

DISCUSSION PAPER SERIES

No. 6620

CONSTRAINED MONOPOLY PRICING WITH RANDOM PARTICIPATION

Gabriel Basaluzzo and Eugenio J Miravete

INDUSTRIAL ORGANIZATION



Centre for Economic Policy Research

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP6620.asp

CONSTRAINED MONOPOLY PRICING WITH RANDOM PARTICIPATION

Gabriel Basaluzzo, Universidad de San Andrés
Eugenio J Miravete, University of Texas, Austin and CEPR

Discussion Paper No. 6620
December 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INDUSTRIAL ORGANIZATION**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Gabriel Basaluzzo and Eugenio J Miravete

ABSTRACT

Constrained Monopoly Pricing with Random Participation*

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 that are commonly used to implement the optimal nonlinear tariff to estimate how demand and cost variables affect the pricing strategies offered by incumbent monopolists in several early U.S. local cellular telephone markets through the different elements of the theoretical model: marginal costs, average price sensitivity of demand, indexing parameters governing the distribution of the two-dimensional type components, support of the distribution of types, and costs associated to the commercialization of tariff options. The sources of identification are the position and shape of each tariff offered by monopolists, the actual number and features of the tariff options used to implement them, as well as a measure of market penetration in each cellular market during the first and last quarter of monopoly regime. We use our model and the structural estimates to provide a performance comparison (profit + welfare) of nonlinear tariffs relative to linear (uniform), optimal two-part, Coasian marginal cost-plus fixed fee, and flat tariffs. We furthermore evaluate the potential welfare gains of implementing universal service requirements.

JEL Classification: C63, D43, D82 and L96

Keywords: nonlinear vs. linear pricing, random participation and universal service

Gabriel Basaluzzo
Departamento de Economía
Universidad de San Andrés
25 de Mayo 587
Buenos Aires
Argentina
Email: gbasaluzzo@udesa.edu.ar

Eugenio J Miravete
University of Texas at Austin
Department of Economics
BRB 1.116
1 University Station C3100
Austin, TX 78712-0301
USA
Email: miravete@eco.utexas.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=163844

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=139116

* 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 2005 Fundación Ramón Areces Conference on “Industrial Organization and Competition Policy,” the 2006 CEPR Conference on Applied Industrial Organization held in Madeira, as well as seminar audiences at ITAM, the Olin School of Business and universities at Auckland, Canterbury–Christchurch, Carlos III–Madrid, Harvard, Northwestern, Otago–Dunedin, Pennsylvania, Purdue, Rutgers–New Brunswick, Southern California, Texas A&M, Toronto, Valencia, Vanderbilt, Victoria–Wellington, and Wisconsin–Madison. The second author acknowledges financial support through NSF Grant SES-0318208. We are solely responsible for all errors that might still remain.

Submitted 11 December 2007

1 Introduction

The adoption of innovations and the introduction of new products and services to the market is a crucial element behind economic development. These new markets are commonly dominated by firms with significant market power which allows them not only to charge high prices but also to take advantage of complex pricing mechanisms in the presence of a heterogeneous customer base. However, these pricing mechanisms, and not only the high price of new products, have a major incidence on the participation of consumers in the market. Thus, policymakers might be interested in knowing whether regulating pricing practices may foster the participation of new consumers and, more importantly, if such policy is welfare enhancing at all.

Regrettably, 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 obtaining general results that would allow us to evaluate the advantages of nonlinear over uniform pricing. These lack of results are a direct consequence of the effect that different pricing strategies have, not only on consumption of active customers, but also on their decision to participate in the market at all. However, and despite this theoretical ambiguity, 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 such an evaluation becomes an important open empirical question that has not been addressed yet.

This paper presents a framework in which we can address a systematic comparison among different nonlinear pricing strategies, addressing explicitly the participation of consumers in the market by means of a model of nonlinear pricing with random participation that uses only the the very limited tariff information commonly available to economists and policymakers. The main finding of this paper is that once we take into account not only the changes in consumption but also participation decisions, linear pricing achieves almost the same welfare level than the optimal fully nonlinear tariff in markets where low valuation consumers are more numerous than high valuation ones. In this same environment linear pricing dominates two-part, flat, and Coasian tariffs.

The importance of considering a model of nonlinear pricing with random participation is that changes in the pricing strategy of firms will not only determine participation decisions the lower end of the distribution of consumer types but rather to all potential customers. Our model is particularly well

¹ Indeed, the 1936 Robinson-Patman Act made it 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. See O’Brien and Shaffer (1994) for further details.

suited to address the distribution and welfare effects of different pricing policies since a non-negligible effect of different pricing mechanisms, such as uniform pricing, flat tariff, two-part tariff, and a Coasian tariff (marginal cost plus a fixed fee) is to induce consumers to subscribe to the cellular service more or less frequently. This is indeed the source of important nonlinearities affecting welfare effects, and the source of most difficulties in ranking the performance of different pricing mechanisms. Thus, our framework allows us to conduct relevant policy evaluations such as measuring the costs *vs.* induced benefits of implementing a universal service requirement.

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.² 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 (2007)— or alternatively, to identify the distribution of unobserved consumer heterogeneity both under monopoly (Miravete (2007)) and duopoly (Miravete and Röller (2004)) from the shape and position of the tariff. 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 compute the value of the structural parameters that generate the closest match to the actual menu of self-selecting two-part tariffs actually offered by several local 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). The use of computing intensive methods allow us to obtain these parameters taking into account the actual implementation of the nonlinear tariff by means of few optional two-part tariffs.

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. This paper offers a way to evaluate the welfare performance of different pricing strategies taking into account consumption and participation decisions. In terms of policy evaluations we choose to focus on the widespread universal service requirement of many telecommunications markets. The basic point to be made is that the optimal tariff under the universal service requirement is not the same than if such regulatory constraint was absent. We thus make use of our model to estimate the basic structural parameters and then take them at face value to compute the optimal tariff under different policy constraints. The results of our evaluations question some commonly held views. For instance, Crandall, Hahn, Litan, and Wallsten (2004) argue that broadband subsidies financed with taxation on other broadband services are particularly damaging to achieve a higher penetration rate in this market. Our evidence suggests, on

² Most empirical applications have only made use of the consumption predictions of nonlinear pricing models while ignoring the less data demanding use of the optimal tariff function. See for instance Ivaldi and Martimort (1994) or Reiss and White (2005) among others.

the contrary, that a small proportion of few high valuation customers give up subscription to the cellular service altogether and consumption among them only reduces slightly. The raised funds allow many of the more numerous low valuation customers to subscribe to this new telephone service. We show that a balanced-budget policy effectively increases market participation. The implicit redistribution of rents lead to moderate welfare loss. But compared with the unrestricted nonlinear tariff, universal service actually enhances welfare by 13.25%. The reason is that with a very small proportion of high valuation customers, the optimal discriminatory pricing policy requires the exclusion of a very large share of potential customers.

1.1 Main Results

Our results show that firms engage in more targeted screening (offering more tariff options) the more numerous are those consumers with high marginal valuation of the cellular service. In general, the value of the estimated structural parameters capture the features of this early industry, *i.e.*, one with non-negligible marginal costs and a small fraction of high valuation customers.

In our model, firms must soften their markups in order to induce some additional consumers to participate in the market because of the existence of an outside option that is not correlated with customers' valuation of the service. Because 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, uniform pricing achieves 94%, and a simple two-part tariff always exceeds 98% of maximum potential profits. Evidently, when the distribution of consumers is concentrated around low valuation customers, there is little gain from nonlinear pricing and most of it comes from excluding them and targeting the few 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. As more numerous options are offered, lower valuation customers subscribe to the cellular service but end up making very limited use of it, thus reducing the overall median airtime usage. Perhaps one of the most interesting findings 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 tariff from a welfare perspective but the fully nonlinear one. Finally, a universal service policy can certainly facilitate subscription, but if it is funded through cross-subsidization, the implied redistribution effects across customers reduce welfare of such a policy by at least 15% relative to the efficient solution.

The paper reports many other results such as the very low marketing costs associated to the commercialization of additional tariffs, that the distribution of consumer types is more spread in markets in which customers commute substantially and with low levels of income, that young commuters are more likely to find other alternatives to cellular service less appealing, and even less as time goes by, and that

the size of the network and cost to finance its deployment affect significantly the marginal cost of cellular carriers although weak, market specific, learning effects are present in this early industry.

1.2 Methodological Considerations

Since consumers may differ in several ways, a monopolist may, in principle, screen them with respect to as many dimensions as he can use to increase profits. In practice, different type dimensions are identifiable by their induced price-independent shifts of individual demands. 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 use in the empirical work.³ In theory almost any tariff behavior is possible under numerous regularity conditions. Thus, multidimensional screening models offer very little guidance to help identify 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) —hereafter *RS*— 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). 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 it attractive enough to participate in the market and consume, say $q(3) = 17$, then a consumer of type $t = 5$, with a higher intrinsic 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.⁴

³ See Rochet and Choné (1998) for the most general version available of this multidimensional screening model.

