Two Part Tariffs with Partial Product Bundling*

Sissel Jensen†

Discussion paper 19/2001
Revised August 2004

Abstract

When a firm operates in an industry with very large differences in consumers' willingness to pay for the service it offers, it faces a challenge in the pricing decision. It wants to engage in price discrimination, but cannot identify a given consumer's market segment ex ante. When consumers' willingness to pay is private information, a widely used sorting mechanism is to offer a menu of two part tariffs, letting high demand and low demand consumers self-select into distinct market segments by their tariff choice. However, when the difference in consumers' willingness to pay is very large, simple two part tariffs are not longer sufficient to discriminate between high and low demand segments; Despite the ability to price discriminate, the firm still prefers to serve high demand consumers only. The model suggests that it might be possible to discriminate between consumers by other means than price-cost distortions; Low demand consumers face a two part tariff with a per unit price possibly above marginal cost, together with a restriction on usage, whereas high demand consumers face an efficient two part tariff.

JEL.CLASS.NO: D42, D82, L96
KEYWORDS: Price discrimination, two part tariffs, quantity discounts, telecommunications

---

*I am grateful for comments from Petter Osmundsen, Fred Schroyen, Lars Sørgard and Jon Vislie. I also thank seminar participants at the Norwegian School of Economics and Business Administration, participants at the 3rd Nordic Workshop in Industrial Organization (NORIO III) in Helsinki. Financial support from Telenor is gratefully acknowledged.

†Institute for Research in Economics and Business Administration. Mailing address: Breiviksveien 40, 5045 Bergen, Norway. Phone: 55 95 94 57, Internet: sissel.jensen@snf.no.

1 Introduction

Some firms operate in industries with very large differences in consumers' willingness to pay for the services they offer. This can be confirmed by observing consumers' actual consumption patterns. For instance, in telecommunications there is a big difference in consumption between low and high demand segments for most services, e.g., call minutes on fixed line connections and mobile phones, and download/upload volume on broadband connections. If a firm sets high price to extract surplus from high demand segments, it will certainly exclude many low demand consumers. If it sets a low price in order to serve low demand segments, it will forego profit from high demand consumers. Hence, it want to practice price discrimination. We know that a monopoly firm can capture more of consumers' surpluses by using a menu of self-selecting two part tariffs instead of uniform pricing, see for instance Goldman, Leland and Sibley (1984), Sharkey and Sibley (1993), and Wilson (1993). However, when a consumer's willingness to pay is private information, the trade-off between exclusion of low demand segments versus rent extraction from high demand segments still exists, and the relative performance of two part tariffs over uniform pricing decline as the demand side heterogeneity increases.

In this paper we explore an extension of a simple two part pricing arrangement by assuming that the firm can observe a customer's usage of its service along more than one dimension. By monitoring consumers' usage patterns, the firm is able to offer a tariff targeted to low demand consumers on terms that differ from the terms on which high demand consumers make their purchases, i.e., the firm uses more than one instrument in the pricing decision. We show that this might imply lower distortions towards low demand consumers in terms of the price-cost margin compared to the bench-mark with one observable dimension. To illustrate the general idea, let us take a closer look at a strategy used in broadband pricing. Users of broadband services have very diverse needs when it comes to internet surfing, e-mailing, music and video downloads, and high quality video and audio streaming, and this reflects their demand for speed and their intensity of usage with respect to download/upload. While surfing the internet is just faster on a high speed connection, high quality video streaming will perform bad on low speed connection. Hence, consumers' willingness to pay for access speed depend partly on which services he uses and partly on his usage intensity. When both dimensions (speed and download) is observable, the firm should consider restrictions along both as part of its screening decision.

Table 1 present an example. The broadband company Tiscali charge high demand consumers £24.99 for broadband at 512kbps, while low demand consumers pay £10 less for the same speed, but on different terms since they must stay online less than 50 hours, or download less than 1Gb, per month. As we

Table 1: *Pricing of broadband, Tiscali (UK, June 2004)*

Product	*Downstream speed*	*Cost per month ($)*
Broadbandx3	150kbps	15.99, Free usage
Broadbandx5	256kbps	19.99, Free usage
Broadbandx10 50 Hours	512kbps	19.99, After 50 hours 2p per minute
Broadbandx10 1Gb	512kbps	19.99, After 1Gb 2p per Mb
Broadbandx10 Unlimited	512kbps	24.99, Free usage

see, Tiscali sorts consumers by self-selection according to two instruments, speed and download. Another example that is consistent with the recommendations in this paper is the widespread practice of various kinds of calling circle tariffs, for instance "Friends and Family" tariffs. Under a calling circle tariff, a subscriber is billed according to aggregate minutes of calling to a restricted set of network subscribers, and the price per minute varies conditional on the node of termination (inside or outside the calling circle). Common in the examples above is that the firm introduces a restriction in the use of the service in low demand consumers' tariff, not really to restrict low types' consumption, but to hurt high demand consumers if they choose a tariff with a low price. The underlying assumption is that consumers with different valuation for the service has distinctly different usage pattern. Specifically, only low demand consumers are willing to accept a severe restriction in the dispersion of calls/monthly download, against a reduction in the cost of usage. Table 2 lists some other examples from telecommunications.

In the analysis, we assume that a monopoly firm sells a single generic good to two consumer groups with different valuation for the good.[1] Although it is clear that there are distinct market segments, it is not possible for the firm to identify a given consumer's market segment ex ante, i.e., consumers' willingness to pay is private information. In order to simplify the interpretations and the intuition, we choose to apply the model to a case with calling circle tariffs.[2] However, with some reinterpretations, the results are valid for most telecommunications and digital services (as indicated in table 2). Consumers usage of the service can be monitored along two dimensions, denoted call duration and call dispersion, both being observed by the monopoly and both being continuous. Call duration is interpreted as a quantity variable, whereas call dispersion is related to

[1]Although telecommunications is subject to competition almost all over the world, we do not add imperfect competition to the framework. The reason for doing this is simply that it adds too much complexity (see Rochet and Stole (2003) and Stole (2001)).

