Constrained Monopoly Pricing with Endogenous Participation*

Gabriel Basaluzzo†

Eugenio J. Miravete‡

April 8, 2006

Abstract

We present a flexible model of monopoly nonlinear pricing with endogenous participation decisions of heterogeneous consumers. We make use of the moments that define the few self-selecting tariff options of a monopolist to estimate how demand and cost variables affect the pricing strategies offered by incumbent carriers in several early U.S. local cellular telephone markets. Intuitively, the sources of identification are the position and shape of each tariff option offered by monopolists, their number, as well as a measure of market penetration in each cellular market during the first and last quarter of monopoly phase in this industry. Our results show that from a welfare perspective, uniform (*linear*) pricing dominate fully *nonlinear* tariffs, the optimal *two-part*, *marginal cost-plus fixed fee* (Coasian), and *flat* tariffs. We furthermore address the potential welfare gains of implementing universal service requirements and the consumption distortion originated by each source of asymmetric information.

Keywords: Nonlinear Pricing; Endogenous Participation; Quantity Underprovision; Nonlinear *vs.* Restricted Pricing; Universal Service.

JEL Codes: C63, D43, D82, L96

---

* We thank Toker Doganoglu, Gautam Gowrisankaran, Phil Haile, Bo Honoré, Elena Krasnokutskaya, Thierry Magnac, Charles Manski, Matthew Shum, and Kenneth Wolpin for their comments and suggestions. We also thank participants at the 2004 Kiel-Munich Workshop on the Economics of Information and Network Industries; the 2004 IO Day conference at the NYU-Stern School of Business; the 2004 conference on "The Economics of Electronic Communication Markets at IDEI–Toulouse; the 2005 Workshop on Mathematical Structures in Economics and Econometrics: "Aggregation and Disaggregation: Characterization and Identification of Collective Demand," at the Banff International Research Station for Mathematical Innovation and Discovery; the 2005 Society of Computational Economics Conference on "Computing in Economics and Finance" at George Washington University; The Fundación Ramón Areces Conference on "Industrial Organization and Competition Policy"; as well as seminar audiences at ITAM and universities at Carlos III, Harvard, Northwestern, Pennsylvania, Purdue, Rutgers at New Brunswick, Southern California, Texas A&M, Toronto, and Vanderbilt. The second author acknowledges financial support through NSF Grant SES-0318208. We are solely responsible for all errors that might still remain.

† Centro de Investigación Económica – Instituto Tecnológico Autónomo de México, Camino a Santa Teresa 930, 10700 México DF, MÉXICO. Phone: +52-55-5628-4000, ext. 2919. E-mail: *gbasaluzz@itam.mx*

‡ Department of Economics, University of Pennsylvania, McNeil Building / 3718 Locust Walk, Philadelphia, PA 19104-6297; and CEPR, London, UK. Phone: 215-898-1505. Fax: 215-573-2057. E-mail: *miravete@ssc.upenn.edu*; http://www.ssc.upenn.edu/~miravete

1 Introduction

There are very few contributions in the economics literature that compare the advantages of nonlinear tariffs relative to uniform pricing. According to Stole (2005, §6) this gap is due to the technical complexity of second-degree price discrimination, which impedes to obtain general results that allow us to evaluate the advantages of nonlinear over uniform pricing. However, important legislation such as the Robinson-Patman Act sets restrictions on the ability of firms to price discriminate, although as of today little is known about the welfare consequences of such long standing policies.[1] In the absence of clear theoretical results, the welfare comparisons among different pricing strategies remain ambiguous and this evaluation becomes an important open empirical question that has not been addressed either.[2] Our paper presents a framework where such comparisons as well as other policy evaluations are feasible while using the very limited tariff information generally available to economists.

The first theoretical comparison among different pricing mechanisms that we are aware of is the work of Spence (1977) who considers the problem of distributing a given amount of output among heterogeneous consumers (capacity pricing without exclusion). He proves the now well-known result that relative to uniform pricing, high valuation consumers purchase more and low valuation consumers purchase less with a nonlinear tariff. Spence also shows that the possibility of introducing discounts increases the monopolist's revenues but nothing is said about relative net welfare effects, neither the possibility of some consumers not purchasing at all is ever contemplated in his analysis. The work of Roberts (1979) first addresses the possibility of optimal exclusion of consumers in a model where consumer types include a vertical valuation dimension as well as income effects. However, no comparison of different pricing strategies are made either.[3] Finally, Katz (1983) presents conditions where an increase in output associated to a change in a pricing strategy is a sufficient indicator for welfare improvement.

In all these models, a monopolist attempts to maximize profits by means of a nonlinear tariff while facing a heterogeneous customer base. Finding the optimal nonlinear tariff has long been stated by economists as a direct revelation mechanism. The solution of such standard problem consists of a pair of functions, $\{T(t), q(t)\}$, where a consumer of —possibly multidimensional— type t gets assigned

[1] Indeed, the 1936 Robinson-Patman Act amended Section 2a of the 1914 Clayton Act to make unlawful for a seller "to discriminate in price between different purchasers of commodities of like grade and quality" where substantial injury to competition may result. In practice this has meant important pricing restrictions for intermediate good markets (secondary line or alleged injury to rivals of the buyer receiving the discriminatory price). See O'Brien and Shaffer (1994) for further details.

[2] We should mention here the recent paper by Chan, Hall, and Rust (2004) where the comparison between a fully nonlinear tariff and a uniform price is briefly considered within a dynamic model of optimal inventory and sale decision.

[3] This is a remarkable paper quite overlooked in the recent literature of nonlinear pricing. Using straightforward mathematical methods it establishes the basic results of nonlinear pricing models later developed by Armstrong (1996) and Rochet and Stole (2002). In particular Roberts (1979) first proves that with two-dimensional types exclusion at the bottom is always optimal. He furthermore proves that the optimal tariff induces efficient pricing not only at the top as usual, but *also* at the bottom.

a consumption level $q(t)$ for an associated payment $T(t)$. Most empirical applications have only made use of the predictions of this general model regarding the behavior of $q(t)$ while ignoring the less data demanding use of the optimal tariff function $T(t)$.

This paper builds upon a recent line of research that makes use of the predictions of a nonlinear pricing model regarding the optimal tariff function $T(t) = T[q(t)]$. The information contained in the position and shape of tariffs has been recently used either to recover a single dimensional index of quality of products —as in Crawford and Shum (2006)— or alternatively, to identify the distribution of unobserved consumer heterogeneity both under monopoly —Miravete (2006)— and duopoly —Miravete and Röller (2004)—. In this paper we broaden this approach by presenting a functionally flexible formulation of the random participation nonlinear pricing model of Rochet and Stole (2002). We solve this model numerically to identify the value of the structural parameters that generate the closest match to a large number of tariffs actually offered by monopoly carriers in the early U.S. cellular telephone industry between 1984 and 1988. The goal is to recover the parameters of the distribution of consumer types defined in a two-dimensional space: usage intensity of preferences (vertical heterogeneity), and individual-specific opportunity costs of participating in the market (horizontal heterogeneity).

Consumers may differ in several ways. A monopolist should, in principle, screen his clients with respect to as many dimensions as he can use to increase profits. Different type dimensions are defined by their induced price-independent shifts of individual demands. In telecommunications, for instance, tariffs may account for accumulated monthly usage, time of use (peak, shoulder, off-peak), distance (local, intrastate, interstate, international), frequency (unlimited calling and substantial discounts when calling family and friends), and identity of the network where calls are terminated (within own-network discounts and fixed/mobile interconnection surcharges) among others. Evidently these are not necessarily the only criteria to screen telephone customers. We should recall here the use of "creative pricing practices" such as short message services, news alert, stock market quotes, downloading of games, icons, and tones, sending and receiving e-mails and pictures, internet access, and other services common to third generation mobile technology. A profit maximizing monopolist will make use of the dimensions that capture most of the heterogeneity of consumers. If consumers show a very similar behavior in one given dimension the monopolist will most likely not screen them with respect to such dimension if implementing the screening device is costly. In an abstract setting where demands are linear we can distinguish, at most, two type dimensions summarized by the intercept and slope of demand, *i.e.*, the maximum willingness to pay for the product and the price responsiveness of each individual. Additional dimensions of consumer types will conform other features of the demand of consumers such as its degree of concavity or convexity or the number of related products that we have to account for. The added difficulty of multidimensional screening problems follows because it is generally impossible to order consumer preferences unequivocally. The identity of the consumer with the highest willingness to pay depends on the interaction of the different

type dimensions and is generally different for each possible price per unit of the product. This leads to optimal exclusion and bunching at the bottom —Armstrong (1996)— and non-monotonicity of optimal tariffs —Wilson (1993, §13-14)—, so that bunching for intermediate regions of the support of consumer types also occurs. The techniques to solve such problems are complicated and in practice difficult to be used in the empirical work.[4] In theory almost any tariff behavior is possible under numerous regularity conditions. Thus, multidimensional screening models offer very little guidance to help identifying the structural parameters that lead to a particular tariff solution with some specific observed features.

The generalized single-dimensional screening model of Rochet and Stole (2002) —*RS* hereafter— is an interesting compromise between the standard single dimensional screening model of Mussa and Rosen (1978) and the general multidimensional screening model described in the previous paragraph and first discussed by Mirrlees (1971). We use a flexible formulation of the *RS* model to identify the structural parameters that best describe the demand and distribution of consumer heterogeneity. In Mussa and Rosen (1978) consumers valuations are ordered along a single dimension that also determines whether individuals participate in the market or not. If a consumer of type $t = 3$ finds attractive enough to participate in the market and consume, say $q(3) = 17$, then a consumer of type $t = 5$, with intrinsic higher valuation for every level of consumption possible, will *necessarily* participate in the market and consume $q(5) > 17$. This feature of the model —the well known single-crossing property— leads to a recursive variational problem that characterizes the optimal nonlinear tariff, a simplification that ensures that the incentive compatibility (*IC*) constraint is binding only upwards and can be enforced locally. Indeed, the optimal tariff is found by solving a first order nonlinear differential equation with a single boundary condition given by the reservation utility of the lowest active consumer type.[5]

In the *RS* model consumers differ in their valuation of the product, t, as well as in the value of their outside option associated to non-participation in the market, x. In this simplified framework, all additional sources of heterogeneity of consumers are summarized by the scalar x that enters additively into the utility function of consumers and is, by assumption, independently distributed from t. Thus, participation becomes endogenous because it depends on the alternatives available to each consumer.[6] The existence of horizontal heterogeneity —the x dimension— reduces the ability of the monopolist to extract consumers' informational rents. Since participation is endogenous, it is optimal for the monopolist

[4] See Rochet and Choné (1998) for the most general version available of this multidimensional screening model. Laffont, Maskin, and Rochet (1987) first provided with a closed form solution for such a problem in a model with two-dimensional types.

[5] The slightly more general model of Maskin and Riley (1984) addresses screening consumers with respect to the different quantities that they purchase instead of the acquisition of a single unit of diverse quality. Footnote 5 of Rochet and Stole (2002) discusses the formal equivalence of both models although they interpret their own results from the perspective of quality discrimination among consumers with unit demands. Given our application to cellular pricing defined on airtime consumption, the interpretation of Maskin and Riley (1984) appears more appropriate to our case.

[6] This feature of the model makes this framework particularly suitable to address competitive environments, a task that we postpone for future research. In a competitive environment the outside option of a consumer might be given by the features of tariffs of competitors.

to leave some additional surplus to a consumer of (vertical) type t in order to increase the likelihood that she participates in the market; a decision that depends on the realization of her (horizontal) type x. Consequently, the solution of the optimal nonlinear tariff when market participation is endogenous leads to a higher level of quality (quantity) for each type t relative to the solution of Mussa and Rosen (1978); a key reference outcome that is replicated under the *RS* framework when the distribution of x becomes degenerate. Ignoring the possibility of horizontal heterogeneity may lead to overestimate the quantity underprovision needed to induce full separation of different consumer types. *Indicate here by how much the the second dimension adds to the average consumption relative to* MR.

However, making the participation decision endogenous comes at a significant cost. The variational problem that characterizes the optimal tariff is no longer recursive and instead it becomes a two-point boundary problem where a nonlinear second order differential equation provides the efficient quality (quantity) for the highest and lowest value in the support of t. Two-point boundary problems are difficult to solve in closed form. This explains why so few general features of the monopoly equilibria can be characterized analytically by Rochet and Stole (2002, §4.1–4.2). In the present paper we use a flexible formulation and solve the optimal monopoly pricing numerically for different values of the structural parameters in order to match our model predictions to the actual tariffs offered by monopolists in several local markets in the early days of the U.S. cellular industry.

Why do we adopt the *RS* random participation framework to build our empirical equilibrium model? There are at least two compelling reasons to chose it over alternative models of single dimensional or multidimensional screening. First, the *RS* solution falls in between the solution of the single dimensional model of Maskin and Riley (1984) —*MR* hereafter— and the first best (*FB*) allocation where the monopolist is able to extract all the informational rents from each consumer type (first degree price discrimination). From an empirical point of view this feature of the *RS* model reduces the likelihood of misspecification. Assuming either full information or a single dimensional screening model would lead to inconsistent estimates if the participation decision is indeed endogenous. In addition, estimating an *RS* based model does not rule out the possibility of concluding that the data is more consistent with either a single dimensional *MR* model or with the *FB* solution that arises in the absence of asymmetry of information. Second (and although this is not our case), nonlinear tariffs frequently include a monthly allowance of free minutes of airtime usage together with the payment of the fixed monthly fee. Wilson (1993, §6.4) shows that such an allowance is a constraint that needs to be imposed exogenously to characterize the equilibrium nonlinear tariff in the *MR* single dimensional screening model. Contrary to *MR*, the *RS* model endogenously explains this feature of the so-called "bucket tariffs."

In this paper we present a method of recovering the structural parameters from a generalized screening model with endogenous participation decisions. We then use these parameters to address pricing issues that have been neglected in the literature and suggest potential policy applications of out methodol-

ogy. We feel that our work contributes to the empirical literature on nonlinear pricing in at least two ways. First, we present an *RS* based model that can be solved numerically to conform actual tariffs. Functional forms are chosen to allow flexible enough solutions, as well as to provide with economically intuitive interpretations of structural parameters. In estimating this model, we make use of the moment conditions that characterize the set of optional tariffs used to implement the optimal nonlinear tariff in practice. These moment conditions serve as the way to estimate a behavioral model where the different core parameters of the theoretical model are conditioned on observable demand and cost variables. Second, we use the structural estimates to evaluate the welfare performance of alternative pricing strategies such as uniform pricing, flat tariff, two-part tariff, and a Coasian tariff (marginal cost plus a fixed fee). An non-negligible effect of these different pricing mechanisms is to induce more or less participation of consumers. This indeed is the source of important nonlinearities affecting welfare effects, and the source of most difficulties to rank the performance of different pricing mechanisms. Thus, our framework allows us to conduct relevant policy evaluations such as to measure the costs *vs.* induced benefits of implementing a universal service requirement.