⁴ The slightly more general model of Maskin and Riley (1984) —hereafter *MR*— 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 between models of quality and quantity discrimination. The latter interpretation is more appropriate for our application.

In our *RS* based 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 , that summarizes all additional sources of heterogeneity of consumers. This scalar x enters additively into the utility function of consumers and is, by assumption, independently distributed from t . Thus, participation is random because it depends on the alternatives available to each consumer. The existence of horizontal heterogeneity (the x dimension) reduces the ability of the monopolist to extract consumers' informational rents. Since participation is random, it is optimal for the monopolist 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). This key reference outcome is replicated under the *RS* framework when the distribution of x becomes degenerate but ignoring the possibility of horizontal heterogeneity may lead one to overestimate the quantity underprovision needed to induce full separation of different consumer types.

Assuming that the participation decision is random comes at a significant cost. The variational problem that characterizes the optimal nonlinear tariff is no longer recursive and instead becomes a two-point boundary problem in which 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. 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.

1.3 Paper Outline

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 model of nonlinear pricing with endogenous participation decisions. Section 4 describes a behavioral model linking the *core* parameters of the computed *RS* model to market specific demand and cost variables. *Core* parameters are first recovered by making use of the equilibrium restrictions implied by firms' specific implementation of the nonlinear tariff solution by means of a menu of self-selecting two-part tariffs. We make use of simulated annealing to compute, market by market, the parameters of the *RS* model that best replicate the observed tariffs as implemented by means of a menu of options. Section 5 evaluates the welfare of alternative pricing mechanisms relative to the unconstrained fully nonlinear tariff. 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 the 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. 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 of thirty local markets, from more to less populated, 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 promised investment commitments proved to be far more costly than initially expected. After awarding licenses for the 30 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, and entry of the second carrier always occurred soon after a firm court decision was made.

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 stage of the industry cellular service was mostly targeting business customers: the handset was initially priced at \$3,000 and peak pricing spanned over a daily 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. Table 1 describes the features of the non-dominated tariff options offered by the firms of our sample. In general, higher fixed monthly fees F_i go together with lower rates per minute p_i . 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. Tariff data are complemented with market

Table 1: Tariff Features

Option No.	Monthly Fee, F_i		Rate per Minute, p_i	
	Mean	Std.Dev.	Mean	Std.Dev.
<i>Markets with ONE option</i> (56 observations)				
1	28.78	(11.50)	0.39	(0.07)
<i>Markets with TWO options</i> (23 observations)				
1	14.13	(4.96)	0.55	(0.13)
2	40.68	(9.40)	0.38	(0.10)
<i>Markets with THREE options</i> (16 observations)				
1	3.24	(1.29)	0.63	(0.14)
2	15.65	(8.62)	0.46	(0.07)
3	33.41	(18.29)	0.31	(0.10)

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

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.⁵

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, nor a precise definition of what constitutes the potential market in this early cellular telephone industry (the latter being a common problem in applied industrial economics). We therefore proceed by defining $COVERAGE = 1.3 \times TCELLS / (BUSINESS + 250 * POPULATION)$, where TCELLS denotes the number of antennae deployed in each market, BUSINESS is the number of business (in thousands) that were considered likely users of cellular telephony, and POPULATION is the population of each market (in millions). Thus, 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 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.⁶ This definition of market penetration is certainly arbitrary, but we feel that it is not unreasonable. We could have considered only the number of businesses 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

⁵ Many other variables were available but we only report here those that are used in the empirical analysis of Section 4.2.

⁶ 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%.

Table 2: Descriptive Statistics

Variables	First Quarter		Last Quarter	
	Mean	Std.Dev.	Mean	Std.Dev.
PLANS	1.4783	0.6909	1.6304	0.7989
COVERAGE	0.0679	0.0681	0.0671	0.0679
TIME	4.1957	2.4094	11.5652	4.0423
MKT-AGE	2.1087	3.1849	24.6087	10.5756
INCOME	28.2225	3.7465	27.5302	3.3053
COMMUTING	26.1609	2.8513	26.2174	2.7778
<i>sdv</i> (COMMUTING)	16.8640	2.9783	16.8627	2.9391
RAIN	3.4107	1.9115	3.0609	2.0246
POVERTY	11.0500	3.0281	11.2304	2.9443
POP-AGE	32.7543	2.8836	32.6826	2.8606
EDUCATION	12.5500	0.2420	12.5370	0.2507
TCELLS	18.1739	18.5320	18.3478	18.4237
HHSIZE	2.6223	0.2877	2.6195	0.2878
<i>sdv</i> (HHSIZE)	1.4665	0.1602	1.4654	0.1604
DENSITY	15.8190	14.0903	15.1110	13.1091
OPERATE	6.5055	1.5614	6.4900	1.4977
PRIME	10.8424	0.6310	9.1809	0.9526
WAGE	6.9580	1.5533	7.3320	1.6649
BELL	0.8261	0.3832	0.8261	0.3832
REGULATED	0.4783	0.5050	0.4565	0.5036
Observations	48		47	

All variables defined in Appendix A.

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 in 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 is a reasonable definition of the potential market, helps identifying the parameters of our model and highlights the methodological contribution of this paper and the value of its potential applications.

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. We believe that a static model of nonlinear pricing without network externalities is an accurate representation of the early U.S. cellular telephone industry. Pricing does not change much in the monopolistic markets of our sample even though the number of subscribers, while still few, increased significantly. Furthermore, the numbering of cellular phones was identical to fixed telephones and thus, cellular carriers were never able to offer within network discounts.

Our model specification makes use of particular functional forms to allow us to implement the *RS* model empirically. Since all our functional forms 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 tariff $P(q)$ to maximize profits:

$$\pi(q) = P(q) - cq. \quad (1)$$

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):

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

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 with $c \leq \bar{t}$.⁷

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

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

From an interior solution to maximizing (2), the demand of consumer type t is linear in the marginal tariff $p(q)$:

$$q(t) = \underset{q}{\operatorname{argmax}} v(t, q, x) = \frac{t - p(q(t))}{\gamma}, \quad (4)$$

for $q > 0$ and where $P'(q(t)) = p(q(t)) > -\gamma, \forall t \in \mathbb{T}$ is a sufficient condition for $q(t)$ being a maximum. It follows that the first best (*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.

⁷ 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.

The Burr type XII distribution 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. Different values of λ identify whether high valuation consumers are more or less numerous than low valuation consumers. If $\lambda = 1$ then t is uniformly distributed (*i.e.*, there is 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 $t = \underline{t}$, asymmetric information turns out to be irrelevant and homogeneous consumers can be efficiently priced by means of a single two-part tariff.⁸

Similarly, the exponential distribution (3b) 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. A small value of ϕ leads to a fast rate of decay of the probability density function of x , and it is thus associated with situations in which consumer horizontal heterogeneity is also small. Thus, the probability associated with 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 ours from a single-dimensional screening model. If $\phi \rightarrow 0$ 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 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 with 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 assume 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 stronger this negative correlation is, the less able the monopolist is to extract informational rents from

⁸ Miravete (2002, Appendix) and Miravete (2005) analyze at length the relationship between λ and the markup charged to different single-dimensional consumers. For further results and properties of the Burr type XII distribution see Johnson, Kotz, and Balakrishnan (1994, §12.4.5).

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:

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

where $f(t) = F'(t)$ and $g(t) = G'(t)$ are the probability density functions associated to (3a) and (3b), respectively.

3.2 Fully Nonlinear Tariff

The monopolist maximizes the expected profits designing the optimal direct revelation mechanism $\{P(t), q(t)\}_{t \in \mathbb{T}}$. This 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 .⁹ A piecewise-smooth function $q(t)$ is implementable by a tariff function $P(t) \equiv P(q(t))$ if and only if $q(t)$ is nondecreasing, $u(t)$ is absolutely continuous, and $\dot{u}(t) = q(t)$ at all continuity points of $q(t)$.¹⁰ After substituting our specific functional form assumptions, this problem can be stated as:¹¹

$$\max_{q(t), u(t)} \int_{\mathbb{T}} M(u(t), t) [P(q(t)) - cq(t)] dt, \quad (6a)$$

$$\dot{u}(t) = q(t) \geq 0, \quad (6b)$$

$$\dot{q}(t) \geq 0, \quad (6c)$$

$$u(t) = tq(t) - \frac{\gamma}{2}q^2(t) - P(q(t)) \geq 0. \quad (6d)$$

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 making it impossible to have it integrated by parts. Rochet and

⁹ Notice that $u(t)$ determines the consumer's decision to participate through the unconditional probability $M(u(t), t)$.

