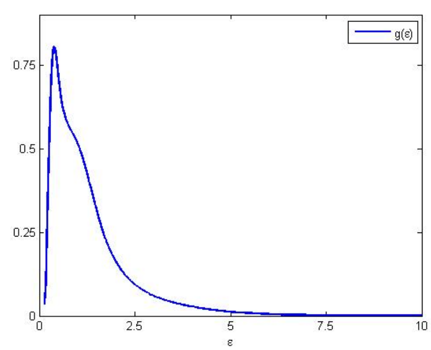


Figure A.6: Tariffs

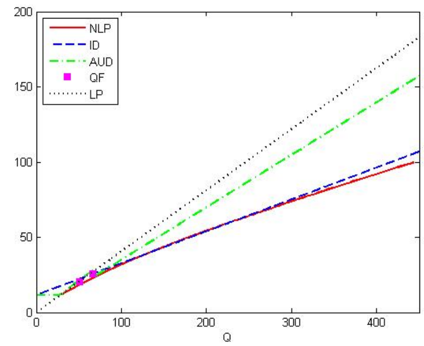


Figure A.7: Consumer surplus

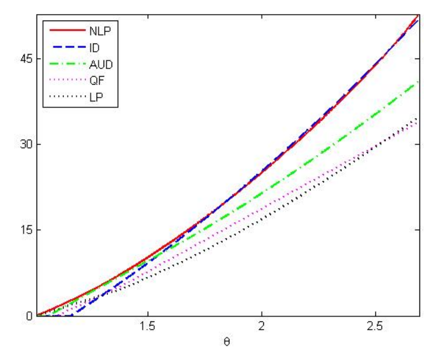


Table A.1: Summary Statistics

Variable[a]	Obs	Mean	Std. Dev.	Min	Max
t	2000	27.27	12.95	12.38	99.67
q_L	2000	571.17	419.43	0	5471
q_D	2000	114.33	183.19	0	1776
q_R	2000	76.13	187.22	0	2626
q	2000	111.28	112.72	2.30	1030.63
Q	2000	85.58	55.29	30.71	442.89
θ	2000	1.45	0.40	1.01	2.69
ϵ	2000	1.34	1.32	0.15	21.48
$rent$	2000	10.37	11.50	0.096	52.84
$rentratio$	2000	0.2939	0.2028	0.0078	0.7083

[a] Note: q_L, q_D and q_R are the quantities of local, distance and roaming calls. q is the aggregation of phone minutes and Q is the effective quantity of mobile service usage. θ is consumer taste while ϵ is unobserved heterogeneity. $rent$ is informational rent left to consumers while $rentratio$ is the ratio of informational rent by payment.

Table A.2: Product Customization

payment range	Variable[a]	Obs	Mean	Std. Dev.
< 24	t	1065	18.64	2.71
	minutes	1065	504.43	338.58
$[24, 34)$	t	518	28.68	2.81
	minutes	518	823.00	282.03
$[34, 44)$	t	219	38.55	2.69
	minutes	219	1127.06	343.36
$[44, 54)$	t	102	48.24	2.90
	minutes	102	1363.82	379.92
$[54, 64)$	t	49	58.21	2.68
	minutes	49	1660.86	464.87
$[64, 74)$	t	24	68.43	2.74
	minutes	24	1668.63	509.19
≥ 74	t	23	85.70	7.92
	minutes	23	2276.26	962.75

[a] Note: t is the payment and *minutes* is the number of total phone minutes used.

Table A.3: Results of Counterfactual Experiments

	NLP[a]	ID	AUD	QF	LP
Expected revenue	27.66	27.04	23.19	19.12	18.42
Expected consumer surplus	10.70	9.73	9.31	7.56	7.26
Expected firm profit	10.42	10.09	9.74	8.74	8.56
Expected purchase	86.63	87.82	66.17	49.81	45.36

[a] Note: See Section 1.5 for explanations of each experiment. NLP means nonlinear pricing. ID means incremental discounts. AUD means all-units discounts. QF means quantity forcing. LP means linear pricing.

Appendix B

Multiproduct Nonlinear Pricing

Proofs

Optimal Exclusion: Conditioning on serving customers with aggregate type in $[h_0, \overline{h}]$ and dropping ϵ to simplify notations, $c(h)$ and $T(c)$ satisfy (8) and (9), respectively. We remark that $c(\cdot)$ does not depend on h_0, while $T(\cdot)$ does through the boundary condition (7). Multiplying (8) by $c'(h)$ and integrating by parts the resulting equation from h_0 to h gives

$$T(c(h)) - T(c(h_0)) = \int_{h_0}^{h} zV_0'(c(z))c'(z)dz = hV_0(c(h)) - h_0V_0(c(h_0)) - \int_{h_0}^{h} V_0(c(z))dz.$$

Using (7), this gives

$$T(c(h)) = hV_0(c(h)) - \int_{h_0}^{h} V_0(c(z))dz.$$

Thus, the provider's expected profit is

$$\int_{h_0}^{h} [T(c(z)) - c(z)]\phi(z)dz = \int_{h_0}^{h} \left\{ \left[zV_0(c(z)) - \int_{h_0}^{z} V_0(c(x))dx \right] - c(z) \right\} \phi(z)dz.$$

Its first derivative with respect to h_0 is

$$
\begin{aligned}
&- \left[h_0 V_0(c(h_0)) - c(h_0)\right] \phi(h_0) + \int_{h_0}^{h} V(c(h_0))\phi(z)dz \\
&= -\phi(h_0) \left\{ \left[h_0 - \frac{1 - \Phi(h_0)}{\phi(h_0)} \right] V_0(c(h_0)) - c(h_0) \right\},
\end{aligned}
$$

which gives the boundary condition (6).

Proof of Proposition 3: We first prove that γ is identified. From the one-to-one mapping between h and t, we have $\Phi(h(t)) = G_t(t)$, which gives $\phi(h(t)) = h'(t)g_t(c)$. Thus,

$$
\frac{1 - \Phi(h)}{\phi(h)} = h'(T^{-1}(t))T^{-1\prime}(t)\frac{1 - G_t(t)}{g_t(t)}
$$

since $h(t) = h(T^{-1}(t))$. Hence, (13) can be written as

$$
\begin{aligned}
\underline{c} &= \left[h_0 - h'(\underline{c})T^{-1\prime}(\underline{t})\frac{1 - G_t(\underline{t})}{g_t(\underline{t})} \right] V_0(\underline{c}) \\
&= \left[1 - \frac{h'(\underline{c})T^{-1\prime}(\underline{t})}{h_0 g_t(\underline{t})} \right] \underline{t},
\end{aligned}
\tag{B.1}
$$

using $\underline{c} = C(h_0) = T^{-1}(\underline{t})$ and $V_0(\underline{c}) = \underline{t}/h_0$ from the boundary condition.

We now compute the derivative $h'(\cdot)$. From (14) and A6, $h'(c) = \xi'(c)$. Using $G_c(C(h_\dagger)) = 1/2$ and taking the derivative of the logarithm of 15 gives

$$
\begin{aligned}
\xi'(c) &= \xi(c) \left[\left(\frac{1}{T'(c)} - 1 \right) \frac{-g_c(c)}{1 - G_c(c)} + \log(2(1 - G_c(c)))\frac{-T''(c)}{T'(c)^2} + \frac{T''(c)}{T'(c)^2}\log(1 - G_c(c)) \right] \\
&= \xi(c) \left[\left(\frac{1}{T'(c)} - 1 \right) \frac{-g_c(c)}{1 - G_c(c)} + \frac{-T''(c)}{T'(c)^2}\log 2 \right].
\end{aligned}
$$

Evaluating this at $\underline{c}$ gives

$$
h'(\underline{c}) = -h_0 \left[\left(\frac{1}{T'(\underline{c})} - 1 \right) g_c(\underline{c}) + \frac{T''(\underline{c})}{T'(\underline{c})^2}\log 2 \right].
$$

Inserting this in (B.1) gives

$$\begin{aligned}
\underline{c} &= \left[1+\left(\left(\frac{1}{T'(\underline{c})}-1\right)g_c(\underline{c})+\frac{T''(\underline{c})}{T'(\underline{c})^2}\log 2\right)\frac{T^{-1\prime}(\underline{t})}{g_t(\underline{t})}\right]\underline{t}\\
&= \left[1+\left(\left(T^{-1\prime}(\underline{t})-1\right)\frac{g_t(\underline{t})}{T^{-1\prime}(\underline{t})}-\frac{T^{-1\prime\prime}(\underline{t})}{T^{-1\prime}(\underline{t})}\log 2\right)\frac{T^{-1\prime}(\underline{t})}{g_t(\underline{t})}\right]\underline{t}\\
&= \left[T^{-1\prime}(\underline{t})-\frac{T^{-1\prime\prime}(\underline{t})}{g_t(\underline{t})}\log 2\right]\underline{t} \qquad \text{(B.2)}
\end{aligned}$$

using $g_c(\underline{c}) = g_t(\underline{t})/T^{-1\prime}(\underline{t})$, $T^{-1\prime}(\underline{t}) = 1/T'(\underline{c})$ and $T^{-1\prime\prime}(\underline{t}) = -(T''(\underline{c})/T'(\underline{c})^2)T^{-1\prime}(\underline{t})$. Since $\log T^{-1}(t) = \log c = \log\kappa + \gamma\log Q_v(t) + (1-\gamma)\log Q_m(t)$, we obtain by taking derivatives

$$\begin{aligned}
T^{-1\prime}(t) &= T^{-1}(t)\left(\gamma\frac{d\log Q_v(t)}{dt}+(1-\gamma)\frac{d\log Q_m(t)}{dt}\right),\\
T^{-1\prime\prime}(t) &= T^{-1\prime}(t)\left(\gamma\frac{d\log Q_v(t)}{dt}+(1-\gamma)\frac{d\log Q_m(t)}{dt}\right)\\
&\quad +T^{-1}(t)\left(\gamma\frac{d^2\log Q_v(t)}{dt^2}+(1-\gamma)\frac{d^2\log Q_m(t)}{dt^2}\right),\\
&= T^{-1}(t)\left[\left(\gamma\frac{d\log Q_v(t)}{dt}+(1-\gamma)\frac{d\log Q_m(t)}{dt}\right)^2+\gamma\frac{d^2\log Q_v(t)}{dt^2}+(1-\gamma)\frac{d^2\log Q_m(t)}{dt^2}\right].
\end{aligned}$$