[2]Firms' use of calling circle tariffs has received some attention in other areas in the economics literature as well. Wang and Wen (1998) consider a duopoly model with demand side heterogeneity, where such pricing behavior enables a new firm to enter the market despite the presence of consumer switching costs. Laffont, Rey and Tirole (1998) examine the effects of discriminatory pricing on the negotiated interconnection agreements between rival network operators. In a recent publication written independently of this, Shi (2003) study the use of calling circle tariffs from a social network theory perspective.

Table 2: *Examples on pricing of telecommunications services (June 2004)*

Company/Product	*Service restriction*	*Pricing arrangement*
Vodafone (UK) Perfect Fit	Lower rates anytime Lower rates daytime Lower rates evening/weekend	Two part tariffs with inclusive minutes
Orange (UK) Your Plan	Lower rates any network anytime Lower rates Orange-Orange anytime Lower rates Orange-Orange off-peak	Two part tariffs with inclusive minutes
O_2 (UK)	Lower rates anytime Lower rates daytime/evening time	Two part tariffs with inclusive minutes
BT (UK) Together 1,2,3	Low rate evening/weekend Free calls evening/weekend Free calls anytime	Two part tariffs
Telenor (Norway) Friends & Family	Lower rates on calls to mobile Lower rates on national calls Lower rates on international calls	Two part tariffs
Tiscali (UK) Dial-up internet access	Unlimited surfing anytime Unlimited surfing daytime, weekdays Unlimited surfing daytime all week	Flat rate Per minute outside hours
BT Broadband 512kbps	Less than 15Gb monthly download Upgrade to free download	Flat rate
BT Broadband 1Mb	Less than 30Gb monthly download Upgrade to free download	Flat rate

the concentration of calls made within the network. In addition, we assume that consumers with different willingness to pay also have distinctly different calling patterns. In particular, high demand consumers make calls to a large number of subscribers, whereas low demand consumers make calls to a small number of subscribers. Hence, call dispersion is an intrinsic part of consumers' preferences, and assumed to be perfectly correlated with consumers' quantity preferences. High dispersion subscribers can be thought of as business consumers while low dispersion subscribers can be thought of as residential consumers. The firm introduces quantity distortions (high per unit price) towards low demand types, according to the well-known model with nonlinear pricing in Mussa and Rosen (1978) and Maskin and Riley (1984). In addition it introduces a restriction on the use of the service in low demand consumers tariff, i.e., it restricts his calling circle by restricting the fraction of the network he can reach.[3]

If we change the interpretation slightly, the model can be used to analyze nonlinear pricing and bundling in a multiproduct monopoly setting. Assuming

[3]The monopoly sells its product in two "versions", a "high-quality" version with unrestricted calling, and a "low-quality" version with a restriction in call dispersion ("damaging"). Deneckere and McAfee (1996) analyze damaging in a framework with third degree price discrimination, and show that the practice can lead to a Pareto improvement. Foros, Jensen and Sand (1999) study the effects of damaging under two part tariffs with exogenous qualities.

that the firm sells a very large number of products,[4] we let the firm bundle a subset of the products and charge units within this product bundle according to a distinct two part tariff. Since the firm does not debundle completely we refer to this practice as partial product bundling. In a model with unit demand, Bakos and Brynjolfsson (1999) study the strategy of bundling a large number of information goods (goods with zero or very low marginal costs of production) and selling them for a fixed price. One of their findings is that the firm should offer a menu of different bundles aimed at each market segment and practice price discrimination when consumers' tastes are positively correlated.[5] The seller can for instance offer an "economy" bundle that is a subset of a "premium" bundle, or "degrade" the quality of the product to disproportionally affect high-demand consumers' valuation (damaging). Armstrong (1999) study optimal multiproduct nonlinear pricing when the firm offers a very large number of products, applicable to telecommunications.[6] When consumers' tastes are correlated across products, he finds that a menu of two part tariffs, each of which have prices proportional to marginal costs, can extract almost all available profits. Note that the tariffs applied in our model follow these recommendations since we assume that the marginal cost do not vary over the product line. However, Armstrong (1999) cover only the case where all products are sold to both consumer types.

Section 2 present consumers' preferences with calling pattern heterogeneity. The focus in section 3 is on the design of two part tariffs, depending on the severity of the usage restriction, and subsequently on the trade offs the firm faces when it decides the relative role of the two instruments at hand, i.e., restriction of call dispersion and price-cost distortion. Section 4 study this in more detail by specifying consumers' distribution functions with respect to call dispersion. Section 5 offer some concluding remarks.

[4] For instance, call minutes to different network nodes are grouped together into the same sub-utility function, and the value a person places on one group of services is independent of whether he consumes the other.

[5] The literature on bundling is large, but since it to a large extent deal with a setting with only two products, and linear pricing, most is not relevant to our model.

[6] Multiproduct nonlinear pricing is also studied elsewhere. Mirman and Sibley (1980) consider a multiproduct monopoly facing consumers who are differentiated by a single characteristic, where the firm offer a menu of commodity bundles together with the price for the bundle. Sibley and Srinagesh (1997) explore the difference between screening the different dimensions of consumer types independently by means of two part tariffs and the alternative of bundling all taste parameters to design a single two part tariff. Miravete (2001) study multidimensional screening where different type components distinguish quality dimensions of products that can be aggregated.

2 A model with calling pattern heterogeneity

The market is served by a monopolist and resale opportunities are absent. The cost function is assumed to be linear, the fixed cost is excluded from the measure of profit and the marginal cost is normalized to zero. On the demand side there are only two consumers (or equally large groups), type 1 with low willingness to pay and type 2 with high willingness to pay. Hence, under a uniform price, type 2 would buy a larger quantity than type 1. A consumer's type is unobservable to the firm, but each type's preferred calling pattern, described by the dispersion of calls, is known. The reservation utility is assumed to be equal for the two consumers and normalized to zero.