¹⁰ This property is also shared by the *MR* model. See Rochet and Stole (2002, Lemma 2), and Wilson (1993, §6.2).

¹¹ For convenience we denote the first and second derivatives 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 .

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 it is in our case, full market coverage by a monopolist is never optimal regardless of whether the support of x is bounded or not. This feature of the model —actually common to Armstrong (1996) and Roberts (1979)— makes it particularly well suited 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.
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}° is ensured to be a compact set. Pooling of different consumer types will never occur in any other region of the type space. 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. In our empirical analysis we normalize $\underline{t} = 0$ to ensure that the support \mathbb{T} is sufficiently spread and low valuation customers are excluded, a feature consistent with the low market penetration of the cellular telephone industry in the early age period of our sample.¹²
3. The monopolist's supply is such that $q(t) \in (q^{\text{mr}}(t), q^{\text{fb}}(t)]$ for all $t \in \mathbb{T}$. Furthermore, $\lim_{\phi \rightarrow 0^+} q(t) = q^{\text{mr}}(t)$, $\forall t \in \mathbb{T}$, so that our estimates can continuously approximate to the *MR* solution, which provides a lower bound consumption for multidimensional consumers.

Thus, from an econometric perspective, our adoption of a random participation framework to build our empirical equilibrium model appears to reduce the likelihood of misspecification relative to alternative models of single dimensional or multidimensional screening. First, as indicated above, the solution falls in between the solution of the single dimensional model of Maskin and Riley (1984) and the first best allocation in which the monopolist is able to extract all the informational rents from each consumer type (first degree price discrimination). 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 this multidimensional *RS* model does not rule out the possibility of concluding that the data are more consistent with either a single dimensional *MR* model or with the *FB* solution that arises in the absence of asymmetry of information. Second, nonlinear tariffs frequently include a monthly allowance

¹²However, 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, so that consumers are efficiently priced both at the *top* as usual, but also at the *bottom*. In short, if the monopolist finds it 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.

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.”

3.3 Screening with a Multipart Tariff

Firms rarely offer a fully nonlinear tariff to their customers, and certainly this never happens in our sample. We thus have to compute the optimal n -part tariff that best screens consumers for each set of structural parameters $(\lambda, \phi, \bar{t}, \gamma, c)$ 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 with 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 fixed commercialization cost associated to each tariff option ensures that the monopolist finds it optimal to offer a menu with only a finite number of tariff options.¹³ 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. The next section incorporates the discrete choice of the number of tariff options within our framework and suggests how to estimate this model.

More formally, let the local monopolist in market i at time τ observe $\omega_{i\tau} = (\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau})$. The monopolist also knows the cost of commercialization per tariff option, $\zeta_{i\tau}$. Since offering tariff options is costly, in addition to solving (6a)–(6d) 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 that ensures *IC*.

A particular two-part tariff is defined by the pair $\{A_k, b_k\}$, where A_k represents the fixed monthly fee and b_k the marginal tariff per minute of the k^{th} -option. A $(\nu + 1)$ -part tariff consists of a menu with ν options defined by a piecewise linear tariff function $\tilde{P}(\mathbf{A}_\nu, \mathbf{b}_\nu)$ where $\mathbf{A}_\nu = (A_1, \dots, A_\nu)$ and $\mathbf{b}_\nu = (b_1, \dots, b_\nu)$. Thus in deciding how many options to offer, the monopolist maximizes:

$$n_{i\tau} \in \underset{v_{i\tau} \in \mathbb{N}}{\operatorname{argmax}} \hat{\pi}(v_{i\tau} | \omega_{i\tau}) - v \cdot \zeta_{i\tau}, \quad (7)$$

¹³ Wilson (1993, §8.3) shows that as long as *IC* is fulfilled and the distribution $F(t)$ is increasing hazard rate, as it is in our case as long as $\lambda > 0$, then the incremental profits of adding a tariff option are positive but rapidly 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 commercialization costs, therefore effectively limiting the optimal number of tariff options offered.

where $\hat{\pi}(v | \omega_{i\tau})$ denotes the expected profits of a monopolist that offers the v self-selecting two-part tariffs that best implements the optimal nonlinear tariff characterized by (6a)–(6d).¹⁴

$$\hat{\pi}(v | \omega_{i\tau}) = \max_{\substack{0 < A_1 < \dots < A_v \\ b_1 > \dots > b_v > c}} \tilde{\pi}(\tilde{P}(\mathbf{A}_v, \mathbf{b}_v) | \omega_{i\tau}), \quad (8)$$

If a firm offers a N self-selecting tariff options in a particular environment characterized by $\omega_{i\tau}$, the incremental profits from offering N instead of $N - 1$ options should exceed the marketing and commercialization cost $\zeta_{i\tau}$. At the same time the incremental profit of offering $N + 1$ rather than N options should not justify introducing the $(N + 1)^{th}$ tariff plan. Hence, for each market i and time period τ the number of tariff options offered fulfills the following condition:

$$N_{i\tau} : \hat{\pi}(N_{i\tau} | \omega_{i\tau}) - \hat{\pi}(N_{i\tau} - 1 | \omega_{i\tau}) > \zeta_{i\tau} > \hat{\pi}(N_{i\tau} + 1 | \omega_{i\tau}) - \hat{\pi}(N_{i\tau} | \omega_{i\tau}), \quad (9)$$

and where $\hat{\pi}(0 | \omega_{i\tau})$ denotes the maximum of the expected profits between uniform pricing and a flat fee. Thus, we can think of the choice of a discrete number of tariff options, $N_{i\tau}$, as determined by the realization of the latent variable $\zeta_{i\tau}$ exceeding different thresholds given by the magnitude of foregone profits of not offering an additional tariff option.¹⁵ We then obtain, as a byproduct, the interval estimates where $\zeta_{i\tau}$ falls as the incremental profits of offering N or $N + 1$ options.

3.4 Qualifications

Before proceeding any further, we should point out a few qualifications to our analysis. First, we are implicitly assuming that the monopolist solves a static problem every period. This equilibrium approach ignores 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 should be below the optimal static markup and thus we would underestimate λ .¹⁶ As the optimal markup is less than the statically optimal one, it induces more participation and thus ϕ would most likely be overestimated.

¹⁴ Notice that $\tilde{\pi}(P | \omega_{i\tau})$ denotes the expected profits of offering the optimal fully nonlinear tariff as written in (6a) while $\tilde{\pi}(\tilde{P}(\mathbf{A}_v, \mathbf{b}_v) | \omega_{i\tau})$ are only the expected profits of offering $\tilde{P}(\mathbf{A}_v, \mathbf{b}_v)$, i.e., a *block declining tariffs* with v blocks. Formally $\tilde{P}(\mathbf{A}_v, \mathbf{b}_v) = \min\{A_1 + b_1q, \dots, A_v + b_vq\}$, $\forall q$.

¹⁵ 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 (2006) studied how the number of tariff options offered changes with the number of firms present in each market. Miravete (2007) evaluates the foregone profits of not offering an additional tariff option within a *MR* framework.

¹⁶ As the hazard rate of $F(t)$ is exactly $h_F(t) = 1/[\lambda(\bar{i} - t)]$, 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 .

Second, 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 constant marginal costs. Thus, for instance, the latter assumption of constant returns to scale in equation (1) ignores the possibility that capacity constraints 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 overestimate marginal costs and λ , as the actual markup in the presence of capacity constraints would exceed that of static pricing with constant returns to scale, thus also underestimating ϕ . We will address the effects of functional form assumptions by recomputing all pricing equilibria for alternative specifications of demand and the distribution of the vertical type dimension, $F(t)$.

4 Empirical Analysis

The empirical implementation of our model comprises three stages. In the first one we estimate market-by-market the *core* parameters $\omega_{i\tau} = (\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau})$, conditional on the actual number of tariff options offered in each market i and time period τ . Next, we compute the incremental profits of offering either $N_{i\tau}$ or $N_{i\tau} + 1$ tariff plans, thus providing with an interval estimate of the commercialization costs that rationalizes that firms only offer a few tariff options. Lastly, we explore how these *core* parameters that fully characterize each market within our theoretical framework can be empirically explained by observable demographics and other exogenous market characteristics. The idea is to explain the observed cross-market variation of our estimated *core* parameters, most of which characterize sources of asymmetric information or marginal costs, *i.e.*, economic variables that are not directly observable. These second stage regressions could serve to extend our policy evaluation to other markets not included in the sample. In Section 5 we analyze the welfare implications of our estimates, which allow us to rank the relative performance (profits, welfare, coverage, *et cetera*) of different nonlinear pricing mechanisms.