Evaluating these derivatives at $\underline{t}$ and inserting them in (B.2) give (16). This equation is quadratic in γ. For it to have a unique solution in $(0,1)$, it is necessary and sufficient that the quadratic form evaluated at 0 and 1 be non zero and of opposite sign. Namely,

$$\begin{aligned}
&\left[\frac{d\log Q_m(\underline{t})}{dt}-\frac{\log 2}{g_t(\underline{t})}\left(\left(\frac{d\log Q_m(\underline{t})}{dt}\right)^2+\frac{d^2\log Q_m(\underline{t})}{dt^2}\right)-\frac{1}{\underline{t}}\right]\\
&\times\left[\frac{d\log Q_v(\underline{t})}{dt}-\frac{\log 2}{g_t(\underline{t})}\left(\left(\frac{d\log Q_v(\underline{t})}{dt}\right)^2+\frac{d^2\log Q_v(\underline{t})}{dt^2}\right)-\frac{1}{\underline{t}}\right]<0.
\end{aligned}$$

Noting that $d\log Q_m(t)/dt = Q_m'(t)/Q_m(t)$ and $d^2\log Q_m(t)/dt^2 = (Q_m''(t)/Q_m(t)) - (Q_m'(t)/Q_m(t))^2$ with $Q_m(\cdot) > 0$ and similarly for $\log Q_v(t)$ give the necessary and sifficient condition of Proposition 3. This identifies γ.

To identify κ, evaluating (11) and (12) at $\overline{h}$ and $\overline{c}$ gives $T'(\overline{c}) = 1$. Thus, taking the derivative of $t = T(\kappa Q_v(t)^\gamma Q_m(t)^{1-\gamma})$ with respect to t and evaluating it at $\overline{t}$ gives (17), where we have used $T'(\overline{c}) = 1$. This equation identifies κ given that γ is identified. To identify $T(\cdot)$, we know that $t = T(\kappa Q_v(t)^\gamma Q_m(t)^{1-\gamma})$ or equivalently $\log\kappa + \gamma \log Q_v(t) + (1-\gamma)\log Q)_m(t) = \log T^{-1}(t)$. Since $Q_v(\cdot)$ and $Q_m(\cdot)$ are identified by Proposition 2, $T^{-1}(\cdot)$ is identified and given by (18).

Aggregation of Voice Consumption

Our data provide the quantities q^L, q^D and q^R of phone calls measured in minutes and the quantity q_m of SMS measured in units. See Table 1. Since some of them can be zero, throughout the paper we add one unit to every quantity. We aggregate these three quantities into a single one $q_v = A(q^L, q^D, q^R)$ to capture voice consumption. We follow Luo (2011) and we make the following assumption on the aggregation function.

Assumption B1: *The aggregation function is of the form* $A(q^L, q^D, q^R) = (q^L)^{\alpha_1}(q^D)^{\alpha_2}(q^R)^{\alpha_3}$, *with* $\alpha_1, \alpha_2, \alpha_3 \geq 0$ *and* $\sum_{j=1}^3 \alpha_j = 1$.

Assumption B1 corresponds to a standard Cobb-Douglas specification. Using A4, the cost-based tariff becomes

$$t = T\left[\{(q^L)^{\alpha_1}(q^D)^{\alpha_2}(q^R)^{\alpha_3}\epsilon_v\}^\gamma \{(q_m)\epsilon_m\}^{1-\gamma}\right],$$

where $T(\cdot)$ is strictly increasing and concave. Considering the inverse and taking the logarithm gives

$$\gamma\left[\alpha_1 \log q^L + \alpha_2 \log q^D + \alpha_3 \log q^R + \log\epsilon_v\right] + (1-\gamma)\left[\log q_m + \log\epsilon_m\right] = \log[T^{-1}(t)],$$

where $\alpha_3 = 1 - \alpha_1 - \alpha_2$.

Luo (2011) shows that the coefficients of α_1, α_2 and α_3 as well as the function $T^{-1}(\cdot)$ are identified. A more general specification is studied in Luo and Xu (in progress). We denote $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ and $q = (\log q^L, \log q^D, \log q^R)'$. The above equation can be written as

$$\alpha' q = r(t;\gamma) - \log\epsilon_v, \tag{B.3}$$

where $r(t;\gamma) = (1/\gamma)\log[T^{-1}(t)/Q_m^{1-\gamma}(t)]$ since $Q_m = q_m\epsilon_m = Q_m(t)$ from Section 4.1.

Hence, $\mathrm{Var}[\alpha' q|t] = \mathrm{Var}[\log \epsilon_v|t]$, where the latter is a constant from A2-(ii) and (3). To estimate α, we partition the range of bills into B bins and define the dummy variable $D_b(t) = 1$ if t belongs to the bin b and zero otherwise. Thus, for any value b, multiplying both sides of (B.3) by $D_b(t)$ and taking the variance conditional on t gives $\mathrm{Var}[D_b(t)\alpha' q|t] = D_b(t)\mathrm{Var}[\log \epsilon_v|t]$. Thus we have the estimating equations

$$\mathrm{Var}[D_b(t)\alpha' q|t] - D_b(t)\mathrm{Var}[\alpha' q|t] = 0$$

for every $b \in \{1, \ldots, B\}$ and $t \in [\underline{t}, \overline{t}]$. Taking expectation, using $D_b^2(t) = D_b(t)$ and $\mathrm{Var}[\alpha' q|t]$ is a constant gives the moment equations

$$\begin{aligned}
0 &= \alpha' \Big[\mathrm{E}\left\{\mathrm{Var}[D_b(t)q|t]\right\} - \mathrm{E}[D_b(t)]\mathrm{E}(\mathrm{Var}[q|t])\Big]\alpha \\
&= \alpha' \Big[\mathrm{E}\Big\{\mathrm{E}\big[D_b(t)(q-\mathrm{E}(q|t))(q-\mathrm{E}(q|t))'|t\big]\Big\} - \mathrm{E}[D_b(t)]\mathrm{E}\Big\{\mathrm{E}\big[(q-\mathrm{E}(q|t))(q-\mathrm{E}(q|t))'\big]|t\Big\}\Big]\alpha \\
&= \alpha' \Big[\mathrm{E}\Big\{D_b(t)(q-\mathrm{E}(q|t))(q-\mathrm{E}(q|t))'\Big\} - \mathrm{E}[D_b(t)]\mathrm{E}\Big\{(q-\mathrm{E}(q|t))(q-\mathrm{E}(q|t))'\Big\}\Big]\alpha \qquad \text{(B.4)}
\end{aligned}$$

for $b \in \{1, \ldots, B\}$.

Replacing $\mathrm{E}[q|t]$ by a nonparametric regression estimator and $\mathrm{E}[D_b(t)]$ by N_b/N, the sample analog of the right-hand side of (B.4) is

$$\alpha' \left[\frac{1}{N}\sum_{i\in b}(q-\hat{\mathrm{E}}(q|t_i))(q-\hat{\mathrm{E}}(q|t_i))'\Big\} - \frac{N_b}{N}\frac{1}{N}\sum_{i=1}^{N}(q_i-\hat{\mathrm{E}}(q|t_i))(q_i-\hat{\mathrm{E}}(q|t_i))'\right]\alpha \qquad \text{(B.5)}$$

where N_b is the number of observations in bin b and N is the total number of observations. Our estimator of α minimizes the sum over b of the the square of (B.5) weighted by $(N/N_b)^2$, namely

$$\min_{\alpha:\alpha_1+\alpha_2+\alpha_3=1} \sum_{b=1}^{B}\left(\alpha' \left[\frac{1}{N_b}\sum_{i\in b}(q-\hat{\mathrm{E}}(q|t_i))(q-\hat{\mathrm{E}}(q|t_i))'\Big\} - \frac{1}{N}\sum_{i=1}^{N}(q_i-\hat{\mathrm{E}}(q|t_i))(q_i-\hat{\mathrm{E}}(q|t_i))'\right]\alpha\right)^2,$$

The conditional expectation of q given t_i is estimated using a standard kernel estimator

$$\hat{\mathrm{E}}[q|t_i] = \frac{\sum_{k=1}^{N} q_k K\left(\frac{t_i - t_k}{h_t}\right)}{\sum_{k=1}^{N} K\left(\frac{t_i - t_k}{h_t}\right)},$$

where $K(\cdot)$ is a kernel function and h_t is a bandwidth. Using 19 bins, a triweight kernel and the rule of thumb bandwidth, the estimated coefficients are $\hat{\alpha}_1 = 0.4295$, $\hat{\alpha}_2 = 0.3199$, $\hat{\alpha}_3 = 0.2505$. Thus, $q_{vi} = (q_i^L)^{0.4295}(q_i^D)^{0.3199}(q_i^R)^{0.2505}$ for $i = 1, \ldots, N$.