The types' call dispersion is exogenous and we shall assume that type 2 has a more dispersed calling pattern than type 1, and that this is common knowledge. Call dispersion is described by consumers' intensity of network node visitation, i.e., how large fraction of a consumer's calls are made to his favorite number, how large fraction to the top-two numbers, top-three and so on. The network node gives the identity of the party called and is not in itself of any interest to the firm. However, the relative intensity of calls made to each network node gives information about consumers' preferences. Since we are only interested in calls made by these two consumers, we can without loss of generality normalize the "entire network" to 1, and say that type 2 always makes calls to the entire network whereas type 1 has a more concentrated calling pattern. Hence, call dispersion can be seen as an intrinsic part of consumers' preferences.[7]

[7] A similar problem is studied in Shi (2003). He describes the structure of a consumer's social network according to interpersonal tie strengths and the relational density of a personal network. Each consumer's demand for communication is determined by the number of strong versus weak ties, and by each consumer's valuation of communication with strong and weak ties. Consumers' willingness to pay for communication with a strong link normally exceeds the willingness to pay for communication with a weak link. However, Shi (2003) includes the possibility that the demand curves for a strong tie and a weak tie crosses. Using a numeric example, Shi (2003) reports two main results: If consumers with many loose ties are the high valuation segment, he finds that the low valuation segment is charged below cost for communication with strong ties and above cost for communication with strong ties. If consumers with many dense ties are the high valuation segment, he finds that the low valuation segment is charged below cost for communication with weak ties and above cost for communication with strong ties. The results are driven by the assumptions about the demand curves together with the assumption that the network sizes are equal.

Shi (2003) do not consider demand heterogeneity due to taste differences, or differences in income levels, between consumers. The only characteristic that differ between consumers is the number of strong versus weak ties in their personal communications network. The distribution of these numbers are known, but the firm can not observe the characteristics of individual personal communications networks. The setting is different in my paper: Demand heterogeneity is due to unobservable differences in consumers' tastes, and observable differences in calling pattern. An additional assumption made in Shi (2003) is that it is optimal for the firm to serve both consumer segments. Hence, he do not examine how increased observability

We define call dispersion according to a cumulative distribution $F_1(n) \geq F_2(n)$ with a probability density function $f_i(n)$, $i = 1, 2$.[8] If a consumer is restricted to make calls only to a n-fraction of the network, he will make calls to those network nodes generating the largest consumer surplus. $F_i(n)$ states how large fraction of consumer i's call minutes that are placed within the n-fraction of the network, and $f_i(n)$ states consumer i's intensity of calls to network node n. The density function $f_i(n)$ is positive and integrable on the support $n \in [0, 1]$, and we will also assume that $f_1 F_2 \leq f_2 F_1$.

We make the assumption that call dispersion is independent of the price per call minute. This assumption may be questioned. However, there is no obvious alternative assumption – i.e., that consumers' calling pattern will be more concentrated or more dispersed when the price per call minute increases. Hence, keeping call dispersion constant seem to be as good as any other assumption. On the other hand, as we will see, call duration to each network node is decreasing in the price per call minute.

A consumer of type θ_i derives a utility from making calls to a given network node according to the following subutility function

$$v_i(x, n) = \theta_i x - \frac{1}{2f_i(n)} x_i^2 \tag{1}$$

We will assume that each consumer is billed according to a single two part tariff $T_i = \{p_i, E_i\}$ (subscript i indicates that the tariff is intended for the consumer of type θ_i, $i = 1, 2$). If a consumer of type θ_i finds it individual rational to pay a fixed fee E, the price for each call minute is equal to p, and the total call length to each network node maximizes the quasilinear subutility function $u_i = v_i(x, n) - px$. Hence, expected call length to a given network node is

$$x_i(p, n) = (\theta_i - p) f_i(n) \tag{2}$$

A consumer's aggregate utility is obtained by aggregating the subutility over all possible network nodes. When each subutility function is quasilinear, the aggregate demand function will appear to maximize the aggregated consumer surplus, and a consumer's gross surplus is represented by the area under the the aggregate demand function. We abstract from the fact that some consumers may have positive utility even in the case when consumption is zero.[9] If the expected

of consumer heterogeneity affect the monopolist's incentive to exclude low demand segments.

[8]For notation we use $f_i(n) \equiv f(n; \theta_i)$, $F_i(n) \equiv F(n; \theta_i)$. The distribution of n conditional on θ_2 first-order stochastically dominates the distribution of n conditional on θ_1, if $\theta_2 \geq \theta_1$. This assumption captures that call dispersion is perfectly correlated with call duration.

[9]A subscriber may want a network connection in order to receive calls only, or to be able to make emergency calls. Oren, Smith and Wilson (1982) study nonlinear pricing under the presence of demand externalities, for instance when the benefit a consumer receives in a communication network depend on their access to communication partners and increase with the size of the network.

net utility from making calls weakly exceeds a consumer's reservation utility he will find it beneficial to subscribe to the network.

Aggregate demand for a consumer of type θ_i over all possible network nodes is given by

$$Q_i(p) \equiv Q(p, \theta_i) = \int_0^1 (\theta_i - p) f_i(n) \, dn = (\theta_i - p), \tag{3}$$

$i = 1, 2$. Since $\theta_2 > \theta_1$ it must be the case that $Q_2(p) > Q_1(p)$ $\forall p$, $Q_i(p)$ is nonincreasing in p, $i = 1, 2$.

Aggregate demand for a consumer of type θ_i over the n-fraction of the network most frequently called is given by

$$\bar{Q}_i(p, n) \equiv Q(p, \theta_i, n) = \int_0^n (\theta_i - p) f_i(n) \, dn = (\theta_i - p) F_i(n), \tag{4}$$

$\bar{Q}_i(p, n)$ is nonincreasing in p and nondecreasing in n, $i = 1, 2$. It is not necessary to impose the single crossing on the subutility functions in (1) since the firm is trying to sort consumers only on the basis of aggregate consumption. Hence the restricted demand curves in (4) can cross.