4.1 First Stage: Structural Parameters

For each market and time we first estimate the five *core* parameters of the model: $\omega_{i\tau} = (\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau})$. For each candidate estimate of $\omega_{i\tau}$ we have to compute the prediction of the monthly fee and marginal rate offered in each market as those that maximize profits within the family of $(\nu + 1)$ -part tariff options (*i.e.*, a menu of ν self-selecting two-part tariffs). Let $\{\hat{A}_\nu(\omega), \hat{b}_\nu(\omega)\} \in \mathbb{R}_+^{2\nu}$ solve the constrained maximization problem (8) for a candidate value of the *core* parameters ω in a market in which ν tariff options are offered

to consumers. The estimate $\hat{\omega}$ should minimize the distance between the features of the actual tariff options and those predicted by our model when restricted to using the piecewise linear function $\tilde{P}(\mathbf{A}_v, \mathbf{b}_v)$:

$$L_1(\omega) = \|\hat{\mathbf{A}}_v(\omega) - \mathbf{A}_v\|, \quad (10a)$$

$$L_2(\omega) = \|\hat{\mathbf{b}}_v(\omega) - \mathbf{b}_v\|. \quad (10b)$$

The model also predicts the share of consumers that participate in the market. Thus, we require that the model not only predicts the tariff properly, but also the market penetration:

$$L_3(\omega) = \left\| \int_{\mathbb{T}} M(t|\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega), t) dt - \text{COVERAGE} \right\|, \quad (11)$$

where $M(t|\cdot) = M(u(t), t|\cdot)$ is the model prediction of unconditional market participation as given by equation (5) evaluated at the optimal consumption $q(t|\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega))$ under the self-selecting menu of tariffs $\{\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega)\}$, and where COVERAGE is our available measure of market penetration for each SMSA and time.

The next two conditions close the model by setting the maximum consumption, q_{\max} and the average monthly bill. Attending to the consumption pattern of this early market we set the maximum usage for each predicted menu of options, $\{\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega)\}$ to $q_{\max} = 500$. Thus:

$$L_4(\omega) = \|q(\bar{t}|\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega)) - 500\|. \quad (12)$$

Finally, we require that the average bill equals \$100 a month.¹⁷ This number partially overcomes our lack of information about the distribution of individual consumption, thus generating meaningful economic predictions. We therefore require that:

$$L_5(\omega) = \left\| \int_{\mathbb{T}} \tilde{P}(q(t|\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega))) M(t|\hat{\mathbf{A}}_v(\omega), \hat{\mathbf{b}}_v(\omega), t) dt - 100 \right\|, \quad (13)$$

We therefore rationalize the position of the tariff options offered by each monopolist by means of the nonlinear pricing model with random participation of Section 3.1. We condition on the actual number of two-part tariffs offered in each market and time to generate such a mapping. Each optional tariff adds

¹⁷ 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.

two more conditions as those in (10a)–(10b), thus exceeding the number of *core* parameters to compute. The estimate $\hat{\omega}$ minimizes the weighted sum of the norms defined by (10a)–(13) independently for each market by means of a simulated annealing algorithm.¹⁸

$$\hat{\omega} \in \underset{\omega \in \mathbb{R}_+^5}{\operatorname{argmin}} \sum_{j=1}^{2v+3} \ell_j L_j(\omega). \quad (14)$$

Thus, conditional on some unobserved market-time fixed commercialization cost, $\zeta_{i\tau}$, that determines the optimal number of tariff options to offer, the vector of structural parameters $(\lambda_{i\tau}, \phi_{i\tau}, \bar{t}_{i\tau}, \gamma_{i\tau}, c_{i\tau}, \zeta_{i\tau})$ fully characterizes market i at time τ . The median values of these estimates are reported in Table 3.¹⁹ The computed structural parameters convey an intuitive characterization of this early industry.²⁰

The few consumers that subscribe to a cellular telephony have a maximum willingness to pay per minute that exceeds substantially the rates commonly charged at that time (see Table 1). These charges are also significantly higher than the estimated marginal costs, which highlights the significant market power of the *wireline* carrier. Few can be considered as “high valuation” consumers (since λ is always less than one) and most non-participants value using other communication methods (high value of ϕ) substantially more than cellular service. As the number of tariff options increase to ease subscription, those still not subscribing value other means of communication in the neighborhood of a thousand dollars per month over cellular telephony. As more lower valuation subscribe the mean telephone usage among active consumers falls and thus demand becomes more inelastic (increasing value of γ with the number of tariff options offered). Finally, commercialization costs are decreasing with the number tariff options offered. These marketing costs are negligible for all but those markets where only one tariff option is offered, where the average commercialization cost amounts to between 8,520 and 37,200 dollars per month in a market with 400,000 inhabitants.

4.2 Second Stage: The Effect of Demographics

Once we have recovered the structural parameters for each market, we can take advantage of the available cross-market variation to relate them to observable demographics. The point of this exercise is to learn how certain unobservable economic variables that are rarely available to econometricians such as the

¹⁸ Weights ℓ_j are only intended to account for the different scales of norms (10a)–(10b).

¹⁹ Indeed to ensure that we are computing the global solution, we solve each market ten times starting from different initial values and assign the median of these then solutions as the value of $\hat{\omega}_{i\tau}$ for market i at time τ . Despite costly computation, we never encountered multiple solutions beyond a very small margin of error.

²⁰ Computing $\omega_{i\tau}$ market by market is reasonable unless consumer decisions were correlated across different cities. Arbitrage across markets was not feasible because of the elevated roaming charges to make or receive calls from other cities. Thus consumers could not benefit from subscribing to the carrier from a neighboring markets and subscribed only to the cellular service offered by the local monopolist. Thus, the pricing decisions of each monopolist can safely be assumed to be independent across markets and carriers.

Table 3: Structural Parameters

	<i>First Quarter</i>	<i>Last Quarter</i>	<i>Both Quarters</i>
<i>ONE Option</i>			
λ	0.4073	0.3924	0.3962
ϕ	407.1654	277.2725	364.8999
\bar{t}	2.4570	2.3708	2.4139
γ	0.0044	0.0042	0.0043
c	0.1326	0.0875	0.1022
ζ_U	0.0930	0.0892	0.0930
ζ_L	0.0213	0.0227	0.0213
Observations	29	27	56
<i>TWO Options</i>			
λ	0.7360	0.5845	0.6019
ϕ	812.2232	683.2963	736.8599
\bar{t}	2.1034	2.1621	2.1392
γ	0.0090	0.0080	0.0081
c	0.2952	0.2228	0.2724
ζ_U	0.0098	0.0129	0.0102
ζ_L	0.0004	0.0002	0.0003
Observations	12	11	23
<i>THREE Options</i>			
λ	0.5466	0.9730	0.7598
ϕ	847.2153	952.0864	899.7136
\bar{t}	1.9891	1.8835	1.9178
γ	0.0098	0.0104	0.0101
c	0.2082	0.1761	0.2077
ζ_U	0.0007	0.0007	0.0007
ζ_L	0.0004	0.0000	0.0002
Observations	7	9	16

Median of the empirical distribution of computed structural parameters. Medians of ζ_U and ζ_L are measured in dollars per inhabitant of each market.

heterogeneous valuation of products, value of the outside option, marginal costs, or commercialization costs are related to available market characteristics.²¹

We assume that there is some measurement error associated to the estimated values of the *core* parameters, *i.e.*, $\hat{\omega} = \omega + \varepsilon$. These errors might arise were demand not linear or if the industry operated under increasing returns to scale. More interestingly in our framework, these errors may also arise 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. Thus, each 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 spe-

²¹ This two-stage approach is known to be valid since the work of Chamberlain (1982) and the second stage estimates are consistent as long as the first stage estimates are consistent and regressors are exogenous. This second stage simply maps the estimated *core* parameters onto a space of exogenous socioeconomic variables through a minimum distance estimator, thus simplifying the estimation by avoiding numerical integration of probabilities and the direct estimation of the covariance matrix of structural error terms.

cific characteristics observable to the econometrician, $\mathbf{Z}_{i\tau}$, as well as others that remain unobservable and summarized by $\varepsilon_{i\tau}$, plus a vector of coefficients δ_m to be estimated for the equation of each parameter:

$$y_{mi\tau} = Y_m(\mathbf{Z}_{mi\tau}, \delta_m, \varepsilon_{mi\tau} | y_{-mi\tau}), \quad (15)$$

where $m = 1, \dots, 6$, 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 $Y_m(\cdot)$ specify different proposed relationships between each structural parameter and observable market characteristics. In our econometric specification we assume that all $Y_m(\cdot)$ are linear functions of market characteristics and some other structural parameters, $y_{-mi\tau}$. In particular, 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 a different number of equilibrium conditions (10a)–(10b) to estimate these parameters depending on the number of tariff options offered. Similarly, all *core* parameters condition the number of tariff options to be offered through their effect on the profitability of each tariff option. These two relations define the exclusion restrictions of our second stage system estimation:

$$y_m = \theta_{m6}\zeta + \mathbf{Z}_m\delta_m + \varepsilon_m, \quad m = 1, \dots, 5, \quad (16a)$$

$$-\zeta = \sum_{j=1}^5 \theta_{6j}y_j + \mathbf{Z}_6\delta_6 + \varepsilon_6. \quad (16b)$$

Furthermore, since regressors of the system of equations (16a)–(16b) are exogenous and thus the consistent estimates of Table 4 can be given a causal interpretation.²²

Observe that the sequential profit maximization process described in (7)–(9) 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 $\theta_{61}, \dots, \theta_{65}$. 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 $\theta_{16}, \dots, \theta_{56}$. Vector δ_m measures the effect of observable firm and market characteristics, $\mathbf{Z}_{i\tau}$,²³ and $\varepsilon_{i\tau}$ captures the effect of any relevant but unavailable information and/or any misspecification error.