Our approach extends the work of Miravete (2006) and Miravete and Röller (2004) regarding the identification of structural parameters of a nonlinear pricing model from the shape and position of the tariff. As for the theoretical evaluations, we intend to fill the existing gap in the literature that has focused mostly on the characterization of equilibrium tariffs rather than on the desirability of price discrimination over uniform pricing. In relation to the policy evaluations, the universal service requirement is a commonly discussed issue in industries such as public utilities where nonlinear tariffs enjoy a widespread use in part because the help recovering the substantial fixed costs common in public utilities by incentivating consumption of large customers through quantity discounts and thus fostering economic efficiency. Policies aimed to achieve universal service are not exclusive of monopoly markets. Indeed, they have been actively used in promoting broadband access after the 1996 U.S. Telecommunications Act. Subsidies, however, may generate important inefficiencies depending on how they are financed.[7] The evaluation carried out in this paper should be understood as an illustrative first attempt to evaluating interesting alternative policies and illustrating the redistribution effects that universal service policies may have among diverse consumers. We certainly hope that the limited information required to conduct such an analysis will turn our approach appealing for policy makers and management of firms and utilities engaged in price discrimination strategies.

[7] Crandall and Waverman (2000) estimate subsidies to ensure that schools, libraries, and rural health facilities have access to internet amount to $2.65bn a year. Crandall, Hahn, Litan, and Wallsten (2004) argue in favor of removing retail price regulation of DSL providers as well as phasing out network sharing requirements as a far more effective strategy to induce investment in broadband capacity than direct subsidies. These authors also argue that broadband subsidies financed with taxation on other broadband services are particularly damaging to achieve a higher penetration rate in this market.

Our results show that firms engage in more targeted screening (offering more tariff options) when the distribution of consumers' valuations is less concentrated around the highest consumer type. In general, the value of the estimated structural parameters capture the features of this early industry, *i.e.*, one with important fixed but negligible marginal costs, and with very few subscribers although of considerably high valuation for new the cellular service.

In our model, and due to the existence of an outside option that is not correlated with customers' valuation of the product, firms have to soften their markups in order to induce some additional consumers to participate in the market. Because of our estimates indicate that the distribution of consumer types is quite concentrated around low valuation customers, the expected profits of a simple flat tariff always exceeds 80% of the maximum potential profits with a fully nonlinear tariff; uniforms pricing achieves 98%; and a simple two-part tariff always exceeds 99% of maximum potential profits. Evidently, when the distribution of consumers is so concentrated around low valuation customers there is little gain from nonlinear pricing and most of it simply consist in excluding them and targeting those with high willingness to pay for the service.

Profits always increase with the addition of tariff options to the firms' menus but they are mostly due to fixed charges because as more numerous options are offered, lower valuation customers subscribe to the cellular service but end up making a very limited use of it, thus overall reducing the median airtime usage. But perhaps one of the most interesting finding is that once we incorporate the effects of consumer surplus these pricing mechanisms are ranked differently and uniform pricing always dominates any of the other nonlinear tariffs from a welfare perspective. Finally, a universal service policy can certainly ease subscription but if it is funded through cross-subsidization, the implied distributional effects reduce the effectiveness of such policy to a third of its potential.

The paper is organized as follows. Section 2 intends to familiarize the reader with the institutions and regulations of the early U.S. cellular telephone industry. Section 3 presents our *RS* based model of nonlinear pricing with endogenous participation decisions. Section 4 presents a behavioral model linking the core parameters of the theoretical model to market specific demand and cost variables. Then, structural parameters are recovered by making use of the moment restrictions implied by firms' specific implementation of the nonlinear tariff solution by means of a menu of self-selecting two-part tariffs. Section 5 evaluates the welfare of alternative pricing mechanisms relative to the unconstrained fully nonlinear tariff and presents the results of a sensitivity analysis aimed to ascertain the robustness of our results. In particular we consider alternative specifications of demand and more general distributions of consumer valuations. This analysis allows us, to quantify the welfare bias of ignoring an existing type independent outside option for each consumer. Section 6 evaluates the welfare effects of enforcing two versions of the universal service requirement. Finally, Section 7 concludes.

2 U.S. Cellular Monopoly Markets

By mid 1980s, the Federal Communications Commission (*FCC*) granted permission to create 305 non–overlapping cellular markets around U.S. standard metropolitan statistical areas (*SMSA*) to be served by two competing carriers. Cellular technology used low powered transmitters that exhausted the allocated bandwidth within small cells. A single conversation with the older high powered transmitters of car phones used a channel within a radius of about 75 miles. The *FCC* required that the low powered transmitters of the new cellular technology used these channels within a maximum radius of 30 miles.[8]

In 1981, the *FCC* set aside 50 MHz of spectrum in the 800 MHz band for cellular services. One of the two cellular channel blocks in each market –the B block or *wireline* license– was awarded to a local *wireline* carrier, while the A block –the *nonwireline* license– was initially awarded by comparative hearing to a carrier other than the local *wireline* incumbent. Licenses were awarded in ten tiers, from more to less populated markets, beginning in 1984. In general the *wireline* licensee offered the service first and enjoyed a temporary monopoly position until the *nonwireline* carrier entered the market, normally within six months of being awarded the license as required by the *FCC*. However, the administrative review process to award licenses among hundreds of contenders only based on technical issues and credibility of the announced investment commitments proved to be far more costly than initially expected. After awarding licenses for the thirty largest markets by means of this expensive and time consuming *beauty contest* —there were up to 579 contenders for a single license—, and while the application review of the second tier of 30 markets was on its way, rules were adopted to award the remaining *nonwireline* licenses through lotteries. Court appeals against the administrative awarding of the *nonwireline* licenses in the early tiers, and legal, technical, or managerial difficulties to start operating the lottery-awarded licenses in subsequent ones led to a situation of temporary monopoly in many of the largest local cellular markets.

In this paper we use data from these monopoly markets. Data include detailed tariff information for about 50 *wireline* monopoly carriers between 1984 and 1988. The length of the monopoly phase in each market can be considered exogenous. Entry of the second firm depended mostly on court decisions regarding the contested administrative award of the *nonwireline* license by the *FCC*. Actually, entry of the second carrier always occurred soon after a firm court decision was made. Thus, predatory pricing incentives can safely be ruled out in this application.

Data include all tariff plans offered by each firm, their monthly subscription fee, rate per minute during peak hours, and the monthly allowance of free minutes, if any. We ignore off-peak pricing because at this early market cellular service is mostly targeting business customers: the handset was initially priced

[8] Hausman (2002, §2.1) documents how the combination of cell splitting and sectorization increased capacity by a factor of 8 relative to the pre-existing, non-cellular, car telephone technology.

Table 1: Tariff Features

	Monthly Fee		Rate per Minute	
Option No.	Mean	Std.Dev.	Mean	Std.Dev.
Markets with ONE option (56 observations)				
1	27.98	(12.53)	0.37	(0.07)
Markets with TWO options (25 observations)				
1	15.39	(6.35)	0.56	(0.16)
2	39.31	(6.61)	0.35	(0.08)
Markets with THREE options (14 observations)				
1	4.00	(2.08)	0.66	(0.15)
2	15.42	(8.88)	0.41	(0.07)
3	32.44	(20.61)	0.33	(0.06)

Mean and standard deviations (between parentheses) of monthly fixed fees F_i and rate per minute p_i are measured in dollars.

at 3,000 dollars and peak pricing spanned over a dayly 11-to-13 hour band at that time. In about one third of the markets, monopolists always offered the same tariff during all sample periods. Thus, we only include the earliest and latest tariff offered by each carrier during the monopoly phase of the cellular industry. In general, higher fixed monthly fees F_i go together with lower rates per minute p_i. Table 1 describes the features of the non-dominated tariff options offered by the firms of our sample. In general the addition of new tariff options to a firm's offering aims to expand his customer base by reducing the monthly fixed fee while increasing the tariff rate per minute of airtime. As we will see in our econometric analysis, this practice induce consumers with lower valuations to subscribe more frequently in those markets where firms offer more tariff options.

In addition, tariff data are complemented with market specific demand and cost information as well as an ownership indicator for each firm. Table 2 presents the descriptive statistics of our data. All money valued magnitudes are expressed in dollars of July 1986. Data definitions and sources are detailed in Appendix A.[9]

Perhaps the only variable that needs a detailed discussion is COVERAGE. This is an important variable that critically determines the dispersion of the distribution of horizontal heterogeneity, x. Unfortunately we do not have any information available on individual subscribers, neither a precise definition of what constitutes the potential market in this early cellular telephone industry (although the latter is a common problem in applied industrial economics). We therefore proceed by defining COVERAGE = 1,3×TCELLS/(BUSINESS+250*POPULATION), *i.e.*, we identify the potential market as that represented by a cellular telephone for each firm and families with an average number of four members. At this early stage

[9] Many other variables were available but we only report here those that are used in the empirical analysis of Section 4.3.

Table 2: Descriptive Statistics

	First Quarter		*Last Quarter*	
Variables	Mean	Std.Dev.	Mean	Std.Dev.
PLANS	1.4894	0.6875	1.6250	0.7889
COVERAGE	0.0677	0.0674	0.0664	0.0666
TIME	4.2553	2.4179	11.6250	3.9873
BUSINESS	45.9747	69.2974	45.3892	68.6780
POPULATION	1797.8452	2899.2626	1784.2193	2869.0236
INCOME	28.1633	3.7277	27.4650	3.2514
COMMUTING	26.0702	2.8878	26.0167	2.8887
RAIN	3.4118	1.8906	3.0431	2.0116
POVERTY	11.0064	3.0099	11.1104	2.9422
POP-AGE	32.7702	2.8541	32.6917	2.8022
EDUCATION	12.5404	0.2482	12.5313	0.2545
COVERAGE	0.0677	0.0674	0.0664	0.0666
HHSIZE	2.6208	0.2848	2.6144	0.2830
sdv(HHSIZE)	1.4650	0.1587	1.4614	0.1582
DENSITY	15.6642	13.9766	14.7665	12.9405
TCELLS	17.8298	18.4807	17.6875	18.3096
OPERATE	6.5543	1.5801	6.5443	1.5228
PRIME	10.8191	0.6442	9.1374	0.9556
WAGE	6.9416	1.5404	7.2745	1.6554
BELL	0.8298	0.3799	0.8333	0.3766
Observations	47		48	

All variables defined in the text.

of development of the industry, the number of subscribers were effectively constrained by the capacity of cellular carriers and a maximum of 1,300 could be simultaneously served by each antenna.[10] This definition of market penetration is certainly arbitrary, but we feel that it is not unreasonable. We could have considered only the number business as the potential market. However, private use of cellular telephony also existed and cellular carriers targeted private users in their marketing campaigns as well. Similarly, subscribing to a cellular service provider was far from common, and thus if we consider the total population as the potential market, coverage becomes minimal, implying necessarily that the distribution of the horizontal heterogeneity dimension, x, is almost degenerate. Evidently, our results depend critically on the definition of COVERAGE. Were a better definition of market penetration available we would have not hesitated using it but in its absence, we feel that the mix of BUSINESS and a fraction of POPULATION that approximates the number of families of each market appears to us as a reasonable definition of the potential market; helps identifying the parameters of our model; and thus highlights the methodological contribution of this paper and the value of its potential applications.

[10] Parker and Röller (1997, §4) report that one antenna could serve between 1,100 and 1,300 subscribers for the average use of cellular telephony at that time. Using a small sample of markets they also report that the correlation between the number of antennas and the number of subscribers exceeds 90%.

3 A Flexible Model of Nonlinear Pricing with Random Participation

This section presents an equilibrium model of monopolistic nonlinear tariff with optimal endogenous participation decisions of consumers. Our specification makes use of particular functional form assumptions to allow us implementing the *RS* model empirically. Since all our functional form assumptions fulfill the requirements of the *RS* model the equilibrium tariff shares all the properties of the general solution that Rochet and Stole (2002) discuss. In this section we summarize the most relevant features of the equilibrium tariff and convey the economic intuition behind our choice of functional forms.