Figures and Tables

Figure B.1: Scatter plot (q_v, t)

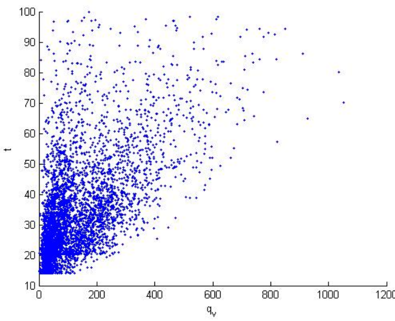


Figure B.2: Scatter plot (q_m, t)

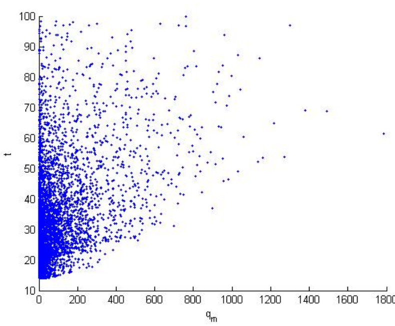


Figure B.3: Scatter plot (q_v, q_m)

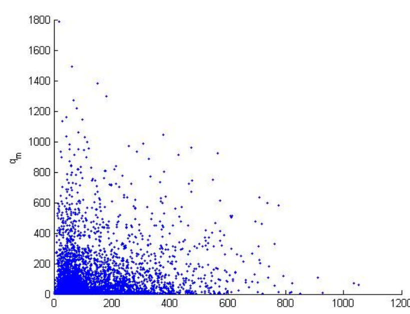


Figure B.4: Fitted Regression of the Tariff

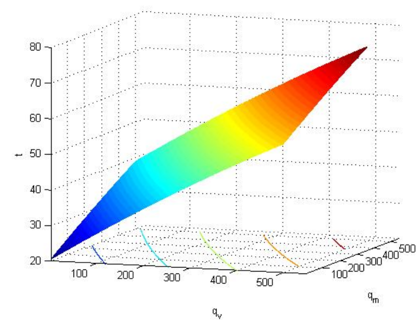


Figure B.5: $\hat{h}(c)$ and $\hat{V}_0'(c)$

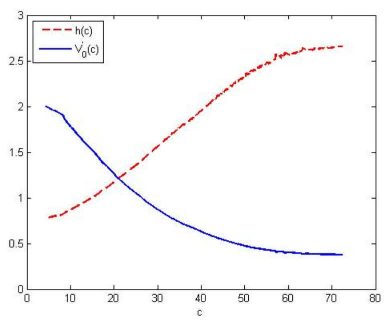


Figure B.6: $\hat{T}(c)$

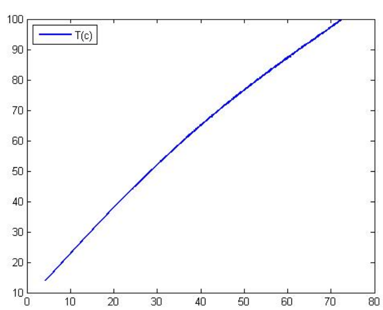


Figure B.7: $\hat{\phi}^*(h)$

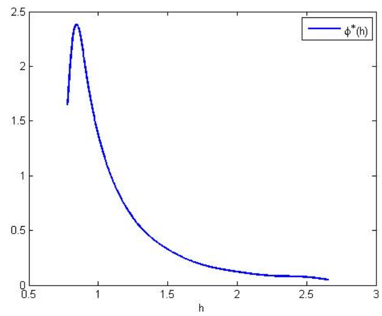


Figure B.8: $1 - \frac{\hat{\Phi}^*(h)}{h\hat{\phi}^*(h)}$

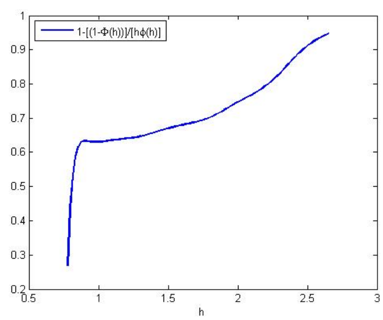


Figure B.9: $\hat{V}_0(t)$, $\hat{h}(t)\hat{V}_0(t)$

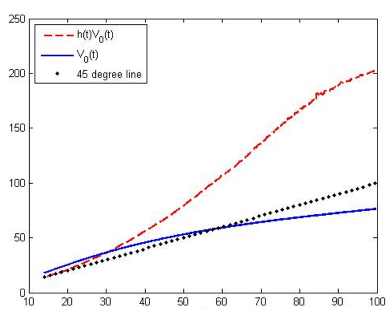


Figure B.10: Density of informational rent by payment

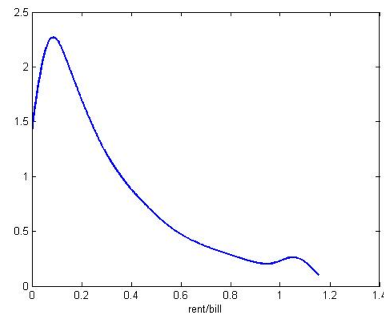


Figure B.11: Densities of unobserved heterogeneity $\hat{f}_{\epsilon_v}(\cdot)$, $\hat{f}_{\epsilon_m}(\cdot)$

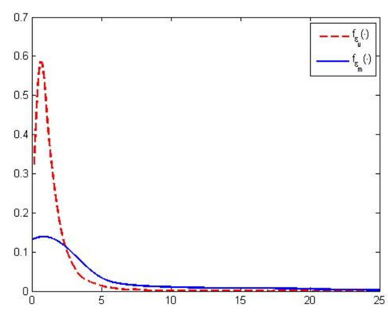


Figure B.12: Cost-based tariff

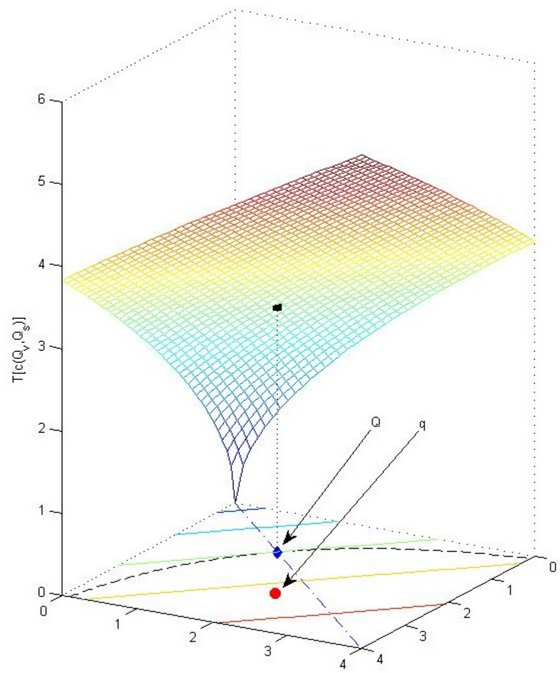


Table B.1: Summary statistics

	Mean	Median	Min	Max	STD
Bill	34.65	30.34	13.97	99.78	16.56
q^L	488.15	420	0	2,673	331.61
q^D	145.41	60	0	3,095	237.68
q^R	160.28	28	0	3,179	288.25
q^{SMS}	100.87	38	0	1,786	161.35
q_v	135.03	86.12	2.24	1,051.07	129.90

N.B.: q^L, q^D and q^R are the quantities of local, distance and roaming calls.

Table B.2: Correlations

Bill Range	Variable	Sample Size	Mean	STD	Correlation Matrix Bill	q_v	q_m
< 24	Bill	1,446	19.57	2.93	1		
	q_v	–	62.56	49.93	0.321	1	
	q_m	–	37.27	51.99	0.207	0.043	1
$[24, 34)$	Bill	1,317	28.77	2.98	1		
	q_v	–	106.50	78.29	0.172	1	
	q_m	–	83.55	108.64	0.074	-0.119	1
$[34, 44)$	Bill	790	38.59	2.76	1		
	q_v	–	156.28	103.76	0.117	1	
	q_m	–	113.69	144.44	0.102	-0.157	1
$[44, 54)$	Bill	479	48.65	2.82	1		
	q_v	–	205.12	140.11	0.090	1	
	q_m	–	162.17	213.50	0.089	-0.231	1
$[54, 64)$	Bill	246	58.47	2.97	1		
	q_v	–	251.41	155.68	0.100	1	
	q_m	–	210.66	253.68	0.133	-0.248	1
$[64, 74)$	Bill	167	68.47	2.89	1		
	q_v	–	308.82	203.85	0.002	1	
	q_m	–	240.27	280.35	0.049	-0.083	1
≥ 74	Bill	166	84.23	7.25	1		
	q_v	–	342.03	227.52	0.068	1	
	q_m	–	279.13	289.87	-0.031	-0.239	1
Full sample	Bill	4,611	34.65	16.56	1		
	q_v	–	135.03	129.90	0.595	1	
	q_m	–	100.87	161.35	0.408	0.112	1

Table B.3: Informational Rents

	Mean	Median	Min	Max	STD
Rents	15.17	6.55	0.00	102.84	21.02
Rent/Bill	0.31	0.22	0.00	1.15	0.28