²² 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 that includes a discrete choice variable to account for the fact that ζ is not observable. The estimation of such a system of equations dates back to Amemiya (1978).

²³ $\mathbf{Z}_{i\tau}$ 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 $\mathbf{Z}_{mi\tau}$ in equation (15) indicates only those K_m exogenous variables that directly affect the behavior of each structural parameter.

Table 4: Second Stage Estimates

	λ	ϕ	\bar{t}	γ	c	PLANS($-\zeta$)
CONSTANT	0.0612 (0.26)	-1594.0500 (2.57)	1.6729 (1.56)	0.0235 (3.29)	-0.9998 (7.91)	-15.0935 (5.14)
TIME	0.3981 (5.55)	1370.3161 (6.09)	0.7893 (5.53)	0.0119 (5.27)	0.1769 (6.66)	-4.6981 (5.31)
MKT-AGE	-0.0034 (1.70)	-53.2701 (15.65)	0.0092 (2.09)	0.0001 (4.18)	-0.0066 (6.80)	
INCOME	-0.0158 (12.38)					
COMMUTING	0.0195 (6.47)	60.5690 (11.46)				
<i>sdv</i> (COMMUTING)	0.0527 (10.06)					
RAIN		52.1590 (6.09)				
POVERTY		36.7105 (4.14)				
POP-AGE		23.7085 (4.07)				
EDUCATION			0.1264 (1.85)			
TCELLS			-0.0088 (6.01)		0.0046 (15.32)	
HHSIZE				0.0008 (1.36)		
<i>sdv</i> (HHSIZE)				-0.0031 (2.35)		
DENSITY				0.0000 (3.46)		0.0078 (3.26)
OPERATE					0.0010 (0.34)	-0.2558 (6.30)
PRIME					0.1062 (12.91)	
BELL						1.1480 (13.86)
WAGE						-0.2946 (13.82)
REGULATED						0.3430 (5.07)
λ						0.2343 (0.51)
ϕ						0.0025 (22.51)
\bar{t}						3.2582 (15.02)
γ						-317.8301 (5.70)
c						18.3280 (19.66)
PLANS($-\zeta$)	0.1049 (7.93)	72.4160 (4.07)	-0.0542 (2.46)	0.0018 (11.63)	0.0774 (23.05)	
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).

Since economic theory provides little guidance on what demographics should be behind parameters indexing the distribution of asymmetric information, we attempted several specifications with some intuitive economic appeal and chose the one presented in Table 4.²⁴ We postulate that λ , ϕ , \bar{t} , and γ are all regressed against demand related variables (not to the same ones as to ensure proper identification of the model). Cost variables enter the regressions of c and the number of tariff plans (directly related to the size of $-\zeta$). All equations include a general and market specific time trends to capture the effect of passage of time that are common to the industry, TIME, or alternatively that are specific to the experience of each particular market, MKT-AGE.

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, with the average COMMUTING time of each market and with the number

²⁴ We do not claim that this is "the specification" that should explain similar empirical implementations of an RS-based model of screening. We report these results because it shows how to use 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.

of tariff options offered. The negative effect of INCOME indicates that affluent markets are overall more homogeneous.

Parameter ϕ increases with TIME and decreases (though far less) with MKT-AGE. Thus, as time goes it is more difficult to find consumers with a reservation value high enough to justify not subscribing to the cellular telephone service. The result is consistent with the availability of more tariff PLANS, some of them tailored for low usage customers with higher rates per minute but lower monthly fees. The lower the monthly fee is the higher the valuation of the outside option needs to be to justify not participating in the market. Similarly, the value of the outside option that justifies not signing up for the cellular service increases with the population age (younger customers are more prone to subscribe) as well as with the share of population under the poverty line (counterintuitive). Bad weather, measured as rainfall precipitation, also appears to ease the subscription decision.²⁵

As the industry matures, the maximum willingness to pay for cellular service, \bar{t} , also increases, thus reinforcing the previous argument regarding the time trend of ϕ . Valuation of the service increases with INCOME and EDUCATION, though marginally but surprisingly decreases with the size of the network as measured by TCELLS.

Parameter γ is inversely related to the average airtime usage that therefore increases over time. As more numerous consumers with lower valuations subscribe to the cellular telephone service, the average telephone usage decreases. This effect is more pronounced in markets located in high density urban areas.

The equation of marginal costs c includes cost related variables such as OPERATE and PRIME, as well as the number of antennae deployed, TCELLS, to control for potential effects of the scale on the unit costs of production. The marginal cost of production appears to increase with the size of the network as well as with the charges needed at this time to finance the building of the network. It appears that firms enjoy some sort of learning-by-doing since costs decrease with experience as measured by MKT-AGE. However, other effects common to the industry, and not only to a particular market dominate and overall c increases with TIME.

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

²⁵ 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 it less enjoyable to look for a pay-phone while away from home or the office, thus triggering a greater demand for cellular telephones.

review by initially introducing several tariff options.²⁶ Consistent with this view, there were more tariff options available in REGULATED markets. Similarly, more tariff options are offered when the customer base is more heterogeneous, the larger the expected value of the outside option (profiting from exclusion), the higher the valuation of cellular telephone service, the expected average usage, as well as with higher marginal costs.

5 The Welfare of Alternative Constrained Pricing

The first theoretical comparison among different pricing mechanisms that we are aware of is the work of Spence (1977), who proves that relative to uniform pricing, high valuation consumers purchase more and low valuation consumers purchase less with a nonlinear tariff. Assuming that all consumers are *always* served, Spence also shows that the possibility of introducing discounts increases the monopolist's revenues. 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 in neither of these two works.²⁷ Finally, Katz (1983) presents conditions under which an increase in output associated to a change in a pricing strategy is a sufficient indicator for welfare improvement.

In comparing the performance of tariffs, an important benchmark is the hypothetical case of full information, *i.e.*, when the monopolist can observe the types of each individual consumer. In such a case a personalized nonlinear pricing solution will be efficient. If the minimum consumer valuation exceeds the marginal cost of production, monopoly profits will 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 solution. The only difference between the two is the distribution of welfare among consumers and the monopolist but in both cases the solution is efficient. 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.

²⁶ 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 a few years eventually disappeared.

²⁷ 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.

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, the closer this nonlinear tariff is to the efficient pricing solution. However, it might also be optimal to exclude low valuation consumers altogether, in which case 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. In this section we study these issues systematically.

5.1 Constrained Tariffs

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 (linear) pricing, flat tariff, the optimal two-part tariff, and a Coasian two-part tariff where the price per additional unit of usage equals the marginal cost of production. We report their performance as measured by their profitability, welfare, coverage, consumption, and average 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. It should be noted that all 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.

Both monthly fees and usage rates are reasonable for this time period and not very different in magnitude from those reported in Table 1. This is most clear in markets with just one option. In markets with more options, monthly fees are lower in order to benefit from the participation of consumers of lower valuation and those with more valuable outside options. The difference between the Coasian and the optimal two-part tariffs is substantial as the latter is allowed to charge a positive markup over marginal costs. Both tariffs have, however, much smaller marginal rates than the linear tariff. The differences in markups become less important as we focus on markets with more numerous tariff options.