The demand curve in (4) resembles the one in (3), except that n affects the intercept and the slope of the individual demand curves. However, these are perfectly (negatively) correlated and the firm can infer about the slope when it knows the intercept (and vice-versa).[10]

Consumer surplus under a two part tariff $T = \{p, E\}$ for some given $n \leq 1$ is given by

$$\begin{aligned} CS_i(p, E, n) &= \int_p^{\theta_i} (\theta_i - p) F_i(n) \, dp - E, \quad i = 1, 2, && (5) \\ CS_2(p, E, n) &> CS_1(p, E, n). && (6) \end{aligned}$$

When both types choose consumption subject to the same tariff, type 2 obtains a larger surplus.

We now proceed in two steps. First, we solve for the optimal two part tariffs, T_1 intended for type 1 and T_2 intended for type 2, treating n as exogenous. Next, having obtained a reduced form profit as a function of n we solve for the optimal size of the allowed calling circle in the two part tariffs T_1 and T_2, assuming that they are excluded from making calls outside the calling circle. This is only a simplifying assumption, and we will also report the results when the assumption is relaxed.

[10] Laffont, Maskin and Rochet (1987) solve for the optimal nonlinear price schedule when a monopolist is uncertain about both the slopes and the intercepts of the individual demand curves it faces, assuming a continuum of types and that the distributions of slopes and intercepts are independent.

3 Two part tariffs

Given the slopes of the demand curves and asymmetric information over θ, the practice that maximizes profit is to offer different two part tariffs intended for the two consumer types. We know the equilibrium in this model as a solution where $p_1 > 0$ and $p_2 = c$. The fixed fee in type 1's tariff is chosen in such a way that he receives his reservation utility, and the fixed fee in type 2's tariff is chosen such that type 2 does not choose the tariff intended for type 1. More formally, consider the model as follows. A two part tariff is characterized by a triple $\{p, E; n\}$, p is the marginal price, E is a fixed fee and $n \leq 1$ is the fraction of the network that can be reached with the tariff. When the reservation utility is normalized to zero, it is individually rational to accept any tariff $\{p, E; n\}$ that yields nonnegative consumer surplus. The two individual rationality constraints are

$$CS_i\left(p_i, E_i, n_i\right) \geq 0, \quad i = 1, 2. \tag{IR_i}$$

Since $CS_2(.) > CS_1(.)$, IR_2 can not bind whenever IR_1 is weakly met. Hence if type 1 is served, IR_1 is the only binding individual rationality constraint. The other relevant constraints are the incentive compatibility constraints

$$CS_i\left(p_i, E_i, n_i\right) \geq CS_i\left(p_j, E_j, n_j\right), \quad i, j = 1, 2, \; i \neq j. \tag{IC_i}$$

The incentive constraint requires that a consumer buys the bundle intended for his type. IC_1 can never bind if IC_2 is weakly met. Hence, the incentive constraint is downward binding only.[11]

It is never profitable to restrict type 2's demand and any restriction in call dispersion will only occur in the tariff intended for type 1. Henceforth we use the notations $n_2 = 1$ and $n_1 = n$. The firm is searching for two part tariffs $\{p_1, E_1, n\}$ and $\{p_2, E_2, 1\} \equiv \{p_2, E_2\}$ in order to maximize profit. If n is fixed we have the following maximization problem

$$\Pi = \max_{p_1, p_2, E_1, E_2} \left\{ E_1 + p_1\left(\theta_1 - p_1\right) F_1\left(n\right) + E_2 + p_2\left(\theta_2 - p_2\right) \right\} \tag{7}$$

subject to $p_i \geq 0, E_i \geq 0$ $(i = 1, 2)$, IR_1, and IC_2

$$E_1 = \int_{p_1}^{\theta_1} \left(\theta_1 - p\right) F_1\left(n\right) dp, \tag{8}$$

$$E_2 = E_1 + \int_{p_2}^{\theta_2} \left(\theta_2 - p\right) dp - \int_{p_1}^{\theta_2} \left(\theta_2 - p\right) F_2\left(n\right) dp. \tag{9}$$

[11] See for instance Tirole (1988) pp 153-154, and Fudenberg and Tirole (1991), pp 247-248.

The outcome is unique with $p_1 \geq p_2 = 0$, and $E_2 > E_1$, whenever $\theta_2 > \theta_1$ and both types are served. The last term in (9) illustrates the two instruments that can be used to reduce the information rent. The firm can increase p_1 or decrease n. If the firm chooses not to serve type 1, the unique outcome is a cost-plus-fixed fee tariff, $p_2 = 0$, and the entire consumer surplus is extracted via the fixed fee.

We now turn to the question of how severe the restriction in call dispersion in type 1's tariff should be. As a benchmark however, we first repeat the profit maximizing two part tariffs in the single-dimensional case with $n = 1$. If the firm has no ability to monitor call dispersion, or to condition a tariff on a calling circle restriction, $n_1 = n_2 = 1$. This is the canonical model with two-types and single-dimensional screening which is examined in, for instance, Sharkey and Sibley (1993).

Lemma 1 (Single-dimensional screening) *A monopoly that is unable to observe anything but individual quantity purchases will increase the unit price in type 1's tariff above marginal cost in order to reduce the information rent to type 2. If consumer heterogeneity is sufficiently large, the monopoly will exclude type 1 from buying.*

(i) For $\frac{\theta_2}{\theta_1} \in \left[1, \frac{3}{2}\right]$ the monopoly will serve both types and offer two different two part tariffs $\{p_1, E_1\}$ and $\{0, E_2\}$ given by

$$p_1 = \theta_2 - \theta_1,\ E_1 = \tfrac{1}{2}\left(2\theta_1 - \theta_2\right)^2,\ E_2 = \tfrac{1}{2}\left(2\theta_1 - \theta_2\right)^2 + \tfrac{1}{2}\left(\theta_2^2 - \theta_1^2\right).$$

(ii) For $\frac{\theta_2}{\theta_1} > \frac{3}{2}$ the monopoly will exclude type 1 and offer a cost-plus-fixed-fee tariff $\{0, E_2\}$ and extract all surplus from type 2. The tariff is given by

$$E_2 = \tfrac{1}{2}\theta_2^2.$$

Lemma 1 is simple to verify by substituting for $F_1(n) = F_2(n) = 1$ in the above maximization problem (7) to (9). The information rent to type 2 is exactly balanced against the gain from serving type 1 when $\theta_2/\theta_1 = 3/2$, i.e., type 1 is served only if $\theta_2/\theta_1 \leq 3/2$ (cut-off rate).