3.1 Basic Elements

A monopolist produces a good or service q at a constant marginal cost c and designs a nonlinear and deterministic tariff $P(q)$ to maximize profits:[11]

$$\pi(q) = P(q) - cq. \tag{1}$$

Consumers' preferences are indexed by a two-dimensional taste parameter (t, x) where t captures the vertical heterogeneity of consumers, *i.e.*, their intensity of preferences for the consumption of a given good; while x captures the horizontal heterogeneity of consumers, *i.e.*, an individual specific valuation of the outside option of not purchasing the product at all. The net utility from trade subtracts the outside opportunity cost from the indirect utility function u that is quadratic in consumption (leading to a linear demand for q):

$$v = tq - \frac{\gamma}{2}q^2 - P(q) - x, \tag{2}$$

where t and x are assumed to be independently distributed according to a Burr type XII distribution with parameter λ and an exponential distribution with parameter $1/\phi$, respectively:[12]

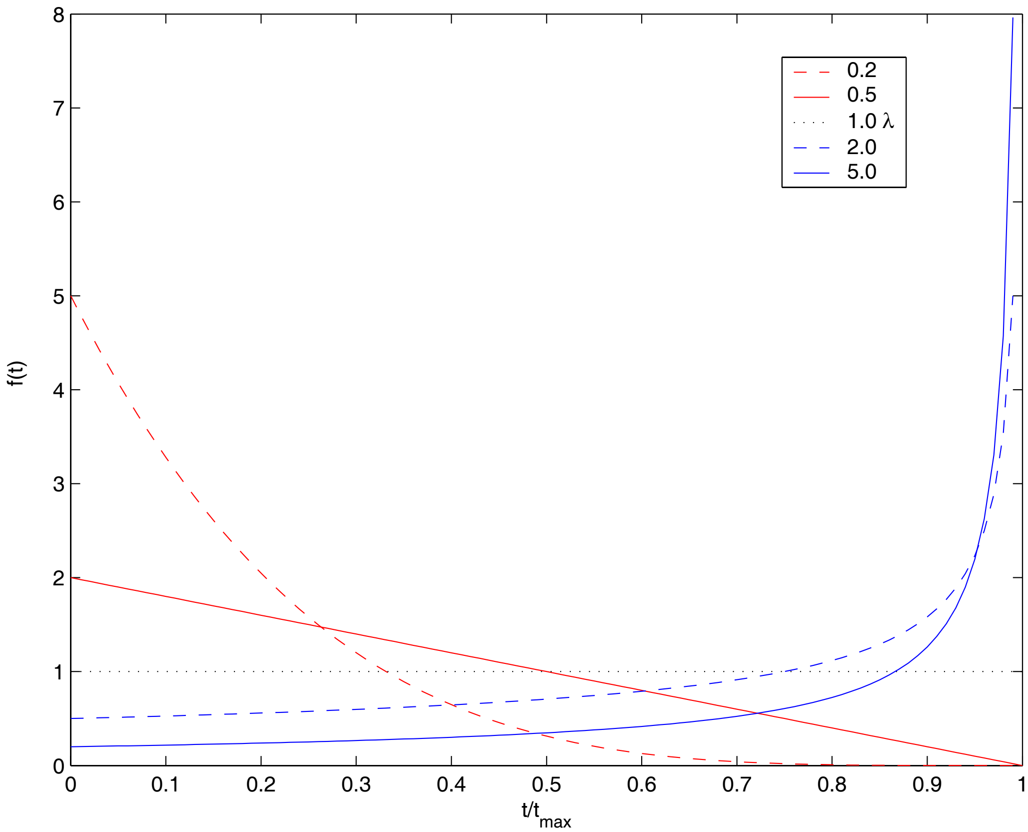
$$t \sim F(t) = 1 - \left[1 - \frac{t - \underline{t}}{\bar{t} - \underline{t}}\right]^{\frac{1}{\lambda}} \quad ;\ \lambda \geq 0,\ \ t \in \mathbb{T} = [\underline{t}, \bar{t}], \tag{3a}$$

$$x \sim G(x) = 1 - \exp\left(-\frac{x}{\phi}\right) \quad ;\ \phi \geq 0,\ \ x \in \mathbb{R}_+, \tag{3b}$$

[11] See Rochet and Stole (2002, footnote 15) for a detailed discussion in favor of deterministic contracts over randomization when consumers have to incur in transportation costs or when they can anonymously return to purchase repeatedly from the principal. Moreover, allocation to consumers using lotteries are not used in cellular telephone pricing.

[12] Rochet and Stole (2002) also include a parameter σ to capture the importance of transportation costs. Thus, their distribution of the horizontal heterogeneity dimension is $G(x) = 1 - \exp(-x/\sigma\phi)$. Using only the tariff information available, we encountered that ϕ and σ cannot be independently identified.

Figure 1: $F(t)$ **— Burr type XII distribution**



with $c \leq \underline{t} \leq \bar{t}$. Finally, adding up (1) and (2) we obtain the joint surplus from trade:

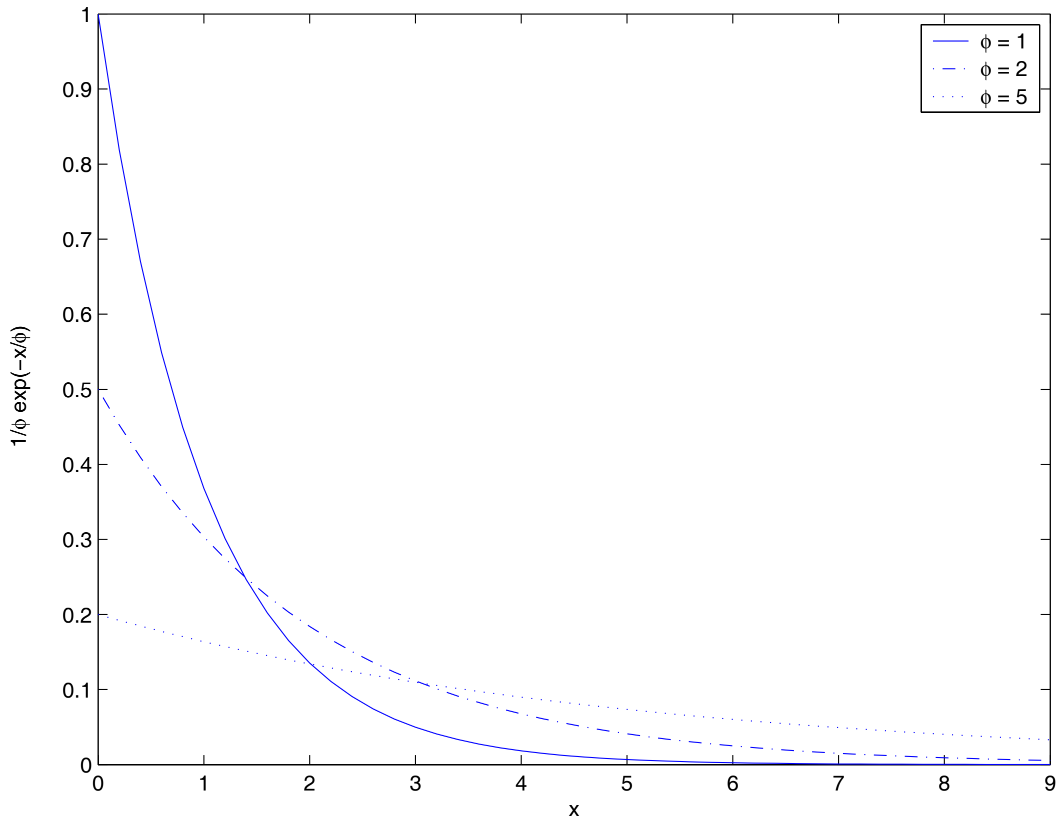
$$S(q, t) = (t - c)q - \frac{\gamma}{2}q^2 - x, \tag{4}$$

from where the *FB* solution is $q^{\text{fb}}(t) = (t - c)/\gamma$, *i.e.*, the purchase level corresponding to the full information, first degree, price discrimination tariff.

The Burr type XII distribution of t in equation (3a) is a restricted beta distribution $F(t \,|\, 1, \lambda^{-1})$. The economic interpretation of λ is appealing and intuitive as it is directly related to the average markups charged by the monopolist to consumers with diverse valuations.[13] Different values of λ identify whether high valuation consumers are more or less numerous than low valuation consumers. This is shown in Figure 1. If $\lambda = 1$ then t is uniformly distributed (*i.e.*, there are the same proportion of high and low valuation customers). If $\lambda > 1$, consumers are more concentrated around higher values of t (more numerous high valuation customers) and *vice versa* when $\lambda < 1$. If $\lambda = 0$, distribution (3a) becomes degenerate at $t = \underline{t}$,

[13] Miravete (2002, Appendix) and Miravete (2005) analyze at length the relationship between λ and the markup charged to different single–dimensional consumers.

Figure 2: $G(x)$ **— Exponential distribution**



asymmetric information turns out to be irrelevant and homogeneous consumers can be efficiently priced by means of a single two-part tariff.[14]

Similarly, the exponential distribution (3a) is a good approximation to model the horizontal heterogeneity dimension, x. This exponential density is monotonically decreasing with a maximum at $x = 0$ and has an expected value of ϕ. This parameter increases when the average consumer has better outside options (perhaps offered by a competing firm in a horizontally differentiated industry). Thus, as it is shown in Figure 2, a small value of ϕ leads to a fast rate of decay of the probability density function of x. In terms of our model, it means that a small ϕ is associated with situations where consumer horizontal heterogeneity is also small. Thus, the probability associated to the event that consumers attach a large value to the outside option is very low. The contrary is true for large values of ϕ. The horizontal heterogeneity summarized by x distinguishes the *RS* model from the standard single-dimensional *MR* screening model. If $\phi \to 0$ the horizontal heterogeneity disappears and we are left with a standard, single-dimensional, nonlinear pricing model. The role of x in the model is to capture the idea that consumers may have different valuations of their outside option of not purchasing q independently of how their preferences are ranked with respect to this

[14] For further results and properties of the Burr type XII distribution see Johnson, Kotz, and Balakrishnan (1994, §12.4.5).

good through t. Thus, for large values of ϕ, it is likely that we find consumers with a high valuation of their outside option. These consumers will, most likely, not participate in the market. Thus, if the monopolist is to maximize profits, he must balance the alternative of extracting substantial informational rents from high t consumers with limited probability *vs.* increasing his customer base by lowering the markup charged to each consumer t. These opposite incentives explain most of the differences of the present model relative to the standard *MR* model, the most important of which is that in the *RS* model the monopolist has a reduced ability to extract informational rents, and thus, markups are lower than in the *MR* model.

Because we are unable to observe t and x we decided to assume, as *RS* did, that type dimensions are independently distributed. If, for instance, these type dimensions were positively correlated, a higher valuation t goes together —although not perfectly— with a higher probability of participation in the market and the solution of the *RS* model will be closer to *MR*. The contrary would happen if t and x were negatively correlated. The strongest this negative correlation is, the less able is the monopolist to extract informational rents from consumers, the further we are from the *MR* tariff, and the closer to the *FB* solution. Thus, if negative correlation exists, our model will overestimate ϕ, while this parameter will be underestimated if the correlation between x and t is positive.

Finally, because of the existence of the individual specific outside option, x, the market share of the monopolist among consumers of type t is given by the following composition of distributions:

$$M(u,t) = \text{Prob}[t, x \leq u] = G(u)f(t) = \left[1 - \exp\left(-\frac{u}{\phi}\right)\right] \frac{1}{\lambda(\bar{t} - \underline{t})} \left(1 - \frac{t - \underline{t}}{\bar{t} - \underline{t}}\right)^{\frac{1}{\lambda} - 1}, \tag{5}$$

where $f(t) = F'(t)$ and $g(t) = G'(t)$ are the probability density functions associated to (3a) and (3b), respectively; $G(x)$ is log–concave —*e.g.*, Karlin (1968, §1.5)—; and the following mean bound condition holds:

$$\lim_{x \to +\infty} xg(x) = \lim_{x \to +\infty} \frac{x}{\phi} \exp\left(-\frac{x}{\phi}\right) = 0. \tag{6}$$

Furthermore, the inverse hazard rate of M over u is nondecreasing in u:

$$H(u,t) = \frac{M(u,t)}{M_u(u,t)} = \phi \cdot \frac{1 - \exp(-u/\phi)}{\exp(-u/\phi)}, \qquad \text{s.t.} \quad H_u(u,t) = \frac{1}{\exp(-u/\phi)} \geq 0. \tag{7}$$

3.2 Fully Nonlinear Tariff

The monopolist maximizes the expected profits designing the optimal direct revelation mechanism $\{P(t), q(t)\}_{t \in T}$. It consists of maximizing the following unconditional expected profits function with respect to $q(t)$ and $u(t)$ provided that individuals of type t report their type truthfully as $u(t)$, *i.e.*, subject to the

participation or individual rationality (*IR*) and incentive compatibility (*IC*) constraints, respectively, that is $\dot{u}(t) = q(t)$ and $\dot{q}(t) \geq 0$ for all t:[15]

$$\int_{\underline{t}}^{\bar{t}} M(u(t), t) \left[S(q(t), t) - u(t) \right] dt, \tag{8}$$

A piecewise-smooth function $q(t)$ is implementable by a tariff function $P(t)$ if and only if $q(t)$ is nondecreasing, $u(t)$ is absolutely continuous, and $\dot{u}(t) = q(t)$ at all continuity points of $q(t)$. This property of the *RS* model is common to the *MR* model.[16] After substituting our specific functional form assumptions, this problem can be stated as:[17]

$$\max_{q(t),u(t)} \int_{\underline{t}}^{\bar{t}} \left[1 - \exp\left(-\frac{u(t)}{\phi} \right) \right] \left(\frac{\bar{t} - t}{\bar{t} - \underline{t}} \right)^{\frac{1}{\lambda} - 1} \left[(t - c)q(t) - \frac{\gamma}{2} q^2(t) - u(t) \right] dt, \tag{9a}$$

$$\dot{u}(t) = q(t) \geq 0, \tag{9b}$$

$$\dot{q}(t) \geq 0, \tag{9c}$$

$$u(t) = tq(t) - \frac{\gamma}{2} q^2(t) - P(t) \geq 0. \tag{9d}$$

As indicated in the introduction, this is not a recursive problem because of the existence of an individual specific outside option so that consumers will only participate when $u(t) \geq x$. It is therefore not possible to incorporate the *IC* constraints directly. In addition, the profit function is not separable in $u(t)$ and $q(t)$ and is nonlinear in $u(t)$, therefore turning impossible to have it integrated by parts. Rochet and Stole (2002) discuss at length the features of this equilibrium that critically depend on the dispersion of the distribution of t. Briefly, tariff features can be summarized as follows:

1. Provided that $G(x)$ is log-concave, as in our case, full market coverage by a monopolist is never optimal regardless of whether the support of x is bounded or not.[18] This is an important feature of the model, —actually common to Armstrong (1996) and Roberts (1979)— because it allows us to evaluate the cost of promoting consumer participation beyond the monopolist's optimal decision. In Section 6 we use the model to study the welfare gains induced by the *Universal Service* requirements commonly enforced in telecommunications.

[15] Notice that $u(t)$ determines the consumer's decision to participate through the unconditional probability of participation $M(u(t), t)$.

[16] See Rochet and Stole (2002, Lemma 2), Tirole (1989, §3.5), and Wilson (1993, §6.2).

[17] For convenience we denote the first and second derivative of the utility function with respect to the vertical type dimension t as $\dot{u}(t)$ and $\ddot{u}(t)$, respectively. Similarly, $\dot{q}(t)$ is the first derivative of consumption with respect to t.

[18] See discussion following Proposition 3 in Rochet and Stole (2002).