The airtime usage per subscriber for the optimal two-part tariff in markets who originally offered two tariff options is about 155 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

Table 5: Welfare, Usage, and Market Penetration

	<i>First Best</i>	<i>Nonlinear</i>	<i>Flat</i>	<i>Linear</i>	<i>Two-Part</i>	<i>Coasian</i>
<i>ONE Option</i>						
MONTHLY FEE	0.0000		150.6434	0.0000	29.1809	108.7219
RATE PER MINUTE	0.1022		0.0000	0.4907	0.3784	0.1022
PROFITS	0.0000	1.0000	0.8579	0.9687	0.9915	0.9147
PROFITS (*)	0.0000	2.1609	1.9360	2.1047	2.1434	1.9680
WELFARE	1.0000	0.7383	0.6869	0.7376	0.7365	0.7171
WELFARE (*)	5.2581	3.8793	3.7176	3.8737	3.8716	3.7650
COVERAGE	0.1356	0.0558	0.0444	0.0633	0.0542	0.0458
AIRTIME USAGE	269.2043	266.2942	388.4190	226.8420	275.9493	356.1682
UNDERSUPPLY	1.0000	0.4169	0.4580	0.4038	0.4134	0.4595
<i>TWO Options</i>						
MONTHLY FEE	0.0000		119.1983	0.0000	24.5869	59.4836
RATE PER MINUTE	0.2724		0.0000	0.6357	0.4851	0.2724
PROFITS	0.0000	1.0000	0.8185	0.9464	0.9881	0.9461
PROFITS (*)	0.0000	0.5816	0.5427	0.5545	0.6021	0.5752
WELFARE	1.0000	0.7406	0.6432	0.7378	0.7375	0.7252
WELFARE (*)	1.3717	0.9817	0.9876	1.0118	1.0101	0.9919
COVERAGE	0.0620	0.0264	0.0220	0.0314	0.0264	0.0239
AIRTIME USAGE	151.4041	152.0511	212.2645	125.4655	154.9054	188.4619
UNDERSUPPLY	1.0000	0.4161	0.4589	0.4034	0.4125	0.4587
<i>THREE Options</i>						
MONTHLY FEE	0.0000		92.9783	0.0000	19.4348	39.8904
RATE PER MINUTE	0.2077		0.0000	0.5403	0.3950	0.2077
PROFITS	0.0000	1.0000	0.8324	0.9368	0.9887	0.9552
PROFITS (*)	0.0000	0.2587	0.2302	0.2350	0.2554	0.2515
WELFARE	1.0000	0.7400	0.6539	0.7374	0.7375	0.7277
WELFARE (*)	0.5684	0.4231	0.3801	0.4194	0.4206	0.4178
COVERAGE	0.0398	0.0178	0.0152	0.0207	0.0173	0.0166
AIRTIME USAGE	121.0355	121.3356	178.0736	96.2239	123.4832	139.3662
UNDERSUPPLY	1.0000	0.4167	0.4588	0.4035	0.4132	0.4592

Median of the distributions across markets. All variables are defined in the text. PROFITS(*) and WELFARE(*) indicate the money value of these variables in millions of dollars for a market with mean potential customer base of about 400,000 customers.

case of a flat tariff. A distribution of types heavily concentrated around low valuation explains that the significant reduction in airtime usage when we compare markets where more tariff options were offered.

Because of the efficiency at the top result, only the highest consumer type pays the marginal cost for the latest unit consumed. All others consume less than under the *FB* solution due to the existence of asymmetric information. Since we do not observe individual subscribers we 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 total market usage 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, usage provision

under discriminatory pricing only amounts to 40% of the *FB* solution due to the existence of asymmetric information. Interestingly, both the optimal flat and Coasian tariffs reduce this distortion substantially as the increase in usage at very low or zero marginal rate compensates the exclusion of many low valuation customers.

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, well above 90% of the expected profits under nonlinear pricing with the exception of the flat tariff. This is reasonable in markets characterized by a small proportion of high valuation customers and within a framework in which 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 an environment there are limited gains from nonlinear pricing; a simple two-part tariff already ensures 98% of all potential profits.

One of the most interesting results of the paper is to document that from a welfare perspective nonlinear pricing adds very little over uniform *linear* pricing and that it dominates any other simple form of nonlinear pricing. The concentration of the distribution $F(t)$ around low values of t and the sharp reduction in consumption as soon as monopolists incorporate low valuation customers may well explain this result. If intermediate types were more numerous, their participation would go hand in hand with a larger airtime usage and the positive effect on consumer surplus of a reduction in marginal rates would tip the balance of welfare in favor of nonlinear tariffs.

6 Policy Evaluation: Implementing the Universal Service Requirement

For half a century, universal service in fixed telephony 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, businesses were also required to pay more for local connections than residents. Thus, the universal service requirement went together with cross-subsidization among the different lines of services that ended with the divestiture of AT&T in 1984.²⁸

Although cellular telephony has not been subject to universal service policies, other competitive services such as broad-band have also been targeted as essential services worth subsidizing, to ensure internet access to most of the population and avoid the feared *digital divide*.²⁹ In light of this, the purpose of this section is to evaluate the economic performance of a universal service policy in the context of the

²⁸ Muller (1997) reviews the origin, history, and development of the universal service in fixed telephony and Riordan (2002, §4.1) summarizes the discussion on cross-subsidization generated by this policy in the U.S.

²⁹ Crandall and Waverman (2000, §8) document that 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.

early cellular telephony industry.³⁰ Having determined the *core* parameters of the model assuming an unregulated profit-maximizing behavior on the part of the monopolist, we are in a position to conduct normative experiments and evaluate the optimal tariff under a constrained regulatory framework. The performance of each policy is measured not only from the point of view of individual consumption, but also taking into account its effect on the participation by different members of the heterogeneous customer base.³¹

We analyze two alternative ways to implement and measure the effects of the universal service requirement. In the *Break Even* approach, the monopolist is asked to maximize market penetration while self-financing the objective of the policy. Alternatively, in the *Free* approach, the monopolist provides the service at no cost to all subscribed customers, while the government simply covers all operational costs through a uniform lump-sum tax.

Under the first universal service requirement, we compute a pricing solution that maximizes market coverage while the monopolist makes non-negative profits. Using the computed structural parameters for each tariff, we solve the following problem for each market and time:

$$\max_{P(q)} \int_{\mathbb{T}} M(u(t), t) dt, \quad (17a)$$

$$\text{s.t.} \quad \int_{\mathbb{T}} M(u(t), t) [(P(q(t)) - cq(t))] dt \geq 0, \quad (17b)$$

$$M(u(t), t) = \left[1 - \exp\left(-\frac{u(t)}{\phi}\right) \right] \frac{1}{\lambda(\bar{t} - \underline{t})} \left(\frac{\bar{t} - t}{\bar{t} - \underline{t}} \right)^{\frac{1}{\lambda} - 1}, \quad (17c)$$

$$u(t) = tq(t) - \frac{\gamma}{2}q^2(t) - P(q(t)) \geq 0, \quad (17d)$$

$$q(t) \in \operatorname{argmax}_q \left\{ tq - \frac{\gamma}{2}q^2 - P(q) \right\}, \quad (17e)$$

$$q_1 \leq q_2 \Rightarrow P(q_1) \leq P(q_2), \quad \forall q_1, q_2. \quad (17f)$$

For convenience, in solving this problem we follow the *taxation principle* of Rochet (1985) rather than the *direct revelation mechanism* approach of Section 3.1. Now, the monopolist does not have any communication with consumers. He chooses a non-decreasing nonlinear schedule $P(q)$ and consumers choose their desired

³⁰ The absence of network externalities is perhaps the most important aspect of this early industry that conditions our evaluation. If network externalities were present, inducing participation will enhance welfare even further. But evidently, the policymaker should use the structural parameters of a model different from ours that has been estimated under the realistic assumption that such network externalities were not present.

³¹ For instance, there are studies documenting that income-targeted programs are not very effective in promoting subscription to fixed telephony. This is the case of Crandall and Waverman (2000, §6), Eriksson, Kaserman, and Mayo (1998), and Garbacz and Thompson (1997). However, all these studies incorrectly 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 adopted by the regulator.

usage level directly. This procedure allows us to maximize participation while ignoring the regularity features of the equilibrium tariff discussed in Section 3.2.³²

From a computational point of view, we restrict $P(q)$ to the class of functions that can be represented by a cubic spline interpolation over a given finite support $\mathbb{Q} = [0, q_{\max}]$, with an equally spaced interpolating grid of size Δ , $Q = \{q_i \in \mathbb{Q}, i = 0, \dots, N_q, q_0 = 0, q_{N_q} = q_{\max}, q_i - q_0 = i \Delta, \Delta > 0\}$:

$$P(q) = S_i(q), \quad \text{if } q_i \leq q \leq q_{i+1}, \quad \forall i = 0, \dots, N_q - 1, \quad (18)$$

with:

$$S_i(q) = s_{i0} + s_{i1}(q - q_i) + s_{i2}(q - q_i)^2 + s_{i3}(q - q_i)^3 \quad (19a)$$

$$S_i(q_i) \leq S_i(q_{i+1}), \quad \forall i = 0, \dots, N_q - 1, \quad (19b)$$

$$S_i(q_{i+1}) = S_{i+1}(q_{i+1}), \quad \forall i = 0, \dots, N_q - 2, \quad (19c)$$

$$S'_i(q_{i+1}) = S'_{i+1}(q_{i+1}), \quad \forall i = 0, \dots, N_q - 2, \quad (19d)$$

$$S''_i(q_{i+1}) = S''_{i+1}(q_{i+1}), \quad \forall i = 0, \dots, N_q - 2, \quad (19e)$$

where condition (19b) indicates that the interpolating function is nondecreasing when considering adjacent nodes, and conditions (19c)-(19e) restrict the optimal solution to be a twice continuously differentiable function at all interior nodes of Q . Under this specification, we search over the possible ranges and densities of Q —parameterized by the pair $\{q_{\max}, \Delta\}$ —, for the tariff $P(q)$ that gives the maximum participation rate, subject to profits being nonnegative.