Table B.4: Unobserved Heterogeneity

	Mean	Median	Min	Max	STD
ϵ_v	1.40	0.97	0.17	45.33	1.62
ϵ_m	4.47	0.74	0.04	166.30	10.44

Appendix C

Bundling and Nonlinear Pricing in Telecommunications

Proofs

Proof of Lemma 1: Due to minimum internet need, $j(\theta,\beta) \geq D(\beta)$ at equilibrium. I establish that any mechanism in which $j(\theta,\beta) > D(\beta)$ for some (θ,β) is dominated by a mechanism in which $j(\theta,\beta) = D(\beta)$ for all (θ,β). Fix any mechanism $\{t(\cdot,\cdot), q(\cdot,\cdot), j(\cdot,\cdot)\}$ satisfying the IC, IR and MN constraints. Suppose that there exists some (θ,β) such that $j(\theta,\beta) > D(\beta)$. I now consider a mechanism $\left\{\tilde{t}(\cdot,\cdot), \tilde{q}(\cdot,\cdot), \tilde{j}(\cdot,\cdot)\right\}$ where $\tilde{t}(\theta,\beta) = t(\theta,\beta) + U(\tilde{q}(\theta,\beta), \tilde{j}(\theta,\beta);\theta) - U(q(\theta,\beta), j(\theta,\beta);\theta)$, $\tilde{q}(\theta,\beta) = q(\theta,\beta)$, and $\tilde{j}(\theta,\beta) = D(\beta)$. The original mechanism is less profitable than the new one because of Assumption 3-(ii).

By definition, the consumer surplus keeps the same. Thus, the IR constraints hold under the new mechanism. I need to show that the IC constraints hold under the new mechanism. Consider the following inequality:

$$U(\tilde{q}(\theta,\beta), \tilde{j}(\theta,\beta);\theta) - \tilde{t}(\theta,\beta) \geq U(\tilde{q}(\tilde{\theta},\tilde{\beta}), \tilde{j}(\tilde{\theta},\tilde{\beta});\theta) - \tilde{t}(\tilde{\theta},\tilde{\beta}),$$

which is equivalent to

$$\begin{aligned} &U(\tilde{q}(\theta,\beta), \tilde{j}(\theta,\beta);\theta) - \Big[t(\theta,\beta) + U(\tilde{q}(\theta,\beta), \tilde{j}(\theta,\beta);\theta) - U(q(\theta,\beta), j(\theta,\beta);\theta)\Big] \\ &\geq U(\tilde{q}(\tilde{\theta},\tilde{\beta}), \tilde{j}(\tilde{\theta},\tilde{\beta});\theta) - \Big[t(\tilde{\theta},\tilde{\beta}) + U(\tilde{q}(\tilde{\theta},\tilde{\beta}), \tilde{j}(\tilde{\theta},\tilde{\beta});\tilde{\theta}) - U(q(\tilde{\theta},\tilde{\beta}), j(\tilde{\theta},\tilde{\beta});\tilde{\theta})\Big]. \end{aligned}$$

The left-hand side equals $U(q(\theta,\beta),j(\theta,\beta);\theta)-t(\theta,\beta)$, whereas the right-hand side equals $U(q(\tilde{\theta},\tilde{\beta}),j(\tilde{\theta},\tilde{\beta});\theta)-t(\tilde{\theta},\tilde{\beta})$ if Assumption 3-(i) is satisfied. In fact,

$$
\begin{aligned}
&\Big\{U(\tilde{q}(\tilde{\theta},\tilde{\beta}),\tilde{j}(\tilde{\theta},\tilde{\beta});\theta)-\big[t(\tilde{\theta},\tilde{\beta})+U(\tilde{q}(\tilde{\theta},\tilde{\beta}),\tilde{j}(\tilde{\theta},\tilde{\beta});\tilde{\theta})-U(q(\tilde{\theta},\tilde{\beta}),j(\tilde{\theta},\tilde{\beta});\tilde{\theta})\big]\Big\}-\\
&\Big\{U(q(\tilde{\theta},\tilde{\beta}),j(\tilde{\theta},\tilde{\beta});\theta)-t(\tilde{\theta},\tilde{\beta})\Big\}\\
&=U(\tilde{q}(\tilde{\theta},\tilde{\beta}),\tilde{j}(\tilde{\theta},\tilde{\beta});\theta)-U(\tilde{q}(\tilde{\theta},\tilde{\beta}),\tilde{j}(\tilde{\theta},\tilde{\beta});\tilde{\theta})+U(q(\tilde{\theta},\tilde{\beta}),j(\tilde{\theta},\tilde{\beta});\tilde{\theta})-U(q(\tilde{\theta},\tilde{\beta}),j(\tilde{\theta},\tilde{\beta});\theta)=0,
\end{aligned}
$$

where the last equation is true if Assumption 3-(i) is satisfied.

The IC constraints hold under the new mechanism because they do under the original one. Thus $j^*(\theta,\beta)=D(\beta)$. I now turn to $q^*(\cdot,\cdot)$ and $t^*(\cdot,\cdot)$. Given $j^*(\theta,\beta)=D(\beta)$, the consumer's problem becomes $\max_{q\in I\!R^+} U(q,D(\beta);\theta)-T^*(q,D(\beta))$. This implies that β only affects phone usage through $D(\beta)$. Hence, $q^*(\theta,\beta)=q^*(\theta,D(\beta))$. Given that I consider non-random nonlinear pricing schedules, it further implies that $t^*(\theta,\beta)=T^*(q^*(\theta,D(\beta)),D(\beta))=t^*(\theta,D(\beta))$. □

Proof of Lemma 2: For a given cutoff taste θ^c, the optimal mechanism $\{q^*(\cdot,\cdot;\theta^c),j^*(\cdot,\cdot;\theta^c),t^*(\cdot,\cdot;\theta^c)\}$ can be derived following Sundararajan (2004) and is defined by (2), $j^*(\theta,\beta;\theta^c)=D(\beta)$, and (3) by replacing θ^c_j with θ^c. An important feature is that the allocation $q^*(\theta,\beta;\theta^c),j^*(\theta,\beta;\theta^c)$ does not depend on θ^c, while the optimal price schedule $t^*(\cdot,\cdot;\theta^c)$ does. The provider's problem is then to find an optimal θ^c to maximize its expected profit

$$
\begin{aligned}
&\int_{\theta^c}^{\overline{\theta}}\big[t^*(\theta,\beta;\theta^c)-c(q^*(\theta,\beta),D(\beta))\big]f(\theta|D(\beta))d\theta+\int_{\underline{\theta}}^{\theta^c}\big[v(0,D(\beta))-c(0,D(\beta))\big]f(\theta|D(\beta))d\theta,\\
&=\int_{\theta^c}^{\overline{\theta}}\Big[U(q^*(\theta,\beta),D(\beta);\theta)-\int_{\theta^c}^{\theta}U_\theta(q^*(z,\beta),D(\beta);z)dz-c(q^*(\theta,\beta),D(\beta))\Big]f(\theta|D(\beta))d\theta\\
&\quad+\big[v(0,D(\beta))-c(0,D(\beta))\big]F(\theta^c|D(\beta)),
\end{aligned}
$$

where the first part is the profit collected from all consumers buying internet and phone services and the second part is the profit collected from consumers buying only internet. The first-order derivative with respect to θ^c is

$$
\begin{aligned}
&-\big[U(q^*(\theta^c,\beta),D(\beta);\theta^c)-c(q^*(\theta^c,\beta),D(\beta))\big]f(\theta^c|D(\beta))+\int_{\theta^c}^{\overline{\theta}}U_\theta(q^*(\theta^c,\beta),D(\beta);\theta^c)f(\theta|D(\beta))d\theta\\
&+\big[v(0,D(\beta))-c(0,D(\beta))\big]f(\theta^c|D(\beta))\\
&=-f(\theta^c|D(\beta))M(\theta^c,D(\beta)).
\end{aligned}
$$

which gives the boundary condition (1). $\square$

Proof of Lemma 4: If the two-dimensional IC constraints hold, the two one-dimensional constraints hold automatically. I now establish that, if the two one-dimensional IC constraints hold, the two-dimensional IC constraints hold as well. To see this, consider any two pairs (θ, β) and $(\tilde{\theta}, \tilde{\beta})$, such that $D(\tilde{\beta}) \geq \beta$. Consider (θ, β) and $(\theta, \tilde{\beta})$. The second one-dimensional IC constraint at (θ, β) implies

$$U(q(\theta, \beta), D(\beta); \theta) - t(\theta, \beta) \geq U(q(\theta, \tilde{\beta}), D(\tilde{\beta}); \theta) - t(\theta, \tilde{\beta}). \tag{C.1}$$

Now consider $(\theta, \tilde{\beta})$ and $(\tilde{\theta}, \tilde{\beta})$. The first one-dimensional IC constraint at $(\theta, \tilde{\beta})$ implies

$$U(q(\theta, \tilde{\beta}), D(\tilde{\beta}); \theta) - t(\theta, \tilde{\beta}) \geq U(q(\tilde{\theta}, \tilde{\beta}), D(\tilde{\beta}); \theta) - t(\tilde{\theta}, \tilde{\beta}). \tag{C.2}$$

Combining (C.1) and (C.2) gives $U(q(\theta, \beta), D(\beta); \theta) - t(\theta, \beta) \geq U(q(\tilde{\theta}, \tilde{\beta}), D(\tilde{\beta}); \theta) - t(\tilde{\theta}, \tilde{\beta})$. Therefore, the two-dimensional IC constraints are satisfied. $\square$

Proof of Lemma 5: First, I show that $q^*(\theta, \beta)$ is decreasing in β. Since $q^*(\theta, \beta) = q^*(\theta, D(\beta))$, without loss of generality I show that $\frac{\partial q^*(\theta, j)}{\partial j} \leq 0$.[1] To simplify the exposition, I suppress the arguments of functions and omit the asterisk superscript hereafter.