Next, we turn to the case of a wider strategy set, i.e., where the tariff intended for type 1 may have a calling circle restriction. According to (3) and (4), a restriction in call dispersion causes a negative horizontal shift in the demand curves. Type 2's gross surplus from consuming the good is evaluated according to type 2's true willingness to pay, $Q_2(p)$, while he is given an information rent as if the heterogeneity was described according to the demand curves $\bar{Q}_1(p, n)$ and $\bar{Q}_2(p, n)$. A distortion in type 1's tariff makes it less tempting for the high demand type to mimic the low demand type. Although type 1 also suffers under such distortions, he is not as seriously affected as type 2.

Lemma 2 (Two-dimensional screening) *If consumers' calling patterns are type dependent, and can be monitored by the monopoly, a restriction on type 1's call dispersion serves as an alternative to a distortion in the unit price to type 1. Assume for now that both types are served and that $\frac{\theta_2}{\theta_1}$ is not too large. For a given restriction $n \leq 1$, type 2 is offered a cost-plus-fixed-fee tariff $\{0, E_2^n\}$ and type 1 is offered a tariff $\{p_1^n, E_1^n, n\}$ where*

$$p_1^n = \begin{cases} 0 & for \quad \frac{\theta_2}{\theta_1} \in \left[1, \frac{F_1(n)}{F_2(n)}\right], \\ \theta_2 - \theta_1 \frac{F_1(n)}{F_2(n)} & for \quad \frac{\theta_2}{\theta_1} \in \left(\frac{F_1(n)}{F_2(n)}, 1 + \frac{F_1(n)}{F_2(n)}\right], \end{cases}$$

E_1^n and E_2^n are determined by (8) and (9).

Lemma 2 is verified by solving maximization problem (7) to (9). Under our assumptions on F_1 and F_2, p_1^n is nondecreasing in n, continuous, and differentiable whenever $p_1^n > 0$. From the pricing rule in Lemma 2 we see that larger heterogeneity in call duration (θ_2 is large relative to θ_1) results in a larger unit price. Further, because type 2 consumers suffer more both from a restriction in call dispersion and from an increase in the unit price, they serve as alternative instruments to relax the incentive constraint. This is reflected in the result that p_1^n is decreased (increased) when n is decreased (increased). In both cases the means is to restrict type 2's consumption if he selects type 1's tariff, by way of a high unit price or access to a smaller network (reduced opportunity set).

Both instruments are costly to use in the sense that type 1's consumption is de facto restricted. In either case the consequence is that type 1 will make fewer calls. The firm loses income from these calls and since type 1 loses surplus on these calls he is not willing to participate unless the fixed fee is reduced. On the other hand, type 1's tariff is no longer as tempting for type 2 and the fixed fee from type 2 can be increased. The optimal trade-off in the firm's use of the two instruments depends on the relative effect they have on the two types' demand.

Assuming that both types are served we use Lemma 2 and write the expected profit in (7) as a function of n

$$\Pi(n) = \begin{cases} \frac{1}{2}\theta_2^2 + \frac{1}{2}\theta_1^2 \frac{F_1(n)^2}{F_2(n)} - F_1(n)\,\theta_1(\theta_2 - \theta_1) & \text{if} \quad p_1^n > 0, \\ \theta_1^2 F_1(n) + \frac{1}{2}\theta_2^2(1 - F_2(n)) & \text{if} \quad p_1^n = 0. \end{cases} \tag{10}$$

The firm maximizes profit with respect to n and the tariffs are determined by Lemma 2.

Lemma 3 (Restriction in call dispersion) *The firm separates between high and low demand consumers by distorting type 1's tariff with respect to call dispersion, alone or together with a distortion in the unit price.*

(i) Type 1 is offered a cost-plus-fixed-fee tariff with a restriction in call dispersion $\tilde{n} \in (0,1]$ if $\tilde{n}$ exists such that

$$\frac{F_1(\tilde{n})}{F_2(\tilde{n})} \geq \frac{\theta_2}{\theta_1} \geq \sqrt{\frac{2f_1(\tilde{n})}{f_2(\tilde{n})}}$$

(ii) Type 1 is offered a two part tariff with a unit price distortion and a restriction in call dispersion $\hat{n} \in (0,1]$ if $\hat{n}$ exists such that

$$\frac{\theta_2}{\theta_1} \geq 1 + \frac{F_1(\hat{n})}{F_2(\hat{n})}\left(1 - \frac{1}{2}\frac{f_2(\hat{n})}{f_1(\hat{n})}\frac{F_1(\hat{n})}{F_2(\hat{n})}\right) \geq \frac{F_1(\hat{n})}{F_2(\hat{n})}$$

The tariffs are subsequently determined according to Lemma 2.

The firm chooses to place a restriction in call dispersion in order to satisfy the condition $\partial\Pi/\partial n \leq 0$. The last inequality in part (i) of Lemma 3 states the condition for $p_1^n = 0$, whereas the first inequality in part (ii) of Lemma 3 states the condition for $p_1^n > 0$. In the first case, the firm only has to trade-off how an increase in n affects the fixed fees. Hence, if the heterogeneity in call duration is low relative to the heterogeneity in call dispersion, type 1 is more likely to be served with a flat rate tariff with a calling circle restriction, i.e., when θ_2/θ_1 is small and/or F_1/F_2 is large. In the opposite case, the firm will offer type 1 consumers a two part tariff with a calling circle restriction together with a distorted unit price.

The firm will restrict type 1's calling circle whenever there is heterogeneity in the types' calling pattern. Since the tariff intended for type 2 has no restriction in call dispersion, the demand curves $\bar{Q}_1(p,n)$ and $Q_2(p)$ never cross if $\theta_2/\theta_1 \geq F_1(n)$, which is always met. It does not matter whether the demand curve $\bar{Q}_2(p,n)$ crosses $\bar{Q}_1(p,n)$ since type 2 is not expected to make his purchases along $\bar{Q}_2(p,n)$. When call dispersion conditional on consumer type θ is known, we can characterize the firm's pricing policy.