The *Break Even* strategy best describes the idea of cross-subsidization that was induced by the rate structure sponsored by the FCC for fixed telephony. Once we know the distribution of vertical and horizontal heterogeneity of consumers in each market, we can predict by how much participation would increase if the monopolist is forced to break-even after solving problem (17a)–(17f). Alternatively, if the government can afford a direct subsidy to an industry, the maximum market penetration possible will be achieved by $P(q(t)) = 0, \forall t \in \mathbb{T}$, *i.e.*, what we call the *Free service* approach. Table 6 shows that these two alternative ways to implement the universal policy broadens market coverage substantially, about 25% above the market penetration of the *FB* solution and more than double the market coverage under a fully nonlinear tariff (compare with coverage in Table 5). However, Figure 1 shows that the source of the increase in market participation is very different depending on which alternative universal service strategies we consider.

³² Thus, it is possible that the coverage maximizing tariff is convex, induces bunching in mid ranges of the support of t or even fail to be efficient at the top as the goal is now not to separate different consumer types to extract rents from them but rather to ensure that many of them participate. This is the reason why in the above problem we do not include the *IC* condition (6c).

Table 6: Average Effects of Universal Service Policy

	<i>First Best</i>	<i>Break Even</i>	<i>Free</i>
<i>ONE Option</i>			
WELFARE	1.0000	0.8262	0.9670
COVERAGE	0.1356	0.1727	0.1616
AIRTIME USAGE	269.2043	148.3324	279.8256
UNDERSUPPLY	1.0000	0.5510	1.0395
<i>TWO Options</i>			
WELFARE	1.0000	0.8425	0.6317
MARKET PENETRATION	0.0620	0.0764	0.0723
AIRTIME USAGE	151.4041	89.5111	172.9245
UNDERSUPPLY	1.0000	0.5912	1.1421
<i>THREE Options</i>			
WELFARE	1.0000	0.8709	0.7902
MARKET PENETRATION	0.0398	0.0468	0.0577
AIRTIME USAGE	121.0355	79.5081	135.5475
UNDERSUPPLY	1.0000	0.6569	1.1199

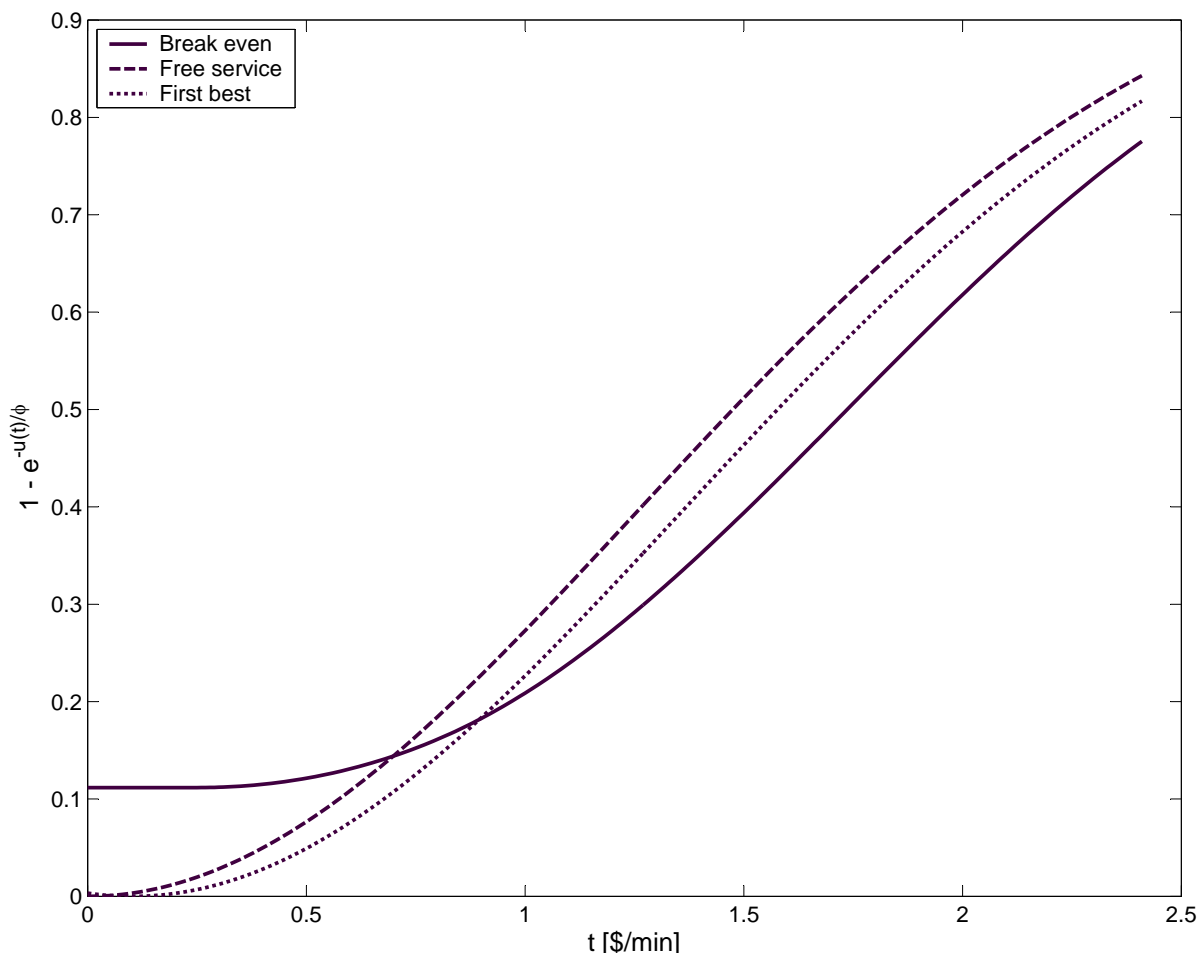
Results reported for the median of the parameters values. Variables are defined as in Table 5.

In order to induce participation a policy of universal service requires important markup reductions. In the *Free* universal service policy reduces the value of the outside option uniformly to all consumer types t , and thus, a larger proportion of them, whether high or low, will participate in the market. This is seen in Figure 1 as an approximately parallel upward shift of the participation line relative to the *FB* solution. Consumers in this market are simply receiving a rent transfer from the government and thus all consumers types may potentially benefit from it as even those willing to pay a positive amount less than c end up subscribing to the service. Since under this *Free* service marginal rates are also lower for all consumer types we can conclude then that the oversupply result reported in Table 6 is widespread and common to all types t , as it is captured by the upward shift of consumption relative to the *FB* solution in Figure 2.³³

The case of a *Break Even* universal service policy is quite different and perhaps more interesting because it involves a substantial redistribution of rents among participants in the market. In this case the goal of maximizing market penetration needs to be achieved without funding from outside the industry. Thus, the monopolist reduces the tariff discount for large customers and increases it for small ones and such re-balancing of the tariff needs to be done optimally so that the number of new low valuation subscribers offsets those high valuation one that decide to leave the market. In environments characterized by our relatively large fraction of low valuation customers, the overall effect of this policy is to increase participation at the cost of reducing consumption of all but the highest consumer type relative to the *FB*

³³ Figures 1 and 2 refer only to those markets where one single option was made available to consumers (more than half of our sample). Qualitative results are identical for the other markets in which more options were offered.

Figure 1: $F(t)$ — Who participates under universal service?