2. If the highest valuation $\bar{t}$ is sufficiently large relative to $\underline{t}$ the equilibrium tariff leads to bunching at the bottom, *i.e.*, the optimal tariff is such that all consumers on $t \in \mathbb{T}^\circ = [\underline{t}, t^\circ]$ purchase the same amount q° at $P^\circ = t^\circ q^\circ - \gamma(q^\circ)^2/2 - u(t^\circ)$. Furthermore, since $G(x)$ is log-concave, $\mathbb{T}^\circ$ is ensured to be a compact set. Pooling of different consumer types will never occur in any other region of the type space. Furthermore, in the upper interval of the type space $\mathbb{T} \setminus \mathbb{T}^\circ = [t^\circ, \bar{t}]$ the optimal utility profile solves the following second-order boundary-value problem:[19]

$$\begin{aligned} 0 &= \frac{1}{\phi} \exp\left(-\frac{u(t)}{\phi}\right)(\bar{t} - t)\left[u(t) - \frac{\gamma}{2}\dot{u}^2(t)\right] + \left[1 - \exp\left(-\frac{u(t)}{\phi}\right)\right](\bar{t} - t)\left[2 - \gamma\ddot{u}(t)\right] \qquad (10a) \\ &\quad + \frac{\lambda - 1}{\lambda}\left[1 - \exp\left(-\frac{u(t)}{\phi}\right)\right]\left[(t - c) - \gamma\dot{u}(t)\right], \\ \dot{u}(t) &= q(t) \geq 0, \quad \text{for} \quad t \geq t^\circ \quad \text{with} \quad \dot{u}(t^\circ) = q^\circ, \qquad (10b) \\ u(t) &= tq(t) - \frac{\gamma}{2}q^2(t) - P(t) \geq 0, \quad \text{for} \quad t \geq t^\circ \quad \text{with} \quad u(t^\circ) = u^\circ, \qquad (10c) \\ q(\bar{t}) &= q^{\text{fb}}(\bar{t}) \geq 0. \qquad (10d) \end{aligned}$$

Thus, the efficiency at the top result holds so that $p(q(\bar{t})) = c$, while low valuation customers might be pooled and possibly excluded in order to increase the informational rents extracted from higher valuation types. However, if the support of $F(t)$ is not too spread, all consumers are served, *i.e.*, $\underline{t} = t^\circ$ or $\mathbb{T}^\circ = \emptyset$; bunching does not exist; and (10c) needs to be replaced by $q(\underline{t}) \geq q^{\text{fb}}(\underline{t}) \geq 0$, so that consumers are efficiently priced both at the *top* as usual, but also at the *bottom*. In short, if the monopolist finds profitable to serve the lowest type $\underline{t}$, he has to offer the most attractive price possible for this consumer to participate in the market for any realization of x. Thus, pricing at marginal cost becomes the optimal strategy.

3. The monopolist's supply is such that $q(t) \in \left(q^{\text{mr}}(t), q^{\text{fb}}(t)\right]$ for all $t \in \mathbb{T} = [\underline{t}, \bar{t}]$. Furthermore:

$$\lim_{\phi \to 0^+} q(t) = q^{\text{mr}}(t)\ ; \quad \forall t \in \mathbb{T}, \qquad (11)$$

so that our estimates can continuously approximate to the *MR* solution, which provides with a lower bound consumption for multidimensional consumers. Thus our empirical analysis will quantify the importance of the distortion due to the existence of the horizontal type x —quantity underprovision for each type t— relative to the *MR* case, as well as the corresponding welfare effects.

[19] This solution determines the lowest active consumer type t°. For advanced introductions to this topic see Ascher, Mattheij, and Russell (1988) or Keller (1968).

3.3 Qualifications

Before proceeding any further, we should point out few qualifications to our analysis. First, we are implicitly assuming that the monopolist solves a static problem every period. This equilibrium approach rules out the possibility of addressing whether these *wireline* firms used the monopoly period to expand their customer base beyond the optimal static monopoly solution in order to deprive the future entrant from the most valuable customers after signing them to long term contracts. If this dynamic consideration were present, the actual markup will be below the optimal static markup and thus we will underestimate λ.[20] As the optimal markup is less than the statically optimal one it induces more participation than the statically optimal, and thus ϕ would most likely be overestimated.

Second, we should recognize the extent to which regulation might induce firms to deviate from profit maximization. It appears that this is not the case. Shew (1994, §4) indicates that the awarding of cellular licenses happened in a political environment that favored less and not more price regulation.[21] In addition, cellular service —a luxury good at that time— was going to be provided competitively. Initially, few states engaged in retail price regulation, some of which did not even disclosed the criteria for testing the reasonableness of cellular service prices. But even in those cases, regulation was vaguely enforced because even when at most state authorities required tariff filing, they let carriers to initially set their own price caps as regulatory bodies where uncertain to evaluate the costs of providing this new service.

Third, our analysis relies heavily on functional form assumptions, which might question the validity of our welfare estimates. In addition to the independence between the distributions of t and x previously discussed we are assuming that demand is linear, that t follows a very particular distribution, and that the industry is characterized by a constant marginal cost. Thus, for instance, the latter assumption of constant returns to scale in equation (1) ignores the possibility that capacity constrains may influence pricing decisions to allocate the cellular service among the highest valuation customers. If the actual tariff incorporated capacity pricing elements, our model would probably overestimate marginal costs and therefore λ, as the actual markup in the presence of capacity constrains would exceed that of static pricing with constant returns to scale. For the opposite reasons of the first paragraph, if capacity constrains are actually present our model would underestimate ϕ. Contrary to the two previous qualifications, which are institutional and affect to the actual modeling of the pricing decision, we will address the effects

[20] The inverse relation between the price markup and the hazard rate of the distribution is well documented in the agency theory literature. See for instance Laffont and Tirole (1993, §1.4-1.5) or Maskin and Riley (1984, §4). As the hazard rate of $F(t)$ is exactly $h_F(t) = 1/[\lambda(\bar{t} - t)]$, λ is directly related to the markup actually charged by the monopolist. Therefore, any downwards deviation from the optimal markup leads to downwards bias in the estimation of the true λ.

[21] For instance, the 1980 Staggers Act had deregulated most railroad rates. Similarly, the 1978 Airline Deregulation Act mandated that the Civil Aeronautics Board terminated setting airfares beginning in 1983, although airfare competition was indeed allowed in May 1980. For further details, see Viscusi, Vernon, and Harrington (2005, §17). A dummy variable indicating whether a particular market was regulated was never found to be statistically significant in the analysis of Section 4.3.

of functional form assumptions by recomputing all pricing equilibrium for alternative specifications of demand, cost technology and distribution of the vertical type dimension, $F(t)$.

But certainly, the most important qualification to the *RS* model is to recognize that firms rarely offer a fully nonlinear tariff to their customers, and certainly never in our sample. In the early U.S. cellular industry many firms only offered a single two-part tariff while others offered two, and a maximum of three self-selecting two-part tariffs. An attempt to fit the lower envelope of a few optional two-part tariff with the prediction of the fully nonlinear tariff described in this section leads to a necessarily misspecified econometric model. Instead, we compute the optimal n-part tariff that best screens consumers for each set of structural parameters $(\lambda, \phi, \underline{t}, \bar{t}, c)$ conditioning on the actual number of tariff options offered by the monopolist of each market. To rationalize this particular number of tariff options as optimal we assume that monopolists in different markets face some commercialization cost ζ per tariff option offered to consumers. These commercialization costs include all non-observable incremental costs associated to the design, marketing, and advertising of each tariff option offered by firms. Since the incremental profits of adding an extra tariff option are positive but decreasing with the number of tariffs, a constant commercialization cost associated to each tariff option ensures that the monopolist find optimal to offer a menu with only a finite number of tariff options. Once we estimate all the structural parameters, our equilibrium approach allows us to recover an interval estimate for these commercialization costs based on the foregone incremental profits of not offering an additional tariff option in each market.[22] Next section incorporates the discrete choice of the number of tariff options within the *RS* framework and suggest how to estimate this model.

4 Empirical Analysis

This section addresses the empirical implementation of the *RS* model described above. We should distinguish three different issues related to this empirical implementation. First, the actual estimation of the structural parameters of the model. Here we encounter the difficulty that in fact firms only offer menus of self–selecting two–part tariffs instead of fully nonlinear tariffs. Our equilibrium approach resolves this issue by assuming the existence of a fixed commercialization cost per tariff option. Second, exploring how a behavioral model could be linked to exogenous demographics and other market characteristics to explain the observed cross–market variation of our estimated structural parameters; most of which characterize sources of asymmetric information or marginal costs, *i.e.*, economic variables that are not directly observable. A successful link between estimated structural parameters and observable demographics would allow

[22] The approximate implementation of nonlinear tariffs by means of a finite menu of self-selecting two-part tariffs has only been addressed theoretically by Wilson (1993, §6.4). Recently Seim and Viard (2005) study how the number of tariff options offered change with the number of firms present in each market. Miravete (2006), within a *MR* framework, evaluates the foregone profits of not offering an additional tariff option.

us to extend our policy evaluation for other markets not included in the sample. And finally, an analysis of the welfare implications of our estimates, which will be postponed to Section 5.

4.1 Screening with a Limited Number of Tariff Options

A monopolist in market i at time τ observes the distribution of consumer preferences given by parameters $(\lambda_{i\tau}, \phi_{i\tau}, \underline{t}_{i\tau}, \bar{t}_{i\tau})$; the slope of consumers' demand $\gamma_{i\tau}$; the marginal cost of production $c_{i\tau}$; and the cost of commercialization per tariff option $\zeta_{i\tau}$. We normalize $\underline{t}_{i\tau} = 0$, which implicitly assumes the systematic exclusion of low valuation types, a feature consistent with the low market penetration of the cellular telephone industry in the early age period of our sample.

Since offering tariff options is costly, in addition to solving (10a)-(10d) the monopolist decides how many tariff options to offer to his customers in order to maximize expected profits, *i.e.*, he has to choose the menu of n two-part tariffs that best approximate the fully nonlinear tariff solution within the set of n affine functions. A pair $\{A, b\}$ where A represents the fixed monthly fee and b the marginal tariff per unit of consumption fully describes any optional two-part tariff. Thus, the monopolist maximizes:

$$n_{i\tau} \in \arg \max_{\nu \in \mathbb{N}} \pi^{\star} \left(\nu \mid \lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau} \right) - \nu \cdot \zeta_{i\tau}. \tag{12}$$

where $\pi^{\star}(\nu \mid \cdot)$ denotes the expected profits of a monopolist that offers the ν self-selecting two-part tariffs that best implements the optimal fully nonlinear tariff option given by equations (10a)-(10d), that is:

$$\pi^{\star} \left(\nu \mid \lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau} \right) = \max_{\substack{A_1 < \ldots < A_\nu \\ b_1 > \ldots > b_\nu}} \pi \left(A_1, \ldots, A_\nu, b_1, \ldots, b_\nu, \mid \lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau} \right). \tag{13}$$

Contrary to other core parameters, we cannot obtain a point estimate of ζ. If a firm offers an n-part tariff in a particular environment characterized by $(\lambda_{i\tau}, \phi_{i\tau}, \underline{t}_{i\tau}, \bar{t}_{i\tau})$, it must be true that the incremental profits from offering n instead of $n-1$ options should exceed the marketing and commercialization cost ζ while at the same time the incremental profit of offering $n+1$ rather than n options should not justify introducing the $n+1$ tariff plan.[23] Once we estimate all structural parameters but ζ we can compute the expected profits of implementing the solution with any number of tariff options, thus obtaining the bounds of ζ as the incremental profits of offering either n or $n+1$ options. Therefore for each market i and time period τ we have the following observation rule:

[23] Wilson (1993, §8.3) shows that as long as IC is fulfilled and the distribution $F(t)$ is increasing hazard rate —and the Burr Type XII distribution fulfills this requirement as long as $\lambda > 0$—, the incremental profits of adding a tariff option are positive but decreasing in the number of tariff options offered. Thus, the incremental profits of offering an additional tariff option will eventually fall short of any fixed costs of commercialization, therefore effectively limiting the optimal number of tariff options offered.

$$n = \mathbf{1}\left[\pi^{\star}(n\,|\,\cdot) - \pi^{\star}(n-1\,|\,\cdot) > \zeta > \pi^{\star}(n+1\,|\,\cdot) - \pi^{\star}(n\,|\,\cdot)\right], \tag{14}$$

and where $\pi^{\star}(0\,|\,\cdot)$ denotes the expected profits with uniform pricing. Thus, we can think of this discrete choice of n as determined by the realization of the latent variable ζ exceeding different thresholds given by the magnitude of foregone profits of not offering an additional tariff option.

To complete the specification of the model, and make it suitable for econometric estimation we need to add some stochastic structure to this monopolist's maximization problem. Thus, each *core* parameter $\lambda_{i\tau}$, $\phi_{i\tau}$ $\bar{t}_{i\tau}$, $\gamma_{i\tau}$, $c_{i\tau}$, and $\zeta_{i\tau}$ is assumed to be a particular function of market specific characteristics observable to the econometrician, $\boldsymbol{Z}_{mi\tau}$, as well as others that remain unobservable and summarized by $\varepsilon_{i\tau}$, plus a vector of coefficients $\boldsymbol{\delta}_m$ to be estimated for each equation so that we can write:

$$y_{mi\tau} = \Upsilon_m\left(\boldsymbol{Z}_{mi\tau}, \boldsymbol{\delta}_m, \varepsilon_{mi\tau}\,|\,y_{-mi\tau}\right), \tag{15}$$

where $m = 1\ldots6$, denotes the equation of each endogenous variable, *i.e.*, $y_{1i\tau} = \lambda_{i\tau}$, $y_{2i\tau} = \phi_{i\tau}$, $y_{3i\tau} = \bar{t}_{i\tau}$, $y_{4i\tau} = \gamma_{i\tau}$, $y_{5i\tau} = c_{i\tau}$, and $y_{6i\tau} = \zeta_{i\tau}$. Functions $\Upsilon_m(\cdot)$ specify different behavioral relationships between each *core* parameter and observable market characteristics. In our econometric specification we assume that all $\Upsilon_m(\cdot)$ are linear functions of market specific characteristics and some other core parameters, $y_{-mi\tau}$. In particular, as discussed above, we assume that the realization of marketing costs, ζ, conditions the value of estimated parameters λ, ϕ, $\bar{t}$, γ, and c as we use different moments to estimate these parameters depending on the number of tariff options offered. All *core* parameters condition the number of tariff options to be offered through their effect on the profitability of each tariff option. In particular we write our model as:

$$y_m = \theta_{m6}\zeta + \boldsymbol{Z}_m\boldsymbol{\delta}_m + \varepsilon_m, \qquad m = 1,\ldots,5, \tag{16a}$$

$$-\zeta = \sum_{j=1}^{5}\theta_{6j}y_j + \boldsymbol{Z}_6\boldsymbol{\delta}_6 + \varepsilon_6. \tag{16b}$$

Observe that the sequential profit maximization process described in (12)-(14) determines the features of the structural form to be estimated. First, the number of tariff options offered is larger the higher are the expected foregone profits. Thus *core* parameters that determine these foregone profits of not adding an additional tariff option, $(\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau})$, enter the equation of the negative of commercialization costs, $-\zeta_{i\tau}$, that is directly linked to the number of options offered. The effect of other structural variables on the number of tariff options is captured by $\theta_{61}, \ldots, \theta_{65}$. Similarly, the number of tariff options conditions how we estimate the core parameters $(\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau})$, and its effect is captured by $\theta_{16}, \ldots, \theta_{56}$.[24] Vector

[24] This is simply a system of equations where $(\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau})$ represent orthogonal dimensions of the screening problem whose estimates are all conditioned on the actual number of tariff options offered by the monopolist. Thus, this is a system of equations

$\boldsymbol{\delta}_m$ measures the effect of observable firm and market characteristics, $\boldsymbol{Z}_{i\tau}$,[25] and $\varepsilon_{i\tau}$ captures the effect of any relevant but unavailable information and/or any misspecification error (and whose components could potentially be correlated). These misspecification errors might indeed arise if demand is not linear, the industry does not operates under constant returns to scale, or more interestingly in our framework, if the true screening model includes more than two-dimensional types, or if within the family of two-dimensional type pricing mechanisms the effect of the outside option x enters non-additively into consumers' utility.