4 A numerical example

For simplicity we assume that type 2 make calls of equal length to all nodes, i.e., $f_2(n)$ is uniformly distributed on the interval $[0,1]$, $f_2(n) = 1$ and $F_2(n) = n$. We assume that type 1 has a more concentrated calling pattern by placing more probability weight to the left tail of the distribution.

The Beta-distribution allows for the possibility that the call length may vary over nodes of call termination. Define type 1's dispersion of calls according to the following p.d.f and the c.d.f.[12]

$$f_1(n,w) = \begin{cases} w(1-n)^{w-1} & \text{if } 0 \le n \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$F_1(n,w) = \begin{cases} 1-(1-n)^{w} & \text{if } 0 \le n \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

The firm seeks to maximize profit with respect to n according to the optimality condition in Lemma 3. Henceforth, we define a variable $t \equiv \frac{\theta_2}{\theta_1}$. The monopoly's pricing strategy is given in the following propositions. Proposition 1, 2, and 3 are obtained by applying Lemma 3 and are proved in Appendix A. Figure 1 summarizes the results.

Proposition 1 *If heterogeneity in call dispersion is sufficiently large relative to the heterogeneity in call duration, type 1 is served with a cost-plus-fixed-fee tariff* $\{0, E_1^n, n^*\}$, $n^* \in [n', n'')$. *This occurs for* $t \le t' \le t''$, *or for* $2 \le w \le 1 + \frac{1}{2}w$. n' *and* n'' *decrease whereas* t' *and* t'' *increase as the heterogeneity in call dispersion increases (*w *increases).*

Proposition 1 shows that a calling circle restriction in type 1's tariff may be sufficient to separate the types. The larger the heterogeneity in call dispersion, the more powerful is a calling circle restriction as an instrument to separate the types. This can be utilized by the firm in two different ways. The firm can achieve less costly separation by decreasing n (reflecting that n'' decreases as w increases), or serve more types with a social efficient tariff (reflecting that t'' increases as w increases). Consumers with different willingness to pay are charged identical unit price (equal to marginal cost), but type 2 pays a larger fixed fee. In terms of pricing, this resembles first degree price discrimination, except that type 1 is not allowed to make calls to the entire network.

Proposition 2 *When the heterogeneity is more moderate and balanced, type 1 consumers are offered a two part tariff* $\{p_1^n, E_1^n, n^{**}\}$, $p_1^n > 0$, $n^{**} \in [0, n')$. *This occurs for* $t' \le t \le t''$ *together with* $w < 2$.

[12] This is the beta distribution over n with shape parameters $v = 1$ and $w > 1$ on $[0, 1]$. The probability density function for the beta distribution is

$$f(n,v,w) = \begin{cases} \dfrac{n^{v-1}(1-n)^{w-1}}{B(v,w)} & \text{if } 0 \le n \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where the shape parameters v and w are positive numbers. With $v = 1$, the shape of the distribution is determined by w, the higher is w the larger is the mass for low n.

Proposition 2 show that when the heterogeneity in call duration increases, it is necessary to increase the restriction in call dispersion (decrease the calling circle) in order to restore incentive compatibility. A calling circle restriction will always be used, either alone (Proposition 1) or in combination with distortionary pricing (Proposition 2).

Proposition 3 *If heterogeneity in call duration is sufficiently large relative to heterogeneity in call dispersion, type 1 is excluded from making purchases. This occurs for $t > \sqrt{2w}$ if $w < 2$ or for $t > 1 + \frac{1}{2}w$ if $w > 2$. Type 1 is served in more cases relative to the single-dimensional case.*

Although increased heterogeneity in call dispersion reduces the incentive to exclude type 1, proposition 3 states that this incentive still exists.[13] Typically, the possibility of type 1 being served increases as the heterogeneity in call dispersion increases because this increases the 'observability' of the two types. The generalization of this is the fact that the firm is always better, or at least equally well, off with an additional observable and instrument at hand.[14]

4.1 Pricing under a three part tariff

I the previous section, low demand consumers were excluded from making calls outside the calling circle. In this section, we allow low demand consumers to make calls outside the calling circle at a separate unit price $\bar{p}_1^n$. Calls to the calling circle is charged at a per unit price p_1^n. Hence, high demand consumers are offered a two part tariff $\{p_2^n, E_2^n\}$ and low demand consumers are offered a three part tariff $\{p_1^n, \bar{p}_1^n, E_1^n\}$.

In order to solve the extended problem, we have to add the utility from these calls to the individual rationality constraint for low demand types (IR_1), to the incentive constraint for the high demand types (IC_2), and revenues from sales to the profit function. The maximization problem can be found in Appendix B, and the problem is stated in B.1 – B.3. The insights are summarized in figure 2(a) – 2(c).

When the heterogeneity in call dispersion is low, it is more likely that low demand types are served with a three part tariff (figure 2(a)). However, at the same time, the two unit prices, p_1^n and $\bar{p}_1^n$ are close to each other when w is close

[13] Instead of saying that $n^* = 0$ we could say that $n^* = 1$ but let p_1^n be sufficiently high to ensure that $Q_1(p_1) = 0$.

[14] Sappington (1983) shows this in a regulation model. A regulator that is uncertain about a multiproduct firm's production technology achieves additional information by observing the production level of each product. Caillaud, Guesnerie, Rey and Tirole (1988) generalize the case with several observable variables.