Taking the total derivative of (2) with respect to j gives

$$\frac{\partial q(\theta, j)}{\partial j} = \frac{-U_{qj} + c_{qj} + U_{q\theta j}\frac{1-F}{f} + U_{q\theta}\frac{\partial \frac{1-F}{f}}{\partial j}}{U_{qq} - c_{qq} - U_{qq\theta}\frac{1-F}{f}} \leq 0,$$

where the inequality holds since $c_{qj} \geq U_{qj}$ (following Assumption 3-(ii)), $U_{q\theta j} = 0$ (following Assumption 3-(i)), $U_{q\theta} > 0$ (following Assumption 2-(ii)), $\frac{\partial \frac{1-F}{f}}{\partial j} \geq 0$ (following Assumption 4). The denominator is negative under Assumption 2. See Sundararajan (2004) for details.

Second, I show that θ^c_j is increasing in j. Suppose the contrary, i.e. $\theta^c_j > \theta^c_{\tilde{j}}$ for some $j, \tilde{j} \in \mathcal{J}$. I show that this leads to a contradiction in all cases.

(i) $\underline{\theta} < \theta^c_j < \overline{\theta}$ and $\underline{\theta} < \theta^c_{\tilde{j}} < \overline{\theta}$. By definition, for all $j \in \mathcal{J}$, $M(\theta^c_j, j) = 0$ if $\underline{\theta} < \theta^c_j < \overline{\theta}$.

[1]For notation convenience, I use differentiation as if the variable j is continuous.

Computing the total differential of $M(\theta^c_j, j) = 0$ with respect to j gives

$$\begin{aligned}\frac{\partial \theta^c_j}{\partial j} &= -\frac{[U_q - U_{q\theta}\frac{1-F}{f} - c_q]\frac{\partial q}{\partial j} + \big[U_j - v_j(0,j)\big] - \big[c_j - c_j(0,j)\big] - U_{\theta j}\frac{1-F}{f} - U_\theta \frac{\partial \frac{1-F}{f}}{\partial j}}{[U_q - U_{q\theta}\frac{1-F}{f} - c_q]\frac{\partial q}{\partial \theta} + U_\theta - U_\theta \frac{\partial \frac{1-F}{f}}{\partial \theta} - U_{\theta\theta}\frac{1-F}{f}} \\ &= -\frac{\big[U_j - v_j(0,j)\big] - \big[c_j - c_j(0,j)\big] - U_\theta \frac{\partial \frac{1-F}{f}}{\partial j}}{U_\theta(1 - \frac{\partial \frac{1-F}{f}}{\partial \theta}) - U_{\theta\theta}\frac{1-F}{f}} \geq 0,\end{aligned}$$

leading to a contradiction as $\theta^c_j > \theta^c_{\tilde{j}}$ by assumption. The second equality holds since $U_q - U_{q\theta}\frac{1-F}{f} - c_q = 0$ (following Equation (2)), $U_{\theta j} = 0$ (following Assumption 3-(i)), while the inequality holds since $U_\theta \geq 0$ (following Assumption 2-(i)), $U_{\theta\theta} \leq 0$ (following Assumption 2-(i)), $H_\theta \geq 0$ (following Assumption 2-(v)), $\frac{\partial \frac{1-F}{f}}{\partial j} \geq 0$. Moreover, Assumption 3-(ii) implies $[U(q, \tilde{j}; \theta) - v(0, \tilde{j})] - [U(q, j; \theta) - v(0, j)] \leq [c(q, \tilde{j}) - c(0, \tilde{j})] - [c(q, j) - c(0, j)]$.

(ii) $\underline{\theta} < \theta^c_j \leq \overline{\theta}$ and $\theta^c_{\tilde{j}} = \underline{\theta}$. By definition, $M(\underline{\theta}, \tilde{j}) \geq 0$. Differentiating $M(\underline{\theta}, j)$ with respect to j gives

$$\begin{aligned}\frac{\partial M(\underline{\theta}, j)}{\partial j} =& [U_q - U_{q\theta}\frac{1}{f(\underline{\theta}|j)} - c_q]\frac{\partial q}{\partial j} + \big[U_j - v_j(0,j)\big] - \big[c_j - c_j(0,j)\big] \\ =& \big[U_j - v_j(0,j)\big] - \big[c_j - c_j(0,j)\big] \leq 0,\end{aligned}$$

which implies that $M(\underline{\theta}, j) \geq M(\underline{\theta}, \tilde{j}) \geq 0$, leading to a contradiction by definition of θ^c_j.

(iii) $\theta^c_j = \overline{\theta}$ and $\underline{\theta} < \theta^c_{\tilde{j}} < \overline{\theta}$. By definition, $M(\theta^c_{\tilde{j}}, \tilde{j}) = 0$. Differentiating $M(\theta^c_{\tilde{j}}, j)$ with respect to j gives

$$\begin{aligned}\frac{\partial M(\theta^c_{\tilde{j}}, j)}{\partial j} =& [U_q - U_{q\theta}\frac{1 - F(\theta^c_{\tilde{j}}|j)}{f(\theta^c_{\tilde{j}}|j)} - c_q]\frac{\partial q}{\partial j} + \big[U_j - v_j(0,j)\big] - \big[c_j - c_j(0,j)\big] \\ =& \big[U_j - v_j(0,j)\big] - \big[c_j - c_j(0,j)\big] \leq 0,\end{aligned}$$

which implies that $M(\theta^c_{\tilde{j}}, j) \geq M(\theta^c_{\tilde{j}}, \tilde{j}) = 0$, leading to a contradiction by definition of θ^c_j.

Third, consider a (θ, j) consumer's indirect utility $S(\theta, j) = \int_{\theta^c_j}^{\theta} U_\theta\big(q(x,j), j; x\big)dx$. The first-order derivative with respect to j is

$$\begin{aligned}\frac{\partial S(\theta, j)}{\partial j} &= -U_\theta\big(q(\theta^c_j, j), j; \theta^c_j\big)\frac{\partial \theta^c_j}{\partial j} + \int_{\theta^c_j}^{\theta} U_{q\theta}\big(q(x,j), j; x\big)\frac{\partial q(x,j)}{\partial j} + U_{j\theta}\big(q(x,j), j; x\big)dx \\ &= -U_\theta\big(q(\theta^c_j, j), j; \theta^c_j\big)\frac{\partial \theta^c_j}{\partial j} + \int_{\theta^c_j}^{\theta} U_{q\theta}\big(q(x,j), j; x\big)\frac{\partial q(x,j)}{\partial j}dx \leq 0,\end{aligned}$$

where the second equality follows from Assumption 3 while the inequality follows from the above results. Therefore, the consumer gets lower indirect utility if he overreport his β.

Finally, I show that $T^*(q,j) - v(q,j)$ is increasing in j. By Lemma 2, we know that

$$T^*(q,j) - v(q,j) = t^*(\theta(q,j),j) - v(q,j) = u(q,\theta(q,j)) - \int_{\theta_j^c}^{\theta(q,j)} u_\theta(q(x,j),x)dx,$$

where $\theta(\cdot,j)$ is the inverse function of $q^*(\cdot,j)$. Differentiating with respect to j gives

$$\begin{aligned}\frac{\partial\big[T^*(q,j) - v(q,j)\big]}{\partial j} =& u_\theta\big(q,\theta(q,j)\big)\frac{\partial\theta(q,j)}{\partial j} - u_\theta\big(q,\theta(q,j)\big)\frac{\partial\theta(q,j)}{\partial j} + \\ & u_\theta\Big(q\big(\theta_j^c,j\big),\theta_j^c\Big)\frac{\partial\theta_j^c}{\partial j} - \int_{\theta_j^c}^{\theta(q,j)} u_{q\theta}\big(q(x,j),x\big)\frac{\partial q(x,j)}{\partial j}dx \\ =& u_\theta\Big(q\big(\theta_j^c,j\big),\theta_j^c\Big)\frac{\partial\theta_j^c}{\partial j} - \int_{\theta_j^c}^{\theta(q,j)} u_{q\theta}\big(q(x,j),x\big)\frac{\partial q(x,j)}{\partial j}dx \geq 0,\end{aligned}$$

where the inequality holds since $u_\theta > 0$, $\frac{\partial\theta_j^c}{\partial j} \geq 0$, $u_{q\theta} > 0$ and $\frac{\partial q(\theta,j)}{\partial j} \leq 0$. $\square$

Proof of Proposition 5: First, I show that $\underline{q}_j > 0$ for all $j = 0,1,2$. Suppose the contrary, i.e. $\underline{q}_j = 0$. Thus $M_j(\theta_j^c) = 0 - u_0[(1-F)/f] - (\kappa_0 - \Delta_j) < \Delta_j - \kappa_0 < 0$, thereby contradicting the definition of θ_j^c. Therefore, $\underline{q}_j > 0$ for all $j = 0,1,2$.