4.2 First Stage: Recovering Core Parameters

Estimation proceeds in two stages. In the first stage we address each local cellular telephone market in isolation and recover a set of values for the parameters that characterizes the *RS* solution. In a second stage, we make use of the variation of these structural estimates across markets to identify the effects of certain economic and demographic data on the set of core parameters, many of which identify magnitudes that are not readily observable to the econometrician such as marginal costs, or the distribution of consumers' types.[26]

Estimating the model market by market is reasonable unless consumer decisions were correlated across different cities. However, arbitrage across markets was not feasible because of the elevated roaming charges to make or receive calls from other cities. Thus consumers could only subscribe to the cellular service offered by the local monopolist, but they will not subscribe in an adjacent market. Thus, the pricing decisions of monopolist, as for screening consumers, will be independent in all markets that they might be present.[27] Furthermore, most regressors of the system of equations (16a)-(16b) are exogenous and thus the consistent estimates of Table 4 can be given a causal interpretation.

How do we compute the values of these *core* parameters? Consider first the case of those (numerous) markets where only a two-part tariff is offered. We need to identify five core parameters of the *RS*

---

that includes a discrete choice one to account for the fact that ζ is not observable. The estimation of such system of equations dates back to Amemiya (1978) and we summarize its implementation for the present paper in Appendix B.

[25] $\boldsymbol{Z}_{i\tau}$ is an $N \times K$ matrix that includes N observations of K all exogenous market variables available to us and whose descriptive statistics are summarized in Table 2, while matrices $\boldsymbol{Z}_{mi\tau}$ in equation (15) indicates only those K_m exogenous variables that directly affect the behaviour of each core parameter.

[26] A single-stage estimation is unnecessarily complicated since structural *core* parameters $\lambda_{i\tau}$, $\phi_{i\tau}$ $\bar{t}_{i\tau}$, $\gamma_{i\tau}$, $c_{i\tau}$, and $\zeta_{i\tau}$ fully capture the features of each market. Thus, as shown by Chamberlain (1982), the second stage simply map the estimated parameters in the first stage onto a lower dimensional space through a minimum distance estimator defined by the exclusion restrictions of the behavioral model presented in Section 4.3. Recovering structural parameters in a first stage to exploit the cross-market variation of demographics to explain their behavior is similar to the procedure used by Crawford and Shum (2006) in a discrete type screening model, and no different from the recovering of the mean utility level in the model of competition with differentiated products of Berry, Levinsohn, and Pakes (1995).

[27] Busse (2000) claims that after the entry of the second firm, cellular duopolist made use of tariff design to coordinate pricing across markets. This effect is not present in monopoly markets, and although firms offered similar tariffs when they served geographically closed markets it would not be possible to distinguish whether these tariffs were similar in order to minimize commercialization costs —Shew (1994)— or because consumer characteristics did not differ much, *i.e.*, along the N.E. corridor as Murray (2002) claims.

model: λ, ϕ, $\bar{t}$, γ, and c. Evidently, a two-part tariff only provides with two pieces of information: the fixed fee A_1 and the rate per minute of airtime b_1. The computed values of the parameters need to predict the actual monthly fee and marginal rate offered in each market as those that maximize profits within the family of ν-part tariff options:

$$\frac{\partial \pi \left(A_1, \ldots, A_\nu, b_1, \ldots, b_\nu \mid \lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau}\right)}{\partial A_j} = 0, \qquad j = 1, \ldots, \nu, \tag{17a}$$

$$\frac{\partial \pi \left(A_1, \ldots, A_\nu, b_1, \ldots, b_\nu \mid \lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau}\right)}{\partial b_j} = 0, \qquad j = 1, \ldots, \nu, \tag{17b}$$

In the case of a single two-part tariff we still need to add at least three more conditions to be able to compute the parameters of interest. A first condition that we use is to require that the model not only predicts the tariff properly, but also the market penetration:

$$\int_{\underline{t}}^{\bar{t}} \left[1 - \exp(-\frac{u(t)}{\phi})\right] \frac{1}{\lambda(\bar{t} - \underline{t})} \left(\frac{\bar{t} - t}{\bar{t} - \underline{t}}\right)^{\frac{1}{\lambda} - 1} dt = \text{COVERAGE}, \tag{18}$$

where $M(u, t)$ is the model prediction of unconditional market participation as given by equation (5), and where COVERAGE is our available measure of market penetration for each *SMSA*and time.

The second condition that we use identifies the maximum type $\bar{t}$ out of the arbitrarily chosen maximum consumption level $q_{\max} = 500$. Thus:

$$\bar{t} = b_\nu + \gamma \cdot 500, \tag{19}$$

where, in our case, $\nu = 1, 2, 3$ depending on whether the monopolist offers one, two or three optional tariffs.

Finally, we require that the average bill equals \$100.00 a month.[28] This number partially overcomes our lack of information about the distribution of individual consumption, thus generating meaningful economic predictions. We therefore require that:

$$\int_{\underline{t}}^{\bar{t}} P[q(t)] \left[1 - \exp(-\frac{u(t)}{\phi})\right] \frac{1}{\lambda(\bar{t} - \underline{t})} \left(\frac{\bar{t} - t}{\bar{t} - \underline{t}}\right)^{\frac{1}{\lambda} - 1} dt = 100. \tag{20}$$

We rationalize the position and shape of the tariffs offered by each monopolist into some metrics. We use the *RS* model conditional on the actual number of two-part tariffs offered in each market and time

[28] The 2002 Semiannual Wireless Survey of the Cellular Telecommunications & Internet Association (CTIA) indicates that for 1988 the average monthly bill for cellular service reached its historical peak at \$98.02. We conducted an extensive search of additional sources of bill information detailed at market level. However, such information, initially collected by CTIA was aggregated at national level and the original detailed data was destroyed because of confidentiality concerns. Our bill data is exactly the same information used by Hausman (2002), and while incomplete, it is unfortunately the best one available.

Table 3: Core Parameters of the *RS* Model

	First Quarter	*Last Quarter*	*Both Quarters*
ONE Option			
λ	0.1688	0.1997	0.18.8
ϕ	56.1254	59.8947	58.4136
$\bar{t}$	3.0998	2.3811	2.7191
γ	0.0054	0.0040	0.0047
c	0.0000	0.0000	0.0000
Observations	29	27	56
TWO Options			
λ	0.0342	0.0355	0.0354
ϕ	69.5234	66.1165	69.5234
$\bar{t}$	23.1224	21.2760	22.7053
γ	0.0454	0.0418	0.0446
c	0.0000	0.0000	0.0000
Observations	13	12	25
THREE Options			
λ	0.0171	0.0242	0.0241
ϕ	58.5320	45.2287	45.8759
$\bar{t}$	59.5089	33.1153	33.1277
γ	0.1183	0.0655	0.0656
c	0.0000	0.0000	0.0000
Observations	5	9	14

Median of the empirical distribution of computed core parameters.

to generate such mapping. Each optional tariff adds two more conditions as those in (17a)-(17b), thus exceeding the number of core parameters to compute. We therefore minimize the square of the sum of the errors implicitly defined by (17a)-(20) in order to compute the value of these core parameters independently for each market by means of a simulated annealing process.[29] Thus, conditional on some unobserved market-time fixed commercialization cost, ζ, that determines the optimal number of tariff options to offer, the vector of *core* parameters $\left(\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau}, \zeta_{i\tau}\right)$ fully characterizes market i at time τ. The median values of the computed core parameters are reported in Table 3, and they are appealingly intuitive:

- Very small value of λ. This is to be expected for the early period of this study as subscribing was expensive. The proportion of high valuation customers is low not only because of the high price of airtime but also because of the astronomical cost of the cellular telephone set. This share of high valuation customers increases over time.

- Increasing price sensitivity. As time goes by γ becomes smaller, thus capturing that as the cellular service becomes more affordable consumers increase their usage significantly.

[29] The implementation of the simulated annealing algorithm is standard and follows Goffe, Ferrier, and Rogers (1994).

- Screening more precisely. Monopolists appear to increase the number of tariff options depending on the relative heterogeneity of the distribution of consumer types and the magnitude of their willingness to pay for cellular telephone service. Thus, relative to the one-option case, monopolists offer three options when the distribution of consumer types is extremely concentrated around few consumers with very high valuation of the service. Offering more options allows monopolist to capture additional rents from other consumers with lower valuation instead of simply excluding them.

- Negligible marginal cost. This is the expected result in an industry where fixed cost and those of increasing capacity are important, but whee transmission does generate little cost other then establishing connection to the network at the switch station.

- Small commercialization costs. *Include here a brief comment on our interval estimates of ζ (include them in the table as well) after we compute the foregone profits of offering $n-1$ and $n+1$ tariff options...*

4.3 Second Stage: Behavioral Model

Economic theory provides little guidance on what demographics should be behind parameters indexing the distribution of asymmetric information. Thus, we attempted several specifications with some intuitive economic appeal and chose the one presented in Table 4 because it produced stable estimates and did not reject the exclusion restrictions highlighted in Section 4.1.[30]

We postulate that demand related variables such as COMMUTING, POP-AGE, POVERTY, EDUCATION, HHSIZE, and INCOME enter into the equations of all demand related parameters of the model: λ, ϕ, $\bar{t}$, and γ. In addition, other regressors enter exclusively in one or few of the equations of some of these demand variables. Similarly, scale — measured by TCELLS— and cost variables such as WAGE, OPERATE, and PRIME enter the specification of marginal and commercialization costs.

Parameter λ captures the degree of vertical heterogeneity among consumers valuations of cellular service and is directly related to the proportion of high valuations customers in the population. We find that λ increases with TIME and with the number of tariff options offered, $(\propto -\zeta)$. This makes sense with the negative effect of INCOME if higher income goes together with more heterogeneous valuations of the service. This result was hinted in Table 3. Higher income allows lower valuation customers to participate, thus reducing the proportion of really high valuation customers among active ones. Given this increased

[30] Evidently, we do not claim that this is "the specification" that should explain similar empirical implementations of a *RS*-based model of screening. We acknowledge that these results might not be robust since we only have 95 observations of tariffs across markets and time. But we still report these results because we feel they convey a compelling explanation and furthermore shows how using demographic information across markets to identify the effects of these variables in a simple manner and thus extend the analysis to out-of-sample observations.

Table 4: Second Stage Estimates

	λ	ϕ	$\bar{t}$	γ	c	PLANS($-\zeta$)
CONSTANT	0.3220 (4.25)	57.9929 (2.11)	−160.1893 (3.32)	−0.0915 (1.31)	0.0026 (0.48)	3.2557 (1.98)
TIME	0.0664 (4.84)	−18.0707 (3.48)	96.6373 (4.82)	−0.0686 (5.00)	−0.0036 (5.12)	1.5874 (4.49)
INCOME	−0.0034 (4.12)					
COMMUTING	0.0010 (1.20)					
POP-AGE		−3.3076 (8.68)				
POVERTY		2.6824 (4.95)				
EDUCATION			28.6150 (2.74)			
COVERAGE			20.8253 (0.68)			
RAIN		−1.4384 (2.05)				
HHSIZE				0.0378 (2.34)		
sdv(HHSIZE)				−0.0726 (2.11)		
DENSITY				0.0005 (3.13)		
TCELLS					−0.0000 (4.50)	
WAGE						0.0479 (3.47)
OPERATE					0.0004 (4.99)	0.0368 (1.39)
PRIME					−0.0009 (3.29)	−0.7529 (0.17)
BELL						0.2412 (5.68)
λ						−11.5537 (9.85)
ϕ						0.0148 (12.90)
$\bar{t}$						0.7053 (0.47)
γ						−340.1889 (0.45)
c						−82.8160 (5.99)
PLANS($-\zeta$)	−0.0509 (13.50)	9.3065 (4.50)	−11.5649 (1.99)	0.0311 (11.93)	0.0009 (4.40)	
MW-R^2	0.9988					
Observations	95					

Generalized least square, system, minimum distance estimates with absolute-value, t-statistics reported between parentheses. MW-$R^2$is the system, moment weighted, R^2 as defined in equation (2.3.16) of Judge, Griffiths, Hill, Lütkepohl, and Lee (1985).

heterogeneity among the customer base, sellers maximize their profits by offering more tariff options to their diverse clientele.

As times goes by the value of the outside option that justifies not signing up for the cellular service decreases, thus prompting the growth in the number of subscribers. Parameter ϕ also decreases with the population age but increases with the share of population under the poverty line. Bad whether, measured as rainfall precipitation, also appears to ease the subscription decision.[31]

As the industry matures, the maximum willingness to pay for cellular service, $\bar{t}$, also increases, thus reinforcing the heterogeneity argument on the effect of INCOME on λ. Parameter $\bar{t}$ also increases with education but fails to increase with COVERAGE. Market penetration is very limited in these early years and it fails to generate substantial network externalities.[32]

[31] Climatology and location effects on the decision to subscribe to fixed local telephony has been documented by Crandall and Waverman (2000) and Riordan (2002, §2). The admittedly weak economic rationale is that people living in more inhospitable climates may find less enjoyable to look for a pay-phone while away from home or the office, thus triggering a greater demand for cellular telephones.

[32] We tried many other specifications and COVERAGE always failed to be statistically significant.

Parameter γ is inversely related to the average airtime usage that therefore increases over time, but decreases (nonlinearly) with the averages size of households and the population density of the market.

The equation of marginal costs c includes cost related variables such as , OPERATE, PRIME, as well as the number of antennae deployed, TCELLS, to control for potential effects of the scale on the unit costs of production. There appears to be both small static and dynamic economies of scale as c decreases both with TIME and TCELLS. Marginal costs respond positively to operating cost indicators and only the effect of the prime lending rate is counterintuitive.