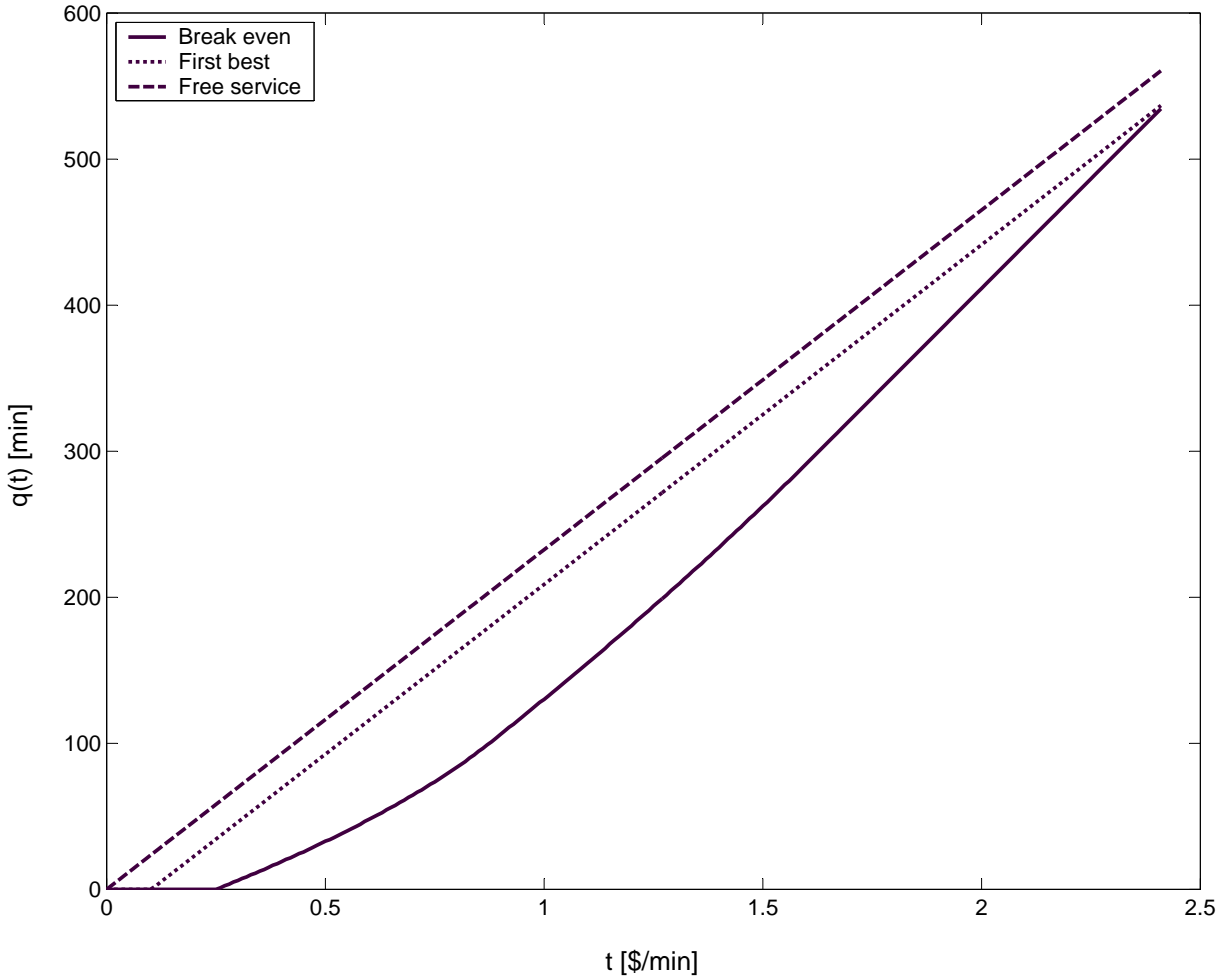
solution. Notice that Figure 2 shows how some of these new low valuation customers do not even consume while all infra-marginal types $t < \bar{t}$ consume less than in the *FB* solution.³⁴

The success of universal service policies is normally measured only by the actual increase in participation but little attention is paid to the efficiency loss and redistributive effects of such policy. Our model takes into account participation and consumption decision and thus can be used to assess the impact of these policies as the combined effect of participation and consumption decisions determine the overall welfare effect.

Under the *Free service* all consumer types are more likely to participate than in the *FB* solution, and conditional on participating all of them increase their consumption. Welfare increases an average of 15.39% relative to the nonlinear tariff solution. Alternatively, we could measure the performance of this policy

³⁴ The efficiency at the top of the *Break Even* policy depicted in Figure 2 is not the result of a constraint imposed while solving problem (17a)–(17f) but rather a feature of the solution.

Figure 2: $F(t)$ — Consumption profile under universal service?



by the trade-off between participation and efficiency. Thus, in order to increase the market penetration by 1% we have to give up 11.89% of the welfare of the *FB* solution. Evidently, all computations of the *Free* strategy ignore the shadow cost of public funds needed to subsidize this industry. Similarly, under the *Break Even* service low types are far more likely to participate than in the *FB* solution despite many of them not consuming at all, and always less than in the efficient solution. Welfare increases relative to the fully nonlinear tariff only by 13.25% and increasing participation by 1% requires to incur in a welfare loss of approximately 8.35% relative to the *FB* solution.

To conclude, our analysis shows that self-financed universal service policy can be even more effective than *Free* service in promoting participation when low types comprise the majority of the customer base. Contrary to the opinion of Crandall et al. (2004), our analysis shows that financing subsidies to low valuation customers through taxes raised from high valuation ones can actually be an effective way to foster participation in telecommunications related industries if there the market is characterized by a large fraction

of low valuation customers. This policy has the effect of adding many low valuation consumers with very modest consumption profiles and excluding some high valuation, intensive consumers. Evidently, such redistribution is responsible for the important welfare loss that a *Break Even* policy represent relative to the *Free* service case. However, even from this perspective, the increase in subscription of low valuation customers at the expense of some high valuation ones and a generalized reduction of consumption that, overall, enhances welfare moderately (by 13.25%) beyond the unregulated fully nonlinear pricing case.

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 endogenous participation decisions induced by each tariff strategy. Thus, our results show, not surprisingly, that nonlinear pricing increases profits only slightly over *linear* pricing in early cellular markets with a majority of low valuation consumers. In environments like this, firms maximize profits by excluding low valuation customers rather than by screening among very diverse customers. In such circumstances it is generally more profitable to offer simple tariffs to an active minority of consumers than adding (costly) options to incorporate a minuscule fraction of low valuation customers and worrying about the design of the right incentives to properly enforce *IC* constrains.

The paper has documented that the early U.S. cellular telephone industry was characterized by firms facing non-negligible cost and heterogeneous consumers regarding both participation and cellular usage. This heterogeneity appears to be related to income, consumer age, and commuting time, among others. Most importantly, low valuation customers are more numerous than high valuation customers, another expected feature of this early market, much characterized by the exclusion of consumers through high fees and expensive equipment. Finally, we have shown that a self-financed universal policy can be a very effective strategy in environments in which consumption and participation are driven by different sources of asymmetry of information although the impact on welfare is rather limited as its effects mostly involve a redistribution of rents among different consumer types.

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 worsened 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 since in competition x , the value of the outside option comes determined mostly by the tariff offering of competitors.

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.
- CHAMBERLAIN, G. (1982): "Multivariate Regression Models for Panel Data." *Journal of Econometrics*, 18, 5–46.
- 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 (2007): "Monopoly Quality Degradation in Cable Television." *Journal of Law and Economics*, 50, 181–219.
- 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.
- 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.
- 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.
- IVALDI, M. AND D. MARTIMORT (1994): "Competition under Nonlinear Pricing." *Annales d'Economie et de Statistique*, 34, 71–196.
- 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.
- KATZ, M. L. (1983): "Non-Uniform Pricing, Output and Welfare under Monopoly." *The Review of Economic Studies*, 50, 37–56.
- MASKIN, E. AND J. RILEY (1984): "Monopoly with Incomplete Information." *RAND Journal of Economics*, 15, 171–196.
- MIRAVETE, E. J. (2002): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optimal Calling Plans." *The Review of Economic Studies*, 69, 943–971.
- MIRAVETE, E. J. (2005): "The Welfare Performance of Sequential Pricing Mechanisms." *International Economic Review*, 46, 1321–1360.
- MIRAVETE, E. J. (2007): "The Limited Gains From Complex Tariffs." Mimeo, University of Texas at Austin.
- 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).
- MIRRELES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies*, 38, 175–208.

- MULLER, M. L., JR (1997): *Universal Service*. Cambridge, MA: MIT-AEI.
- 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.
- REISS, P. C. AND M. W. WHITE (2005): "Household Electricity Demand, Revisited." *The Review of Economic Studies*, 72, 853–883.
- RIORDAN, M. (2002): "Universal Residential Telephone Services." In M. E. 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. (1985): "The Taxation Principle and Multi-Time Hamilton-Jacoby Equations." *Journal of Mathematical Economics*, 14, 113–128.
- 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." *The Review of Economic Studies*, 69, 277–311.
- SEIM, K. AND V. B. VIARD (2006): "The Effect of Market Structure on Cellular Technology Adoption and Pricing." Mimeo, Stanford Graduate School of Business.
- 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.
- 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 and MKT-AGE is the number of months each particular market has been in operation.
- 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,³⁵ the average commuting time in minutes, COMMUTING, 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*(·)” 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;³⁶ and an index of average annual wages per employee for the cellular industry, WAGE. This latter source of information also reports the total number of antennae deployed in each market, TCELLS.
- 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.³⁷ 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.

³⁵ BUSINESS refers to what was considered at that time to be highly potential customers by cellular industry experts: business service firms, health care, professional, and legal services, contract construction, transportation, finance, insurance, and real estate.

³⁶ 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.

³⁷ See Easterling, D.R., T.R. Karl, E.H. 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.