Second, I show that fixed cost parameters $\kappa_0 - \Delta_j$ can be identified for $j = 1,2$. The cutoff consumer receives no information rent. Namely, $t_j(\theta_j^c) = \theta_j^c u_0(q_j^*(\theta_j^c)) + v_j(q_j^*(\theta_j^c))$, which can be written as

$$T_j(\underline{q}_j) = \theta_j^c u_0(\underline{q}_j) + v_j(\underline{q}_j). \tag{C.3}$$

By definition of the cutoff taste, I have

$$\big[\theta_j^c u_0(\underline{q}_j) + v_j(\underline{q}_j) - v_j(0)\big] - u_0(\underline{q}_j)\frac{1-F_j(\theta_j^c)}{f_j(\theta_j^c)} = \big(\kappa_0 + \kappa_j - \Delta_j + \gamma\underline{q}_j\big) - \kappa_j. \tag{C.4}$$

Equations (C.3) and (C.4) imply

$$\begin{aligned}\kappa_0 - \Delta_j &= T_j(\underline{q}_j) - T_j(0) - \gamma\underline{q}_j - u_0(\underline{q}_j)\frac{1-F_j(\theta_j^c)}{f_j(\theta_j^c)} \\ &= T_j(\underline{q}_j) - T_j(0) - \gamma\underline{q}_j - \theta_0(\underline{q}_j)u_0(\underline{q}_j)\frac{T_j'(\underline{q}_j) - \gamma}{T_0'(\underline{q}_j)},\end{aligned} \tag{C.5}$$

where the last equality holds since

$$\frac{1-F_j(\theta_j^c)}{f_j(\theta_j^c)} = \frac{T_j'(\underline{q}_j)-\gamma}{u_0'(\underline{q}_j)} = \theta_0(\underline{q}_j)\frac{T_j'(\underline{q}_j)-\gamma}{T_0'(\underline{q}_j)}.$$

Similarly, if $j=0$, then $\kappa_0 = T_0(\underline{q}_0) - 0 - \gamma\underline{q}_0 - T_0(\underline{q}_0)\frac{T_0'(\underline{q}_0)-\gamma}{T_0'(\underline{q}_0)} = \frac{\gamma}{T_0'(\underline{q}_0)}T_0(\underline{q}_0) - \gamma\underline{q}_0$, where the first equality following from $\theta_0(\underline{q}_0)u_0(\underline{q}_0) = T_0(\underline{q}_0)$. Thus κ_0 is identified, leading to the identification of Δ_1 and Δ_2 by (C.5). $\square$

Estimation of Tariff Functions and Construction of Phone Usage

The data provide the quantity of phone calls Q measured in minutes, the internet speed j measured in Mbps and the payment t measured in RMB. Following Luo (2011), I aggregate phone call minutes, add-ons and additional features into a single index $q = Q \times \epsilon$ to capture phone usage. The term ϵ captures the add-ons and additional features, which are unobserved by the analyst. Thus the tariff for group-j becomes $t = T_j(Q\epsilon)$, where $j \in \mathcal{J}$, and $T_j(\cdot)$ is strictly increasing and concave. Considering the inverse and taking the natural logarithm gives

$$\log Q = \log T_j^{-1}(t) - \log \epsilon. \tag{C.6}$$

Following Luo (2011), I assume that $\epsilon \perp \theta$. The tariff function $T_j(\cdot)$ is identified.

To estimate $T_j(\cdot)$, I approximate its inverse function with splines and find the optimal approximate spline that minimizes the sum of squared errors in (C.6). Since $T_j^{-1}(\cdot)$ is increasing and convex, I use constrained smoothing regression splines proposed by Dole (1999) to approximate it

$$\psi(\cdot;\delta_j) \equiv \sum_{l=1}^{n_j} \delta_j^l s_j^l(\cdot),$$

where δ_j is a vector of parameters δ_j^l, s_j^l is a cubic basis function, and n_j is the number of interior knots. The function $\psi(\cdot;\delta_j)$ is increasing and positive if and only if $\delta_j \geq 0$.

I then solve the following problem:

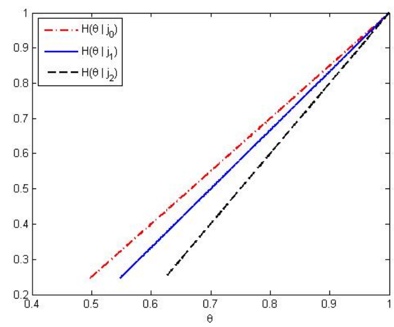
$$\min_{\delta_0,\delta_1,\delta_2 \geq 0} \quad \sum_{j\in\mathcal{J}} \sum_{i=1}^{N_j^*} \left[\log Q_j^i - \log \psi(t_j^i;\delta_j) \right]^2.$$

I estimate $T(\cdot)$ as $\hat{T}(\cdot) = \psi^{-1}(\cdot;\hat{\delta}_j)$. Figure C.3 displays the estimated tariff functions $T_0(\cdot), T_1(\cdot), T_2(\cdot)$. I construct a bundle-$j$ user's phone usage as $q = \hat{T}_j^{-1}(t) = \psi(t;\hat{\delta}_j)$ for all $t \in [\underline{t}, \overline{t}]$. The data on bundle-$j$ users are $\left\{(q_j^i, t_j^i)\right\}_{i=1}^{N_j^*}$ and $\hat{T}_j(\cdot)$.

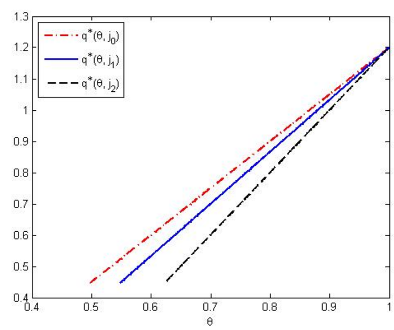
Figures and Tables

Figure C.1: Numerical Example

(a) $\theta - \frac{1-F(\theta|j)}{f(\theta|j)}$



(b) $q^*(\cdot,\cdot)$



(c) $T^*(\cdot,\cdot) - v(\cdot,\cdot)$

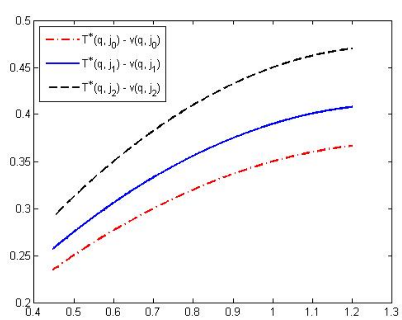
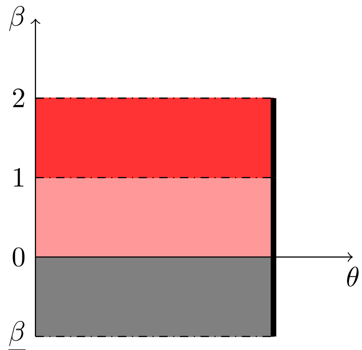
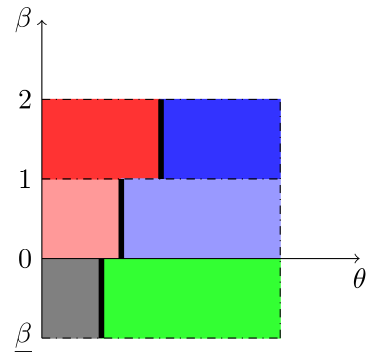


Figure C.2: Numerical Example: Bundling

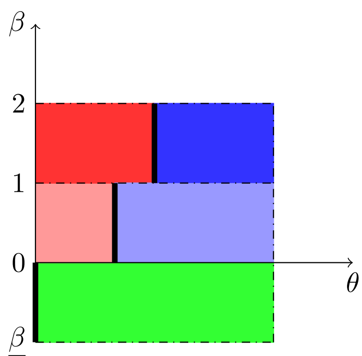
(a) Component Pricing
$\Lambda > 1$



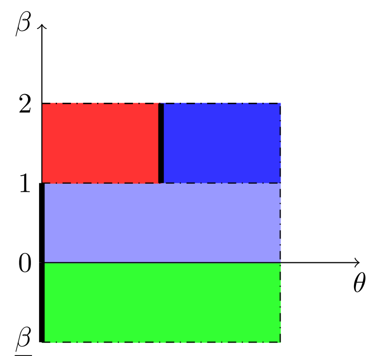
(b) Mixed Bundling
$\Lambda \in (-\frac{1}{b_0}, 1]$



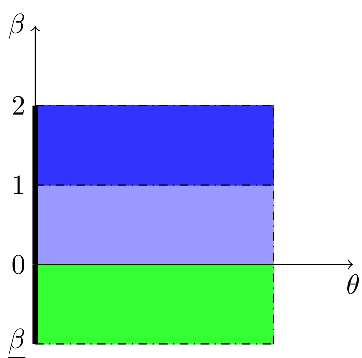
(c) Mixed Bundling
$\Lambda \in (-\frac{1}{b_1}, -\frac{1}{b_0}]$



(d) Semi-mixed Bundling with Exclusion
$\Lambda \in (-\frac{1}{b_2}, -\frac{1}{b_1}]$