Finally, we have to address the determinants of the number of tariff options offered, which are inversely related to commercialization costs ζ. Carriers from the BELL system, the wide majority of incumbents, are more prone to offer several tariff options. This might be due to the BELL companies aiming to obtain licenses in markets with a more heterogeneous customer base; previous experience in pricing fixed line telephony; or as Shew (1994) argues, an explicit attempt to avoid the constraints of future regulatory review by initially introducing several tariff options.[33] Similarly, more tariff options are offered when the distribution of valuations is less concentrated around high values (smaller λ); the larger the expected outside option; and the smaller the marginal cost. The only unexpected result is the negative effect of WAGE.

5 The Welfare of Alternative Constrained Pricing

If a monopolist has full information, *i.e.*, if he could observe the types of each individual consumer, a personalized nonlinear pricing solution will be efficient. If the minimum consumer valuation exceeds the marginal cost of production, monopoly profits would reach the maximum possible level and equal total welfare, while consumer surplus is zero. In an environment of full information, nonlinear pricing clearly dominates the welfare performance of linear pricing unless the uniform price is competitive and coincides with marginal cost c, *i.e.*, the first best (*FB*) solution. However, beyond this very limited scenario, little is known about the relative performance of different pricing mechanisms to screen a population of heterogeneous consumers whose types remain private information.

Consider a standard nonlinear pricing mechanism *vs.* a uniform price strategy. Under very general conditions the optimal tariff leads to quantity discounts in order to separate large from small customers. Thus, large consumers will purchase more than under uniform pricing and the contrary will be true for small customers. In principle, the larger the fraction of high valuation consumers is, the closer this nonlinear

[33] Initially almost half of the markets in the present sample were regulated, even though a competitive environment was envisioned. Still, as there were no references to judge the performance of this new market, regulators simply approved all tariffs initially submitted with the idea of evaluating their performance in later periods. This might have led to a proliferation of tariff plans in this initial stage of the market in an attempt to preempt the effect of potentially restrictive future rate reviews. Price regulation was never seriously enforced in this industry and after few years eventually disappeared.

tariff is to the efficient pricing solution. However, it might also be optimal to exclude low valuation consumers altogether, and thus, the welfare comparison of a nonlinear tariff relative to a uniform price becomes ambiguous and subject to severe nonlinearities. The lack of robust theoretical predictions on this matter and the difficulty to compute general solutions of different pricing mechanisms turns this comparison into an essentially empirical question that has, so far, attracted little attention.[34] In this section we study these issues systematically. In addition we evaluate the welfare effects of ignoring a second source of asymmetric information and check whether our results are robust to the parametric assumptions needed to estimate the present version of the model using such a limited information.

5.1 Constrained Tariffs

Sometimes, and mostly attending to fairness claims, price discrimination is ruled out altogether. But in most cases, firms choose how to implement their pricing mechanism attending to screening costs considerations and common practices in the industry. Table 5 compares the welfare, profits, and market penetration induced by a set of constrained pricing strategies with the most general nonlinear tariff computed in Section 3.2. The set of constrained tariffs includes: marginal cost pricing (*FB*), uniform pricing, flat tariff, the optimal two-part tariff, and a Coasian two-tariff where the price per additional unit equals the marginal cost of production. It should be noted that these computations do not take the share of active consumers as given but they rather account for the different subscription level induced by each type of tariff.

Table 5 summarizes the most important results of this paper. We report the performance of six tariffs as measured by their profitability, welfare, coverage, consumption, and distortion needed to screen consumers due to the existence of asymmetric information. For the sake of completeness we report the median values of all these items distinguishing among markets depending on the number of tariff options offered. In Table 5 monthly fee and rate per minute are measured in dollars while expected profits and welfare are measured in percentages (100%=1) relative to the fully nonlinear tariff case (ignoring marketing costs) and the *FB* solution, respectively.[35]

Both, monthly fees and usage rates are reasonable for this time period and not very different in magnitude from those reported in Table1 In markets with more options, where the distribution of t is more spread (lower λ) monthly fees are lower in order benefit from the participation of consumers of lower

[34] Miravete (2006) studies the foregone profits and welfare of implementing a standard nonlinear tariffs by a monopolist only approximately by means of a variable number of self-selecting, optional, two-part tariffs within the *MR* framework. Miravete (2005) evaluates the relative performance of sequential pricing mechanisms relative to standard, nonlinear, monopoly pricing, where the monopolist may offer either a menu of two-part tariffs or a menu of fully nonlinear options. Finally, Miravete and Röller (2004) evaluate the welfare and profit performance of nonlinear tariffs *vs.* two-part tariff, flat tariff, and uniform pricing in duopoly.

[35] WELFARE(*) indicates the money value in million of dollars for a market with mean potential customer base of about 400.000 customers.

Table 5: Welfare, Usage, and Market Penetration

	First Best	*Nonlinear*	*Flat*	*Linear*	*Two-Part*	*Coasian*
ONE Option						
MONTHLY FEE	0.0000	− − −	63.8000	0.0000	5.4891	63.7387
RATE PER MINUTE	0.0000	− − −	0.0000	0.3241	0.2897	0.0000
PROFITS	0.0000	1.0000	0.8578	0.9895	0.9953	0.8578
WELFARE	1.0000	0.7050	0.6627	0.7072	0.7038	0.6657
WELFARE (*)	17.2993	12.1694	11.4663	12.2239	12.1641	11.4663
COVERAGE	0.2898	0.1168	0.0824	0.1268	0.1145	0.0826
AIRTIME USAGE	162.0358	161.6836	249.2312	144.6096	165.0389	249.2312
UNDERSUPPLY	1.0000	0.4000	0.4465	0.3951	0.3966	0.4469
TWO Options						
MONTHLY FEE	0.0000	− − −	45.2000	0.0000	0.5271	45.2000
RATE PER MINUTE	0.0000	− − −	0.0000	0.8267	0.8176	0.0000
PROFITS	0.0000	1.0000	0.8262	0.9994	0.9996	0.8262
WELFARE	1.0000	0.6733	0.6327	0.6758	0.6739	0.6327
WELFARE (*)	6.4561	4.3469	4.0780	4.3547	4.3541	4.0743
COVERAGE	0.1557	0.0583	0.0390	0.0596	0.0580	0.0390
AIRTIME USAGE	39.7798	39.8167	65.7041	38.8167	39.9175	65.7041
UNDERSUPPLY	1.0000	0.3746	0.4155	0.3742	0.3740	0.4155
THREE Options						
MONTHLY FEE	0.0000	− − −	25.8000	0.0000	0.2360	25.8000
RATE PER MINUTE	0.0000	− − −	0.0000	0.7418	0.7328	0.0000
PROFITS	0.0000	1.0000	0.8247	0.9996	0.9997	0.8247
WELFARE	1.0000	0.6638	0.6220	0.6641	0.6640	0.6220
WELFARE (*)	3.2535	2.2333	2.0962	2.2361	2.2341	2.0954
COVERAGE	0.1298	0.0487	0.0331	0.0504	0.0486	0.0331
AIRTIME USAGE	28.5731	28.6058	47.5176	28.0778	28.5837	47.5176
UNDERSUPPLY	1.0000	0.3725	0.4124	0.3719	0.3720	0.4124

Values for the median parameter estimates. Variables explained in the text.

valuation. Since the estimates of marginal cost are almost nil, marginal rates of the Flat and Coasian tariffs are almost identical.

The airtime usage per subscriber for the optimal two-part tariff in markets who originally offered just a two-part tariff is about 165 minutes. This usage level is almost identical to the average cellular usage reported by Hausman (2002) for 1992. Consumption is substantially larger if airtime is not priced as in the case of a flat tariff. But more interestingly is the significant reduction in airtime usage when we compare markets where more tariff options were offered. Remember that in these markets, participating consumers had a lower valuation of the service and thus, while profits always increase (although only slightly in our case) when adding tariff options, most of them are generated by fixed charges.

There is efficiency at the top both in the *MR* and the *RS* model, *i.e.*, only the highest consumer type is efficiently priced. All others consume less than under the *FB* solution due to the existence of asymmetric information. Since we do not observe individual subscribers we have need to define some

aggregate measure of the undersupply of airtime in each market when using different tariffs. In Table 5 undersupply indicates the median of the ratio of total minutes sold in a market (average airtime usage $\times$ coverage) under each tariff relative to the sales with the *FB* solution. Thus, as the number of tariff options increases from uniform to two-part tariff and then to the fully nonlinear tariff, the distortion gets reduced as these mechanisms target heterogeneous consumers more accurately. Overall, there is about a 40% of undersuply due to the existence of asymmetry of information.

As argued by Miravete (2006) and Wilson (1993, §6.4) simple tariffs accrue most of the potential profits of complex nonlinear pricing strategies. We find that simple pricing strategies capture the vast majority of potential profits in markets characterized by a very small proportion of high valuation customers and with a framework where firms have to reduce their markups in order to balance the rent extraction from high valuation customers with the probability of inducing participation to medium and low valuation ones. In such environment there is limited gains awaiting to nonlinear pricing: uniform pricing already ensures 98% of all potential profits and implementing the optimal two-part tariff only adds, at most, less than a 1% of potential profits.

This high concentration of the distribution $F(t)$ around $\underline{t}$ and the sharp reduction in consumption as soon as monopolist incorporate low valuation customers to their clientele leads to one of the most interesting results of the paper: from a welfare perspective uniform *linear* pricing dominates any form of nonlinear pricing. Still, the difference with fully *nonlinear* pricing is minimal and might be the consequence of an extreme concentration of the distribution around low valuation customers. If intermediate types were more numerous, their participation would go hand in hand with a larger airtime usage and thus, the positive effect on consumer surplus of a reduction in marginal rates would tip the balance of welfare in favor of nonlinear tariffs.

5.2 Rochet–Stole *vs.* Mussa–Rosen

In this paper we assumed that participation decision were not necessarily correlated with the valuation of the product. This is what distinguishes *RS* from the *MR* approach. How important is that distinction? Estimates of ϕ in Table 3 are far larger than zero and thus, results should be significantly different when a single type dimension ranks consumption and participation decision of consumers.[36] However, we can go a step further and compute the implied coverage, optimal nonlinear tariff, and airtime usage when $\phi = 0$ to answer the following questions: What is the welfare effect of ignoring a second source of asymmetric information? Which type dimension, t or x is responsible for a larger share of the inefficiency induced by asymmetric information?

[36] Indeed, the estimates of *core* parameters of Miravete (2006) are obtained just under the constraint that $\phi = 0$.

Table 6: Sensitivity Analysis

q_{max}	450	500	550	600	700
λ	0.1354	0.1184	0.1056	0.0856	0.0856
ϕ	52.0640	56.1254	62.5620	60.9582	66.2881
$\bar{t}$	4.3922	5.3758	5.3868	7.5619	6.9143
γ	0.0089	0.0101	0.0092	0.0119	0.0095
c	0.0000	0.0000	0.0000	0.0000	0.0000
a	0.80	0.90	1.00	1.10	1.20
λ	0.1111	0.1152	0.1184	0.1230	0.1275
ϕ	82.0029	68.1724	56.1254	50.7121	44.0909
$\bar{t}$	4.9364	4.9311	5.3758	5.0633	4.3145
γ	0.0091	0.0091	0.0101	0.0094	0.0079
c	0.0000	0.0000	0.0000	0.0000	0.0000
ρ	0.50	0.75	1.00	1.25	1.50
λ	......	......	0.1184	......	......
ϕ	......	......	56.1254	......	......
$\bar{t}$	......	......	5.3758	......	......
γ	......	......	0.0101	......	......
c	......	......	0.0000	......	......
ω	0.50	0.75	1.00	1.25	1.50
λ	......	......	0.1184	......	......
ϕ	......	......	56.1254	......	......
$\bar{t}$	......	......	5.3758	......	......
γ	......	......	0.0101	......	......
c	......	......	0.0000	......	......

Median of the empirical distribution of computed core parameters.

To be completed...

5.3 Sensitivity Analysis

A major advantage of our approach is that it does not requires individual consumption data to estimate the importance of asymmetric information when firms engage in multidimensional screening. The drawback is that results rely heavily on functional form assumptions. In this section we introduce some flexibility in the formulation of the distribution of the vertical type, demand and cost functions, as well as the assumed maximum consumption possible in order to calibrate how important are these assumptions and whether they compromise the robustness of our results.

We first modify the maximum quantity purchased, q_{max}. In equation (19) we simply assumed that $q_{max} = 500$. We chose this level because it was sufficiently above the documented average monthly usage at this period and because for this usage level, consumers will always pay according to the "last tariff option," the one with highest fixed monthly fee and lowest rate per minute. First top section of Table 6 evaluates the solution at $q_{max} = \{450, 500, 550, 600, 700\}$. Marginal cost is independent of q_{max} and again always zero. The

rest of *core* parameters vary mostly monotonically but substantially. This result points out of the importance of accurately identifying the relevant range of maximum consumption to obtain precise welfare estimates. Although quantitative results are highly dependent on the assumed value of $q_{\max}$, qualitative ones, the relative performance of different tariffs still remain valid because all *core* parameters vary monotonically with $q_{\max}$.

The second extension deals with $F(t)$. We now generalize the model by assuming that the vertical type dimension, t, is distributed according to a beta distribution with parameters a and λ^{-1}, that is:

$$F(t|a,\lambda^{-1}) = \frac{\Gamma(a+\lambda^{-1})}{\Gamma(a)\Gamma(\lambda^{-1})} \cdot \frac{(t-\underline{t})^{a-1}(\bar{t}-t)^{\lambda^{-1}-1}}{(\bar{t}-\underline{t})^{a+\lambda^{-1}-2}}, \tag{21}$$

which coincides with the assumed Burr Type XII distribution of equation (3a) when $a = 1$. For values of a smaller than 1 the distribution of t is even more concentrated around $\underline{t}$ while for values of a larger than 1 the distribution gains a substantial mass of probability for intermediate values of t. The second section of Table 6 indicates that λ, $\bar{t}$ and c are quite robust to changes in a. Parameters ϕ and γ vary a little more significantly but proportionally less than with changes of $q_{\max}$. However, the most important result is that λ is very stable to changes in a. This means that in general the distribution of the vertical taste parameter t is quite well approximated by the assumed Burr Type XII distribution of equation (3a).