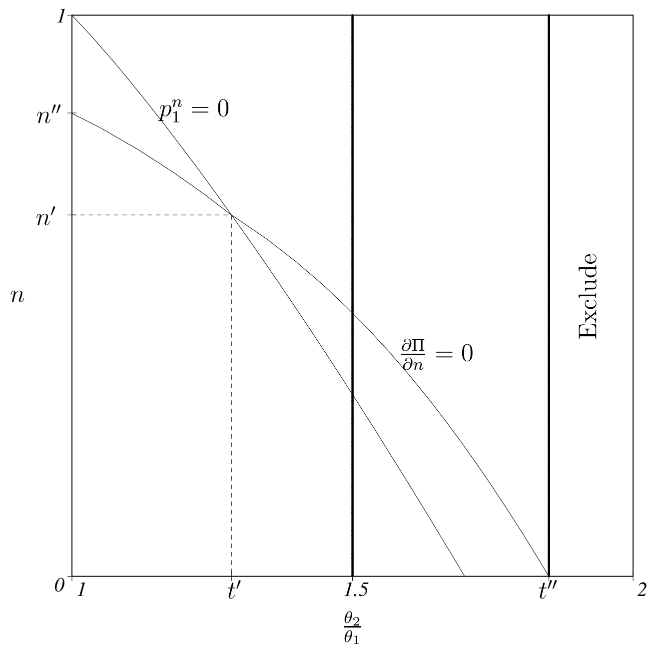
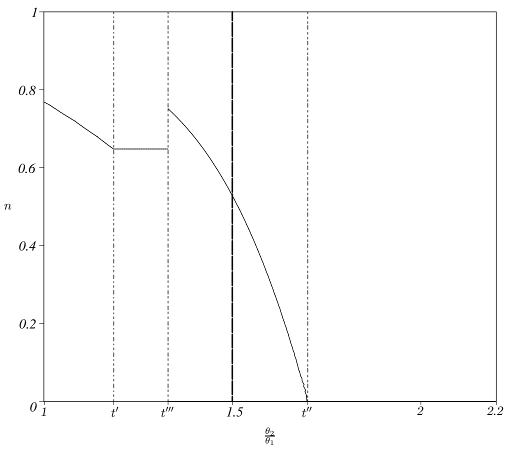


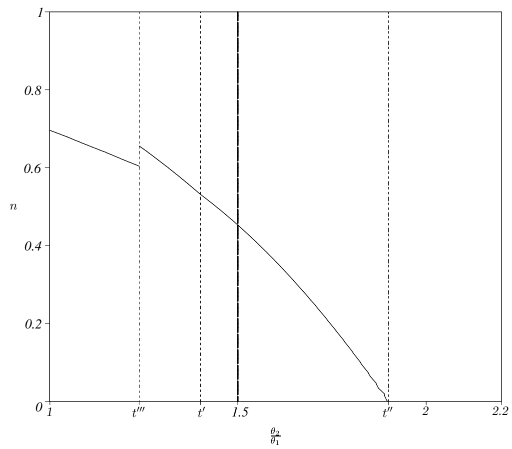
Figure 1: *Pricing policy towards type 1,* $w = 1.7$*. The larger the heterogeneity in call dispersion (high* w*) the larger is the possibility that type 1 is served and that he is served with an efficient tariff, i.e., a cost-plus-fixed-fee tariff.*

to 1. When the heterogeneity in call dispersion increases, it is more likely that low demand types are charged at marginal cost for call to the calling circle (in figure 2(b), t' is closer to t''). At the same time it becomes less likely that the firm will use a three part tariff (t''' closer to 1). When the heterogeneity in call dispersion is large, low demand types' calls to the calling circle is always charged at marginal cost (figure 2(c)). At the same time, it becomes less likely that low demand types are served with a three part tariff.

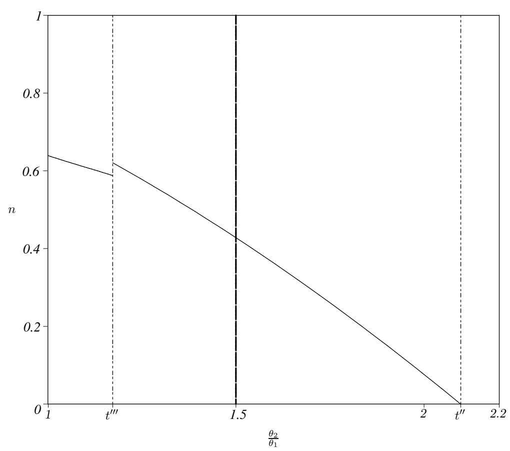
The example above shows that the firm moves from a simple two part tariff to a calling circle tariff as the heterogeneity in call dispersion increases. In contrast to Shi (2003), we find that both unit prices are above marginal cost. In addition, by allowing more flexibility in the firm's pricing strategy, we show that a three part tariff is optimal only when the heterogeneity in call dispersion (w) or call duration (θ_2/θ_1) is relative small.



(a) Low heterogeneity in call dispersion, three part tariff for $t \leq t'''$ $(1 \leq w \leq 1.6)$.



(b) Moderate heterogeneity in call dispersion, $p_1^n = 0$ for $t \leq t'$ $(1.6 \leq w \leq 2)$.



(c) Large heterogeneity in call dispersion, $p_1^n = 0$ whenever type 1 is served $(w \geq 2)$.

Figure 2: *Low demand types are served with a three part tariff for $\theta_2/\theta_1 \leq, t'''$. If $\theta_2/\theta_1 \leq t'$, calls to the calling circle is charged at marginal cost. For $\theta_2/\theta_1 \in [t''', t'']$, low demand types are served with a calling circle tariff only. Low demand types are excluded for $\theta_2/\theta_1 > t'''$.*

5 Concluding remarks

Price discrimination is socially desirable if it induces the firm to serve market segments it otherwise would have excluded. Under second degree price discrimination, a widely used sorting mechanism is to give quantity discounts to high demand segments; Low demand segments pay a low monthly fee and a higher per unit price, whereas high demand segments pay a high monthly fee and a per unit price equal to marginal cost, and high demand and low demand consumers self-select into distinct market segments by their tariff choice. However, when the difference in consumers' willingness to pay is very large, pure quantity discounts are not longer sufficient to discriminate between high and low demand segments; The firm still prefers to serve high demand consumers only. Under such circumstances it is desirable to use additional, or alternative sorting mechanisms. In this paper, this is done by using information about the two consumers' calling patterns in addition to monitoring their quantity purchases. Both dimensions are observable, and both vary systematically with consumers' true willingness to pay; High demand consumers buy many units and have a dispersed calling pattern; Low demand consumers buy fewer units and have a concentrated calling pattern. Therefore, they are both appropriate as sorting mechanisms.