(e) Semi-mixed Bundling without Exclusion
$\Lambda \leq -\frac{1}{b_2}$



(f) Pure Bundling
$\Lambda \leq -\frac{1}{b_2}$

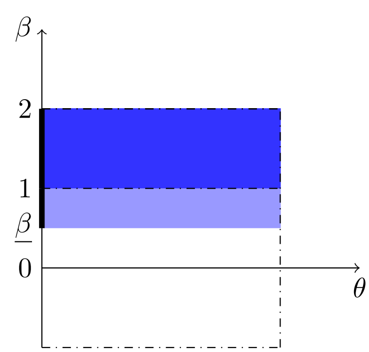


Figure C.3: Tariffs $\hat{T}_0(\cdot)$, $\hat{T}_1(\cdot)$, $\hat{T}_2(\cdot)$

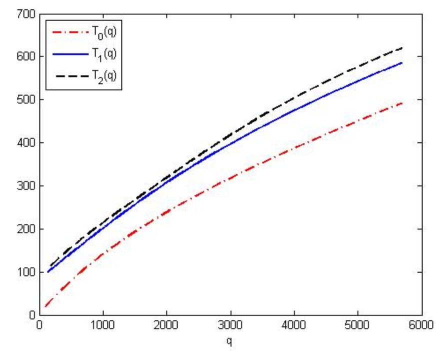


Figure C.4: Marginal Intrinsic Utility $\hat{u}_0'(\cdot)$

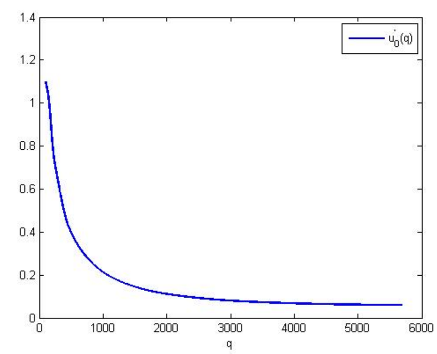


Figure C.5: Marginal Complementary Utilities $\hat{v}_1'(\cdot)$, $\hat{v}_2'(\cdot)$

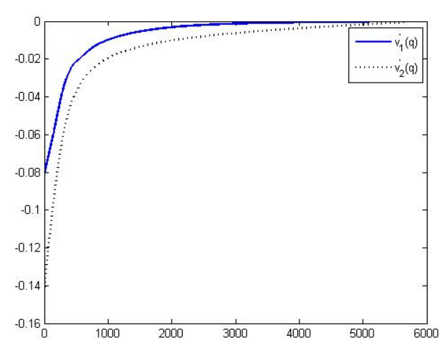


Figure C.6: Phone Service Assignments $\hat{q}_0(\cdot)$, $\hat{q}_1(\cdot)$, $\hat{q}_2(\cdot)$

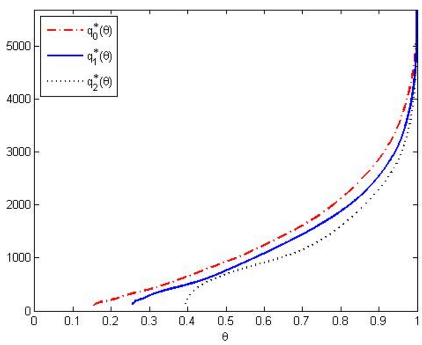


Figure C.7: Conditional Type Densities $\hat{f}_0^*(\cdot)$, $\hat{f}_1^*(\cdot)$, $\hat{f}_2^*(\cdot)$

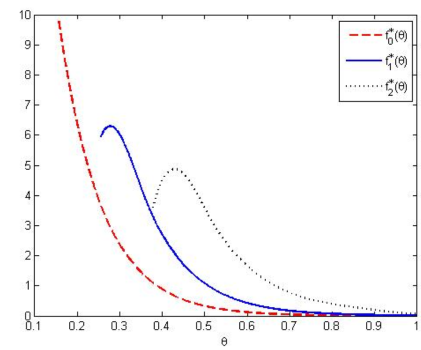


Figure C.8: $\hat{H}_0(\cdot)$, $\hat{H}_1(\cdot)$, $\hat{H}_2(\cdot)$

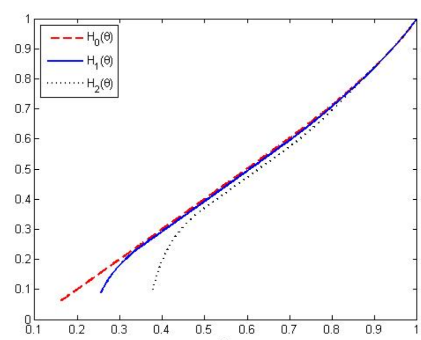


Figure C.9: Example with Mixing Two Groups on the Same Interval

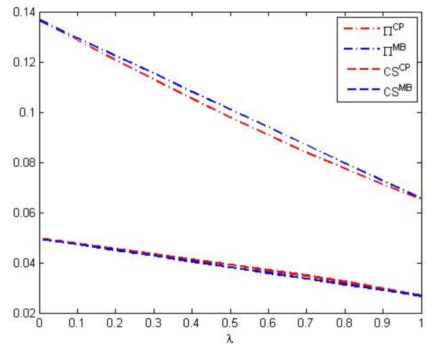


Figure C.10: Tariff Functions under Mixed Bundling and Component Pricing

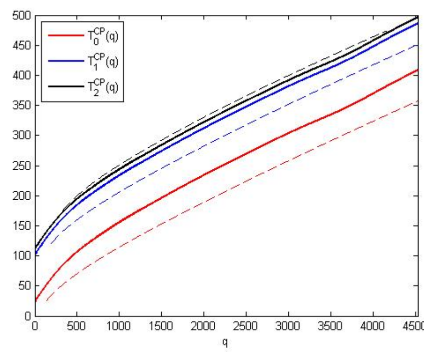


Figure C.11: Breakdown of Welfare

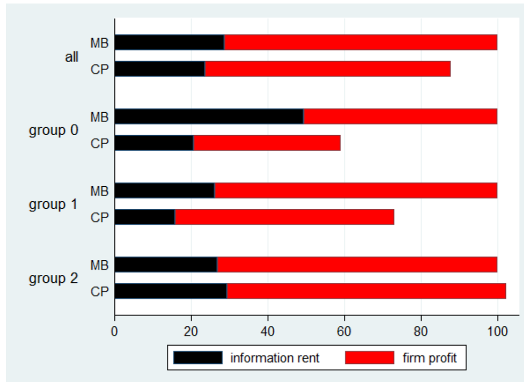


Figure C.12: Breakdown of Bill

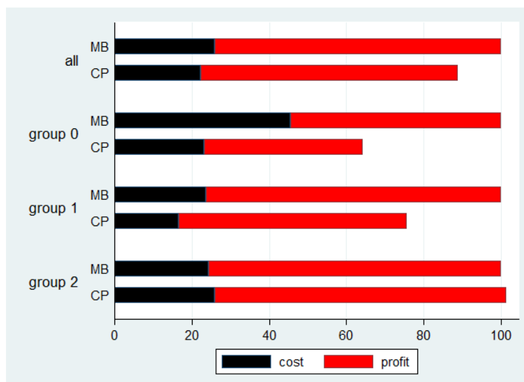


Table C.1: Number of Subscribers as of Dec 31, 2009 (in millions)

	Fixed Line	Broadband	Mobile Service
China Telecom	189	53	56
China Unicom	103	39	145
China Mobile	25	6	522

Table C.2: Summary Statistics

Internet	Variable	N	Mean	S.D.	Min	Max
0	bill	7683	74.91	76.17	19	972.70
	total minutes	–	600.21	720.84	10	4990
	per minute rate	–	0.2129	0.2981	0.0057	9.2333
1	bill	11206	131.06	48.70	99.92	998.84
	total minutes	–	627.88	717.66	10	4956
	per minute rate	–	0.2496	0.4637	0.0060	10.4996
2	bill	12406	176.47	60.85	108	985.99
	total minutes	–	973.13	1045.65	10	4992
	per minute rate	–	0.2727	0.6037	0.0077	23.5360

Table C.3: Regression of the Bill

Internet (Mbps)	0	1	2
bill			
$TotalMin$	0.0859	0.0978	0.1049
	(0.0010)	(0.0013)	(0.0011)
$TotalMin^2$	-3.61e-06	-4.12e-06	-7.01e-06
	(1.04e-07)	(1.61e-07)	(1.40e-07)
$Constant$	81.1750	112.1811	150.8603
	(0.8021)	(0.9854)	(1.1487)
Adjusted R^2	0.2711	0.4186	0.4086

Bibliography

Adams, W. and J. Yellen (1976): "Commodity Bundling and the Burden of Monopoly," *The Quarterly Journal of Economics*, 90, 475–498.

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Armstrong, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75.

——— (1999): "Price Discrimination by a Many-Product Firm," *Review of Economic Studies*, 66, 151–168.

——— (2006): "Recent Developments in the Economics of Price Discrimination," In R. Blundell, W. Newey and T. Persson, eds., *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Volume 2, Cambridge University Press.

Armstrong, M. and J.-C. Rochet (1999): "Multi-Dimensional Screening: A User's Guide," *European Economic Review*, 43, 959–979.

Armstrong, M. and J. Vickers (2010): "Competitive Non-Linear Pricing and Bundling," *The Review of Economic Studies*, 77, 30–60.

Aryal, G. and I. Perrigne (2011): "An Insurance Model with Bidimensional Screening," Working Paper, Pennsylvania State University.

Aryal, G., I. Perrigne, and Q. Vuong (2009): "Identification of Insurance Models with Multidimensional Screening," Working Paper, Pennsylvania State University.

Athey, S. and P. Haile (2007): "Nonparametric approaches to auctions," In J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Volume VI, Amsterdam: North Holland.