Next, we generalize the utility function (2) to:

$$v = \frac{t^{\rho+1}}{\gamma(\rho+1)}\left[1-\left(1-\frac{\gamma}{t}q\right)^{\rho+1}\right] - P(q) - x, \tag{22}$$

which leads to a convex (concave) demand function when $\rho > 1$ ($\rho < 1$) and coincides with the linear demand function of the *RS* model when $\rho = 1$. If demand is convex (concave), the *RS* model will overestimate (underestimate) welfare effects as consumers' willingness to pay for the service are smaller (larger) than with a linear demand function.[37] *The third section of Table 6 shows...*

Finally, we modify the total cost function to:

$$TC(q) = \frac{c}{1+\omega}q^{1+\omega}, \tag{23}$$

so that the monopolist enjoys increasing, constant, or decreasing returns to scale depending whether ω is less than, equal, or larger than 1, as marginal cost is now cq^{ω}. Values of ω larger than one may capture the effect of capacity limits while when we assume $\omega < 1$ we are considering that learning and scale effects are more important. *The last section of Table 6 shows...*

[37] In this case, the *IC* condition (9b) becomes $\dot{u}(t) = (t^{\rho} - (t-\gamma q(t))^{\rho})/\gamma$.

6 Policy Evaluations: Enforcing the Universal Service Requirement

The meaning of universal service in fixed telephony has changed quite substantially over the past century. Theodore Vail, CEO of AT&T, first used the term *universal service* in AT&T's 1907 Annual report to support the idea of a single nationally integrated telephone service. Universal service was the declared strategy of a dominant carrier to eliminate competitors with incompatible and not interconnected systems until the 1920s. In later decades, universal service meant a deliberate policy of underpricing local residential connections by overpricing long-distance calls in order to ease access to residential customers. To achieve this goal, business were also required to pay more for local connections than residents.[38] Thus, universal service went together with cross-subsidization among the different line of services that ended with the divestiture of AT&T in 1984.[39]

Although cellular telephony has not been subject to universal service policies, other competitive services such as broad-band have been targeted as essential services worth subsidizing to ensure access to internet to most of the population and avoid the feared *digital divide*. As Crandall and Waverman (2000, §8) document, the 1996 U.S. Telecommunications Act ensures the subsidized access to internet for schools, libraries and rural health facilities at an estimated cost of \$2.65bn a year, far more than the traditional support for universal service in fixed telephony. The analysis that follows shows the hypothetical effects of implementing a universal service requirement in the early stages of development of the cellular telephone industry while *simultaneously* accounting for the unobserved heterogeneity of consumers and the effect that changes in tariffs have on subscription decisions.

Our monopoly model predicts optimal exclusion at the bottom. This feature of the model is also a very convenient because although market penetration of fixed telephony is over 90% it never reaches full market coverage. We then use the model to provide with two alternative ways to implement and measure the effects of the universal service requirement:

1. We first compute a balanced-budget pricing solution that maximizes market coverage while the monopolist makes non–negative profits, $\mathbb{S} \geq 0$. Thus, using the structural parameters from each tariff, we solve the following problem for each market and time:

$$\max_{q(t),u(t)} \int_{t^\circ}^{\bar{t}} \left[1 - \exp(-\frac{u(t)}{\phi})\right] \frac{1}{\lambda(\bar{t}-\underline{t})} \left(\frac{\bar{t}-t}{\bar{t}-\underline{t}}\right)^{\frac{1}{\lambda}-1} dt, \tag{24a}$$

[38] Muller (1997) discusses in detail the origin, history, and development of the universal service policy in fixed telephony in the U.S. and Crandall and Waverman (2000, §1) summarize recent developments of the universal service policy.

[39] For a definition of subsidy-free pricing see Faulhaber (1975). Riordan (2002, §4.1) briefly summarizes the discussion on the cross-subsidization generated by universal service policies in fixed telephony.

Table 7: Average Effects of Universal Service Policy

	Balanced	*Free*
ONE Option		
WELFARE	0.7901	1.0000
WELFARE (*)	14.1584	17.3188
COVERAGE	0.3952	0.2922
AIRTIME USAGE	68.9520	159.3380
UNDERSUPPLY	0.5999	1.0000
TWO Options		
WELFARE	0.5908	1.0000
WELFARE (*)	16.2874	39.7798
MARKET PENETRATION	0.1835	0.1557
AIRTIME USAGE	16.2874	39.7798
UNDERSUPPLY	0.4653	1.0000
THREE Options		
WELFARE	0.8689	1.0000
WELFARE (*)	2.8235	3.2535
MARKET PENETRATION	0.1797	0.1298
AIRTIME USAGE	14.0731	28.5731
UNDERSUPPLY	0.6874	1.0000

Median sample values across markets. Welfare and market penetration is reported as percentage of the profits of the fully nonlinear tariff solution. The rest of variables are defined as in Table 5.

$$\mathbb{S} \leq \int_{t^\circ}^{\bar{t}} \pi[q(t)] \left[1 - \exp(-\frac{u(t)}{\phi})\right] \frac{1}{\lambda(\bar{t} - \underline{t})} \left(\frac{\bar{t} - t}{\bar{t} - \underline{t}}\right)^{\frac{1}{\lambda} - 1} dt, \tag{24b}$$

$$\dot{u}(t) = q(t) \geq 0, \tag{24c}$$

$$\dot{q}(t) \geq 0, \tag{24d}$$

$$u(t) = tq(t) - \frac{\gamma}{2} q^2(t) - P(t) \geq 0. \tag{24e}$$

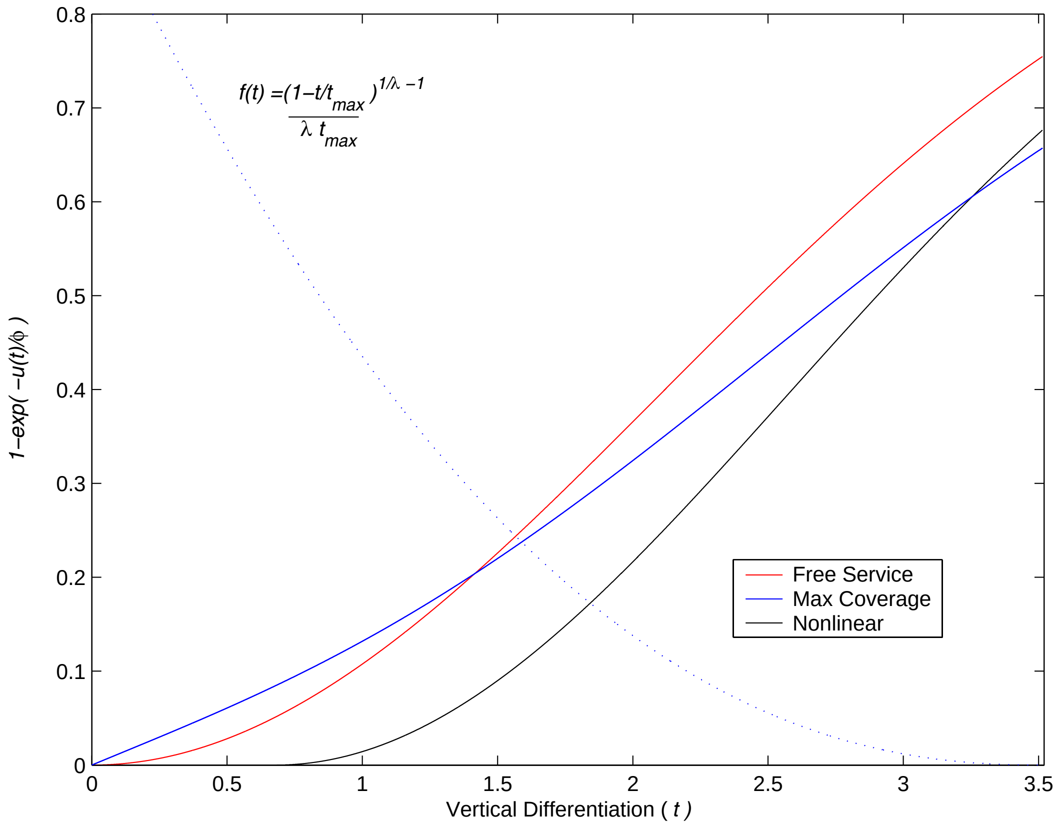
This is the best way in which our single product model can capture the idea of the actual cross-subsidization that was induced by the rate structure sponsored by the *FCC*. In 1950 the *FCC* set the criteria to divide non-traffic sensitive fixed costs of the telephone network between interstate and intrastate services. According to Crandall and Waverman (2000, §1), long-distance cross-subsidized local calling only by the amount of 3% of the non-traffic sensitive, local-network costs but by mid-1970s, the subsidy represented already 20% of such costs. Provided the distribution of vertical and horizontal heterogeneity of consumers in each market, solving problem (24a)–(24e) we can predict by how much participation would increase if the monopolist is forced to break-even, *i.e.*, when $\mathbb{S} = 0$: $M_1^{US}(u, t^\circ)$. Alternatively we could also compute the market penetration for a given level of direct subsidy from the government, $\mathbb{S}$. We understand our break-even policy as a lower bound estimate of the effect of the universal service policy.

2. The maximum market penetration possible is to ensure that anybody willing to pay a positive price for telephone service actually subscribes to it. As for our model, we have to solve a modified version of (24a)–(24e) subject to the additional constraint that $P(t^{\circ}) = 0$ regardless of whether there is full market coverage, $t^{\circ} = \underline{t}$, or not. This provides us with the new market penetration $M_2^{US}(u, t^{\circ})$ and the associated cost of providing this level of universal service, $\mathbb{S}$. We therefore view this second measure of market penetration as an upper bound estimate of the effect of universal service policy.[40]

In general, the existing studies document that income-targeted programs are not very effective in promoting subscription to fixed telephony.[41] However, all these studies take the tariff offered by the monopolist as given. Our structural approach allows us to recalculate what the optimal tariff of the monopolist would be in the presence of two alternative ways to implement the universal service policy.

Table 7 shows that both alternative ways to implement the universal policy broadens market coverage substantially. It also reduces the common distortion to screen consumers as can be seen by the important increase in the median ratio of undersupply. Welfare effects, although increasing relative to the nonlinear pricing solution, are quite different depending on whether the universal service policy is financed through a

Figure 3: $F(t)$ **— Who participates under universal service?**



that now low valuation customers are more numerous and the monopolist cannot reduce the distortion as much as under the "free" universal policy in order to ensure the *IC* constraint and separation of t types. The "balanced" policy adds more low valuation customers than the "free" one as the tariff discounts are more pronounced for them. Subscription reaches its maximum level but at the cost of reducing the average usage —as now very low consumer types become active— and thus increasing welfare relative to the nonlinear pricing solution only by a third of what the "free" universal service policy can achieve. In terms of practical policy matters this result advice against using pricing of public utilities as a way to convey any redistribution policy. It thus call into question the welfare effectiveness of traditional cross-subsidization in telecommunications (and other public utility industries) where high income customers were overcharged in long distance service to generate funding for to ensure that low income customers would subscribe to the local telephone service.

7 Summary and Conclusions

This paper contributes in several ways to the literature of empirical screening models. Most importantly, this is the first systematic study that compares the performance of different pricing mechanisms by explicitly addressing the different participation induced by each tariff strategy. Thus, our results show, not surprisingly, that nonlinear pricing has very little to increase profits over uniform or *linear* pricing in early markets. The introduction of a new and expensive product only triggers only few customers to purchase the product and thus, the equilibrium distribution of valuations is very skewed and concentrated around low types. In such circumstances it is generally more profitable to offer simple tariffs that exclude low valuation consumers than adding (costly) options to incorporate a minuscule fraction of low valuation customers

We show that allowing nonlinear pricing models with sufficient flexibility —including more than one type dimension— is important to reduce the estimation bias and the correct welfare evaluation of pricing policies. In particular if we ignore the existence of an independent outside option to consumers we show...

This paper is also the first attempt to estimate a structural nonlinear pricing model with several sources of asymmetric information. We do this relying on a minimal amount of information: the available tariff in each market and a measure of market penetration. We feel that this is a promising approach to circumvent the lack of detailed individual subscriber information, an issue that only gets aggravated when several competing firms offering nonlinear tariffs to their customers. This paper has also contributed in developing the methods that would make the estimation of such exclusive agency model feasible. That is the next step in our research agenda.

References

AMEMIYA, T. (1978): "The Estimation of a Simultaneous–Equation Generalized Probit Model." *Econometrica*, 46, 1193–1205.

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

ASCHER, U. M., R. M. M. MATTHEIJ, AND R. D. RUSSELL (1988): *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations, (Classics in Applied Mathematics, Vol. 13)*. Englewood Cliffs, NJ: Prentice-Hall.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium." *Econometrica*, 63, 841–890.

BUSSE, M. R. (2000): "Multimarket Contact and Price Coordination in the Cellular Telephone Industry." *Journal of Economics & Management Strategy*, 9, 287–320.

CHAMBERLAIN, G. (1982): "Multivariate Regression Models for Panel Data." *Journal of Econometrics*, 18, 5–46.

CHAN, H. M., G. HALL, AND J. RUST (2004): "Price Discrimination in the Steel Market." Mimeo, Charles River Associates, Yale University, and University of Maryland.

CRANDALL, R., R. W. HAHN, R. E. LITAN, AND S. WALLSTEN (2004): "Universal Broadband Access: Implementing President Bush's Vision." Mimeo, AEI-Brookings Joint Center for Regulatory Studies, Working Paper # 04-01.

CRANDALL, R. W. AND L. WAVERMAN (2000): *Who Pays for "Universal Service"? When Telephone Subsidies Become Transparent*. Washington, DC: Brookings Institution Press.

CRAWFORD, G. S. AND M. SHUM (2006): "Monopoly Quality Degradation in Cable Television." *Journal of Law and Economics*, 49, forthcoming.

ERIKSSON, R. C., D. L. KASERMAN, AND J. W. MAYO (1998): "Targeted and Untargeted Subsidy Schemes: Evidence from Post-Divestiture Efforts to Promote Universal Service." *Journal of Law and Economics*, 41, 477–502.

FAULHABER, G. R. (1975): "Cross-Subsidization: Pricing in Public Enterprises." *American Economic Review*, 65, 966–977.

GARBACZ, C. AND H. G. THOMPSON, JR (1997): "Assessing the Impact of FCC Lifeline and Link-Up Programs on Telephone Penetration." *Journal of Regulatory Economics*, 11, 67–78.

GOFFE, W. L., G. D. FERRIER, AND J. ROGERS (1994): "Global Optimization of Statistical Functions with Simulated Annealing." *Journal of Econometrics*, 60, 65–99.

HAUSMAN, J. A. (2002): "Mobile Telephone." In M. Cave, S. K. Majumdar, and I. Vogelsang (eds.): *Handbook of Telecommunications Economics*, Vol. I. Amsterdam, The Netherlands: North-Holland.

JOHNSON, N. L., S. KOTZ, AND N. BALAKRISHNAN (1994): *Continuous Univariate Distributions,* Vol. 1, 2nd edn. New York, NY: John Wiley & Sons.

JUDGE, G. G., W. E. GRIFFITHS, R. C. HILL, H. LÜTKEPOHL, AND T.-C. LEE (1985): *The Theory and Practice of Econometrics*, 2nd edn. New York, NY: John Wiley & Sons.

KARLIN, S. (1968): *Total Positivity,* Vol. I. Stanford, CA: Stanford University Press.

KATZ, M. L. (1983): "Non-Uniform Pricing, Output and Welfare under Monopoly." *Review of Economic Studies*, 50, 37–56.

KELLER, H. B. (1968): *Numerical Methods for Two-Point Boundary-Value Problems*. Waltham, MA: Blaisdell.

LAFFONT, J.-J., E. MASKIN, AND J.-C. ROCHET (1987): "Optimal Nonlinear Pricing with Two-Dimensional Characteristics." In T. Groves, R. Radner, and S. Reiter (eds.): *Information, Incentives, and Economic Mechanisms*. Minneapolis, MN: University of Minnesota Press.

LAFFONT, J.-J. AND J. TIROLE (1993): *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

LEE, L.-J. (1981): "Simultaneous Equations Models with Discrete and Censored Dependent Variables." In C. F. Manski and D. L. McFadden (eds.): *Structural Analysis of Discrete Data with Applications*. MIT Press.

MASKIN, E. AND J. RILEY (1984): "Monopoly with Incomplete Information." *RAND Journal of Economics*, 15, 171–196.

MIRAVETE, E. J. (2002): "Choosing the Wrong Calling Plan? Ignorance and Learning." *American Economic Review*, 93, 297–310.

MIRAVETE, E. J. (2005): "The Welfare Performace of Sequential Pricing Mechanisms." *International Economic Review*, 46, forthcoming.

MIRAVETE, E. J. (2006): "Are All Those Calling Plans Really Necessary? The Limited Gains From Complex Tariffs." Mimeo, University of Pennsylvania.

MIRAVETE, E. J. AND L.-H. RÖLLER (2004): "Competitive Nonlinear Pricing in Duopoly Equilibrium: The Early U.S. Cellular Telephone Industry." Mimeo, University of Pennsylvania and Wissenschaftszentrum Berlin für Sozialforschung (WZB).

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38, 175–208.

MITCHELL, B. M. AND I. VOGELSANG (1991): *Telecommunication Pricing. Theory and Practice*. Cambridge, UK: Cambridge University Press.

MULLER, M. L., JR (1997): *Universal Service*. Cambridge, MA: MIT-AEI.

MURRAY, J. B., JR (2002): *Wireless Nation. The Frenzied Launch of the Cellular Revolution in America*. Cambridge, MA: Perseus Publishing.

MUSSA, M. AND S. ROSEN (1978): "Monopoly and Product Quality." *Journal of Economic Theory*, 18, 301–317.

O'BRIEN, D. P. AND G. SHAFFER (1994): "The Welfare Effects of Forbidding Discriminatory Discounts: A Secondary Line Analysis of Robinson-Patman." *Journal of Law, Economics, and Organization*, 10, 296–318.

PARKER, P. M. AND L.-H. RÖLLER (1997): "Collusive Conduct in Duopolies: Multimarket Contact and Cross-Ownership in the Mobile Telephone Industry." *RAND Journal of Economics*, 28, 304–322.

RIORDAN, M. (2002): "Universal Residential Telephone Services." In M. Cave, S. K. Majumdar, and I. Vogelsang (eds.): *Handbook of Telecommunications Economics*, Vol. I. Amsterdam, The Netherlands: North-Holland.

ROBERTS, K. W. S. (1979): "Welfare Considerations of Nonlinear Pricing." *Economic Journal*, 89, 66–83.

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

ROCHET, J.-C. AND L. A. STOLE (2002): "Nonlinear Pricing with Random Participation." *Review of Economic Studies*, 69, 277–311.

SEIM, K. AND V. B. VIARD (2005): "The Effect of Entry and Market Structure on Cellular Pricing Tactics." Mimeo, Graduate School of Business, Stanford University.

SHEW, W. B. (1994): "Regulation, Competition, and Prices in the U.S. Cellular Telephone Industry." Mimeo, The American Enterprise Institute.

SPENCE, M. (1977): "Nonlinear Prices and Welfare." *Journal of Public Economics*, 8, 1–18.

STOLE, L. A. (2005): "Price Discrimination and Imperfect Competition." In M. Armstrong and R. H. Porter (eds.): *Handbook of Industrial Organization,* Vol. 3. New York, NY: North-Holland.

TIROLE, J. (1989): *Industrial Organization*. Cambridge, MA: MIT.

VISCUSI, W. K., J. M. VERNON, AND J. E. HARRINGTON, JR (2005): *Economics of Regulation and Antitrust*, 4th edn. Cambridge, MA: MIT Press.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, 50, 1–25.

WILSON, R. B. (1993): *Nonlinear Pricing*. New York, NY: Oxford University Press.

Appendix

A Variables Definition and Data Sources

- Tariff information is reported by *Cellular Price and Marketing Letter*, Information Enterprises, various issues, 1984–1988. TIME indicates the number of months since the first monopolist started offering cellular service.

- Socioeconomic and demographic data of each market comes from the 1989 *Statistical Abstracts of the United States*; U.S. Department of Commerce, Bureau of the Census, using the *FCC* Cellular Boundary Notices, 1982–1987, available in *The Cellular Market Data Book*, EMCI, Inc.; as well as the 1990 U.S. Decennial Census. Variables include the size of households, HHSIZE; thousands of high potential business establishments, BUSINESS;[42] the average commuting time in minutes, COMMUTING; total population of the *SMSA* in millions, POPULATION; the population density of the market (people per square mile), DENSITY; median income in thousands of dollars, INCOME; percentage of households with income below the poverty level, POVERTY; median age of population in years, POP-AGE; and median number of years of education, EDUCATION. Variables marked "$sdv(\cdot)$" indicate the within market standard deviation of the corresponding demographic.

- Industry cost indicators for each market are obtained from the Bureau of Labor Statistics; U.S. Department of Energy; *BOMA Experience Exchange Report: Income/Expense Analysis for Office Buildings*, various issues, 1985–1989; and *Cellular Price and Marketing Letter*, Information Enterprises, various issues, 1984–1988; and 1990 U.S. Census. They include the one-period lagged prime lending rate, PRIME; an index of operating expenses per square foot of office space, OPERATE;[43] and an index of average annual wages per employee for the cellular industry, WAGE.

- Weather data is available on the web at `http://cdiac.esd.ornl.org`, and includes average temperature and precipitation for 1,221 stations in the contiguous continental states plus those of Alaska.[44] Data include the average quarterly precipitation in inches, RAIN.

- Largest shareholder information is available from the *FCC*. We identify with BELL those carriers owned by firms from the former Bell system to distinguish them from independently owned carriers.

[42] BUSINESS refers to what it was considered at that time as highly potential customers by cellular industry experts: business service firms, health care, professional, and legal services, contract construction, transportation, finance, insurance, and real state.

[43] These expenses include cleaning, repair and maintenance, administrative costs, utilities, local taxes, security and ground services, office payroll, as well as other leasing expenses associated with running an office.

[44] See Easterling, D.R., T.R. Karl, EH. Mason, P.Y. Hughes, D.P. Bowman and R.C. Daniels, *United States Historical Climatology Network (U.S. HCN) Monthly Temperature and Precipitation Data*. ORNL/CDIAC-87, NDP-019/R3, 1996. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

B Estimation of the Behavioral Model

Regardless of the particular behavioral model put forward by the econometrician, *core* parameters $\lambda_{i\tau}$, $\phi_{i\tau}$ $\bar{t}_{i\tau}$, $\gamma_{i\tau}$, $c_{i\tau}$, and $\zeta_{i\tau}$ fully capture the features of the model according to the present specification of the *RS* problem. The goal now is to make these *core* parameters a function of market characteristics $\boldsymbol{Z}_m$ and other *core* parameters according to the discussion of Section 4.1 and summarized in equations (16a)-(16b). For convenience we define the $K \times K_m$ selection matrix $\boldsymbol{J}_m$ that contains arrays of 0's and 1's such that:

$$\boldsymbol{Z}_{mi\tau} = \boldsymbol{Z}_{i\tau}\boldsymbol{J}_m. \tag{B.1}$$

Since ζ is not observable, and only the number of tariff functions is available, (16a)-(16b) corresponds to a simultaneous equations model with discrete endogenous variables first considered by Amemiya (1978). The estimation of such model is straightforward and relies on imposing the identifying restrictions of substituting the reduced form parameters into the structural form (16a)-(16b).[45] Briefly, the estimation procedure is the following. We first estimate a set of reduced form equations of the endogenous variables on all K exogenous regressors:

$$y_m = \boldsymbol{Z}\boldsymbol{\beta}_m + u_m, \qquad m = 1, \ldots, 5, \tag{B.2a}$$

$$-\zeta = \boldsymbol{Z}\boldsymbol{\beta}_6 + u_6. \tag{B.2b}$$

Next, making use of (B.1) and substituting (B.2a)-(B.2b) into (16a)-(16b) we obtain:

$$y_m = \theta_{m6}\boldsymbol{Z}\boldsymbol{\beta}_6 + \boldsymbol{Z}\boldsymbol{J}_m\boldsymbol{\delta}_m + (\varepsilon_m + \theta_{m6}u_6), \qquad m = 1, \ldots, 5, \tag{B.3a}$$

$$-\zeta = \sum_{j=1}^{5}\theta_{6j}\boldsymbol{Z}\boldsymbol{\beta}_j + \boldsymbol{Z}\boldsymbol{J}_6\boldsymbol{\delta}_6 + \left(\varepsilon_6 + \sum_{j=1}^{5}\theta_{6j}u_j\right), \tag{B.3b}$$

and thus, comparing the coefficients of (B.2a)-(B.2b) and (B.3a)-(B.3b) we obtain the set of identifying restrictions that allow us to estimate the structural parameters. Equations (B.2a) are estimated by ordinary least squares while (B.2b) is estimated as a maximum likelihood ordered probit model, thus producing consistent estimates $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_6)'$ so that the identifying restrictions are evaluated as:

$$\tilde{\beta}_m = \theta_{m6}\tilde{\beta}_6 + \boldsymbol{J}_m\boldsymbol{\delta}_m + \eta_m, \qquad m = 1, \ldots, 5, \tag{B.4a}$$

$$\tilde{\beta}_6 = \sum_{j=1}^{5}\theta_{6j}\tilde{\beta}_j + \boldsymbol{J}_6\boldsymbol{\delta}_6 + \eta_6. \tag{B.4b}$$

[45] In addition to Amemiya (1978), Lee (1981) further studies how to improve the asymptotic efficiency of this generalized least square estimator.

Since estimates $\tilde{\beta}$ are consistent and asymptotically normally distributed, the error terms $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_6)$ are also asymptotically normally distributed:

$$\boldsymbol{\eta} \sim N\left[0, (\boldsymbol{\Theta} \otimes \boldsymbol{I}_K)\, V(\tilde{\beta})\, (\boldsymbol{\Theta} \otimes \boldsymbol{I}_K)'\right], \tag{B.5}$$

where $V(\tilde{\beta})$ is the covariance matrix of the reduced form estimates $\tilde{\beta}$ and $\boldsymbol{\Theta}$ is the matrix of coefficients affecting the endogenous variables defined as:[46]

$$\boldsymbol{\Theta} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -\theta_{16} \\ 0 & 1 & 0 & 0 & 0 & -\theta_{26} \\ 0 & 0 & 1 & 0 & 0 & -\theta_{36} \\ 0 & 0 & 0 & 1 & 0 & -\theta_{46} \\ 0 & 0 & 0 & 0 & 1 & -\theta_{56} \\ -\theta_{61} & -\theta_{62} & -\theta_{63} & -\theta_{64} & -\theta_{65} & 1 \end{pmatrix}. \tag{B.6}$$

The system of identifying restrictions (B.4a)-(B.4b) can therefore be written in matrix form as:

$$\tilde{\beta} = \boldsymbol{D}\boldsymbol{\theta}^\star + \boldsymbol{\eta}, \tag{B.7}$$

where:

$$\boldsymbol{D} = \begin{pmatrix} \tilde{\beta}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{\beta}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\beta}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\beta}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{\beta}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \tilde{\beta}_1 & \tilde{\beta}_2 & \tilde{\beta}_3 & \tilde{\beta}_4 & \tilde{\beta}_5 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_6 \end{pmatrix}, \tag{B.8a}$$

$$\boldsymbol{\theta}^\star = \left(\boldsymbol{\theta}_{16}, \boldsymbol{\theta}_{26}, \boldsymbol{\theta}_{36}, \boldsymbol{\theta}_{46}, \boldsymbol{\theta}_{56}, \boldsymbol{\theta}_{61}, \boldsymbol{\theta}_{62}, \boldsymbol{\theta}_{63}, \boldsymbol{\theta}_{64}, \boldsymbol{\theta}_{65}, \boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \boldsymbol{\delta}_3', \boldsymbol{\delta}_4', \boldsymbol{\delta}_5', \boldsymbol{\delta}_6'\right)', \tag{B.8b}$$

so that estimating (B.7) by generalized least squares to account for correlation among the elements of $\boldsymbol{\eta}$ provides with the system minimum distance estimator of the structural parameters:

$$\tilde{\theta}^\star = \left(\boldsymbol{D}'\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{D}\right)^{-1} \boldsymbol{D}'\boldsymbol{\Sigma}_\eta^{-1}\tilde{\beta}, \tag{B.9a}$$

$$V(\tilde{\theta}^\star) = \left(\boldsymbol{D}'\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{D}\right)^{-1}. \tag{B.9b}$$

[46] The estimation, equation by equation of the system (B.2a)-(B.2b) is not efficient because it ignores potential correlation among the different error terms $u_1, \ldots, u_6$. Thus, according to White (1982), for this potentially misspecified model, the proper covariance matrix is $V(\tilde{\beta}) = H^{-1}G'GH^{-1}$ where $H = H(\tilde{\beta})$ is a block-diagonal matrix of the Hessians and $G = G(\tilde{\beta})$ is a matrix of stacked gradients where each equation (B.2a)-(B.2b) is evaluated at $\tilde{\beta}$.