Bakos, Y. and E. Brynjolfsson (1999): "Bundling Information Goods: Pricing, Profits, and Efficiency," *Management Science*, 45, 1613–1630.

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

Berry, S. and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48, 1193–1225.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

Blum, J., J. Kiefer, and M. Rosenblatt (1961): "Distribution free tests of independence based on the sample distribution function," *The annals of mathematical statistics*, 485–498.

Busse, M. and M. Rysman (2005): "Competition and Price Discrimination in Yellow Pages Advertising," *RAND Journal of Economics*, 36, 378–390.

Campo, S., E. Guerre, I. Perrigne, and Q. Vuong (2011): "Semiparametric Estimation of First-Price Auctions with Risk-Averse Bidders," *The Review of Economic Studies*, 78, 112.

Che, Y. and I. Gale (2000): "The Optimal Mechanism for Selling to a Budget-Constrained Buyer," *Journal of Economic Theory*, 92, 198–233.

Chen, L. and Y. Luo (2012): "Nonlinear Pricing with Network Effects in Yellow Pages," Working Paper, Pennsylvania State University.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," In J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Volume VI, Amsterdam: North Holland.

Chu, C., P. Leslie, and A. Sorensen (2011): "Bundle-Size Pricing As an Approximation to Mixed Bundling," *American Economic Review*, 101, 263–303.

Cohen, A. (2008): "Package Size and Price Discrimination in the Paper Towel Market," *International Journal of Industrial Organization*, 26, 502–516.

Crawford, G. and M. Shum (2007): "Monopoly Quality Degradation and Regulation in Cable Television," *Journal of Law and Economics*, 50, 181–219.

Crawford, G. S. and A. Yurukoglu (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 643–85.

Debreu, G. (1959): *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, Yale University Press.

D'Haultfoeuille, X. and P. Février (2007): "Identification and estimation of incentive problems: Adverse selection," *Unpublished Manuscript, CREST-INSEE.[1216].*

Dole, D. (1999): "Cosmo: A Constrained Scatterplot Smoother for Estimating Convex, Monotonic Transformations," *Journal of Business & Economic Statistics*, 17, 444–455.

Dubin, J. and D. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

Dunn, K. and K. Singleton (1986): "Modelling the Term Structure of Interest Rates Under Habit Formation and Durability of Goods," *Journal of Financial Economics*, 17, 27–55.

Economides, N., K. Seim, and V. Viard (2008): "Quantifying the Benefits of Entry Into Local Phone Service," *RAND Journal of Economics*, 39, 699–730.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models." *Journal of Political Economy*, 112, 60.

Fang, H. and P. Norman (2006): "To Bundle or Not to Bundle," *Rand Journal of Economics*, 37, 946–963.

Grubb, M. and M. Osborne (2012): "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock," MIT Sloan Research Paper.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-price Auctions," *Econometrica*, 68, 525–574.

——— (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77, 1193–1227.

Hanemann, W. (1984): "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541–561.

Ho, K., J. Ho, and J. H. Mortimer (2012): "The Use of Full-Line Forcing Contracts in the Video Rental Industry," *American Economic Review*, 102, 686–719.

Horowitz, J. L. (1996): "Semiparametric estimation of a regression model with an unknown transformation of the dependent variable," *Econometrica*, 103–137.

Horrigan, J. (2010): "Broadband Adoption and Use in America," OBI Working Paper, Federal Communications Commision.

Ivaldi, M. and D. Martimort (1994): "Competition Under Nonlinear Pricing," *Annales d'Economie et de Statistique*, 71–114.

Kobayashi, B. (2005): "Does Economics Provide a Reliable Guide to Regulating Commodity Bundling by Firms? A Survey of the Economic Literature," *Journal of Competition Law and Economics*, 1, 707–746.

Lavergne, P. and V. Patilea (2010): "Smooth Minimum Distance Estimation and Testing with Conditional Estimating Equations: Uniform in Bandwidth Theory," Working Paper, Toulouse School of Economics.

Lavergne, P. and Q. H. Vuong (1996): "Nonparametric selection of regressors: The nonnested case," *Econometrica: Journal of the Econometric Society*, 207–219.

Leslie, P. (2004): "Price Discrimination in Broadway Theater," *RAND Journal of Economics*, 35, 520–541.

Liu, H., P. Chintagunta, and T. Zhu (2010): "Complementarities and the Demand for Home Broadband Internet Services," *Marketing Science*, 29, 701–720.

Luo, Y. (2011): "Nonlinear Pricing with Product Customization in Mobile Service Industry," Working Paper, NET Institute.

——— (2012): "Bundling and Nonlinear Pricing in Telecommunications," Working Paper, Pennsylvania State University.

——— (in progress): "Nonparametric Estimation of a Function Under Both Shape and Bound Restrictions," .

Luo, Y., I. Perrigne, and Q. Vuong (2012): "Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS," Working Paper, Pennsylvania State University.

Luo, Y. and H. Xu (in progress): "Two-Sided Single-Index Models," .

Maskin, E. and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 171–196.

Mazzeo, M. (2002): "Product Choice and Oligopoly Market Structure," *RAND Journal of Economics*, 33, 221–242.

McAfee, R., J. McMillan, and M. Whinston (1989): "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *The Quarterly Journal of Economics*, 371–383.

McManus, B. (2007): "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee," *The RAND Journal of Economics*, 38, 512–532.

Meyer, M. C. (2008): "Inference using shape-restricted regression splines," *The Annals of Applied Statistics*, 1013–1033.

Miravete, E. (2002): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans," *Review of Economic Studies*, 69, 943–971.

——— (2005): "The Welfare Performance of Sequential Pricing Mechanisms," *International Economic Review*, 46, 1321–1360.

Miravete, E. and L. Röller (2004): "Competitive Nonlinear Pricing in Duopoly Equilibrium: The Early US Cellular Telephone Industry," Working Paper, University of Texas at Austin.

Miravete, E. J. (2007): "The limited gains from complex tariffs," Working Paper, University of Texas at Austin.

Murphy, A. (2007): "A Dynamic Model of Housing Supply," Working Paper, Washington University in St. Louis.

Narayanan, S., P. Chintagunta, and E. Miravete (2007): "The Role of Self Selection, Usage Uncertainty and Learning in the Demand for Local Telephone Service," *Quantitative Marketing and Economics*, 5, 1–34.

Perrigne, I. and Q. Vuong (2011a): "Nonlinear Pricing in Yellow Pages," Working Paper, Pennsylvania State University.

——— (2011b): "Nonparametric identification of a contract model with adverse selection and moral hazard," *Econometrica*, 79, 1499–1539.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric estimation of index coefficients," *Econometrica*, 1403–1430.

Rochet, J.-C. and P. Chone (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, 783–826.

Rochet, J.-C. and L. Stole (2003): "The Economics of Multidimensional Screening," In D. Mathias, L.P. Hansen and S.J. Turnovsk, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Volume 1, Cambridge University Press.

Salinger, M. (1995): "A Graphical Analysis of Bundling," *Journal of Business*, 68, 85–98.

Schmalensee, R. (1981): "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review*, 71, 242–247.

——— (1984): "Gaussian demand and commodity bundling," *Journal of Business*, 211–230.

Schwartz, M. (1990): "Third-Degree Price Discrimination and Output: Generalizing a Welfare Result," *American Economic Review*, 80, 1259–1262.

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics*, 37, 619–640.

Solow, R. (1957): "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39, 312–320.

Stole, L. (2007): "Price Discrimination and Competition," In M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, Volume 3, Amsterdam: North Holland.

Sundararajan, A. (2003): "Network Effects, Nonlinear Pricing and Entry Deterrence," New York University.

——— (2004): "Nonlinear Pricing of Information Goods," *Management Science*, 50, 1660–1673.

Wilson, R. (1997): *Nonlinear Pricing*, Oxford University Press.

Wong, A. C. L. (2012): "Menu Implementability and Practical Pricing Schemes," Working Paper, Shanghai University of Finance and Economics.

Vita

Yao Luo

EDUCATION:

- Ph.D. Economics, The Pennsylvania State University, 2013
- M.A. Economics, Shanghai University of Finance and Economics, 2008
- B.S. Economics and Mathematics, Wuhan University, 2006

FIELDS:

- Primary: Empirical Industrial Organization, Applied Microeconometrics
- Secondary: Microeconomic Theory

GRANTS & FELLOWSHIPS:

- Robert C. Daniels Graduate Fellowship, 2012-2013
- NET Institute Summer Research Grant, 2011
- Bates White Fellowship, 2010 - 2011

TEACHING EXPERIENCE:

- Instructor, Econometrics (UG), summer 2010
- Teaching Assistant, Econometrics (G), 2009-2010; Intermediate Macroeconomics (UG), 2008-2009

RESEARCH EXPERIENCE:

- Research Assistant for Professor Herman J. Bierens, summer 2012
- Research Assistant for Professor Isabelle Perrigne and Professor Quang Vuong, 2011-2012
- Research Assistant for The Center for Research on International Financial and Energy Security, 2010-2011

PRESENTATIONS

- "Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS"
  - Stanford Institute for Theoretical Economics (SITE), Palo Alto, USA, July 2012
  - 8th CIREQ Ph.D. Students' Conference, Montreal, Canada, June 2012
- "Nonlinear Pricing with Product Customization in Mobile Service Industry"
  - Econometric Society North American Summer Meeting, Evanston, USA, June 2012
  - University of Montreal, Montreal, Canada, May 2012
  - JXUFE (Institute of Industrial Economics), Nanchang, China, December 2011