In our model, the firm can introduce quantity distortions towards low demand types, according to the well-known model with nonlinear pricing. Another instrument is to introduce a restriction on the use of the service. The firm typically finds it optimal to combine distortions along the two dimensions. Then, type 1 consumers face a two part tariff with a marginal price possibly above marginal cost, together with a restriction on the usage pattern (the calling circle). However, the calling circle restriction allows the firm to reduce the distortion in the pricing rule in the low-demand type's tariff. Whenever the monopoly firm finds it profitable to serve type 1, and there is observable heterogeneity in the use of the service, it will always impose a calling circle restriction in type 1's contract. Moreover, this restriction is sometimes sufficient to achieve separation.

The model shows that the firm chooses to serve market segments it otherwise would have excluded with an additional sorting dimension. Hence, not only will profit increase, but sorting along this dimension may also lead to a Pareto improvement. Finally, the model suggests that it might be possible to practice a pricing strategy closer to flat rate pricing by separating consumers by other means than price-cost distortions. Hence, the outcome would be closer to first degree price discrimination.

Appendix

A Proof of Propositions 1, 2, and 3

The profit function given in (10) is continuous and concave. From Lemma 2 and Lemma 3 we derive the conditions $p_1^n = 0$ and $\Pi'_n = 0$, which are the two curves in figure 1. The slopes of these are given by

$$\left.\frac{dn}{dt}\right|_{p_1^n=0} = \frac{n^2}{nf_1 - F_1} \leq 0 \tag{A.1}$$

$$\left.\frac{dn}{dt}\right|_{\Pi'_n=0} = \begin{cases} \dfrac{\sqrt{2f_1}}{f_{1n}} \leq 0 & \text{if} \quad p_1^n = 0 \\[2ex] \dfrac{2n^3 f_1^2}{2f_1(nf_1 - F_1)^2 + nF_1^2 f_{1n}} \leq 0 & \text{if} \quad p_1^n > 0 \end{cases} \tag{A.2}$$

with notation $f_{1n} \equiv df_1(n,w)/dn$, $f_{1w} \equiv df_1(n,w)/dw$ and so on.

When w increases there will be a positive shift in the curve defining $p_1^n = 0$.

$$\left.\frac{dt}{dw}\right|_{p_1^n=0} = \frac{F_{1w}}{n} \geq 0 \tag{A.3}$$

The shift in the curve defining $\Pi'_n = 0$ is negative for larger values of n and positive for smaller values of n.

$$\left.\frac{dt}{dw}\right|_{\Pi'_n=0} = \begin{cases} -\dfrac{f_{1w}}{f_{1n}} & \text{if} \quad p_1^n = 0 \\[2ex] \dfrac{F_{1w}}{n} - \dfrac{1}{2n^2}\dfrac{F_1(2F_{1w}f_1 - f_{1w})}{f_1^2} & \text{if} \quad p_1^n > 0 \end{cases} \tag{A.4}$$

When w increases it places more probability weight to the lower end. Hence, f_{1w} is positive for smaller values of n and negative for higher values of n, while f_{1n} is negative for all $n \in [0,1]$.

Next, we evaluate the shift along the t-axis

$$lim_{n\to 0^+}\left[\left.\frac{dt}{dw}\right|_{\Pi'_n=0}\right] = \begin{cases} \frac{1}{\sqrt{2w}} & \text{if} \quad p_1^n = 0 \\[2ex] \frac{1}{2} & \text{if} \quad p_1^n > 0 \end{cases} \tag{A.5}$$

Hence, since the shift is positive along the t-axis, the shift along the n-axis must be negative, implying that t'' is increasing and n'' is decreasing in w.

We can show that n' decreases when the heterogeneity in call dispersion increases by differentiating the condition

$$\frac{F_1(n',w)}{n'} = \sqrt{2f_1(n',w)} \tag{A.6}$$

which gives us

$$\frac{dn'}{dw} = -\frac{n^2 f_{1w} - nF_{1w}\sqrt{2f_1}}{n^2 f_{1n} - \sqrt{2f_1}(nf_1 - F_1)} \leq 0 \tag{A.7}$$

Since we have $t' = \frac{F_1(n',w)}{n'}$, which is monotonic with $dt'/dn' < 0$ (by A.1), t' is increasing in w. By inspection we can conclude that the firm offers a cost-plus-fixed-fee tariff for $t < t'$ and $n > n'$. This completes the proof of Propositions 1 and 2.

When $w = 2$ the curves are tangent at the point $(t, n) = (2, 0)$ and $t' = t''$.

$$lim_{n\to 0^+}\left[\frac{F_1}{n}\right] = w \tag{A.8}$$

$$lim_{n\to 0^+}\left[\sqrt{2f_1}\right] = \sqrt{2w} \tag{A.9}$$

$$lim_{n\to 0^+}\left[1 + \frac{F_1}{n}(1 - \frac{1}{2}\frac{F_1}{f_1})\right] = 1 + \frac{1}{2w} \tag{A.10}$$

The shift in the curve defining $p_1^n = 0$ along the t-axis is given by

$$lim_{n\to 0^+}\left[\left.\frac{dt}{dw}\right|_{p_1^n=0}\right] = 1 \tag{A.11}$$

The shift in (A.11) is larger than (A.5). Since $w > 1$ type 1 is for certain served when $t < 3/2$. Together with the preceding statements this completes the proof of Proposition 3.

B Pricing under a three part tariff

The maximization problem is given by

$$\begin{aligned}\Pi = & \max_{p_1^n, \bar{p}_1^n, p_2^n, E_1^n, E_2^n} \Big\{ E_1^n + p_1^n(\theta_1 - p_1^n) F_1(n) + \\ & \bar{p}_1^n(\theta_1 - \bar{p}_1^n)(1 - F_1(n)) + E_2^n + p_2^n(\theta_2 - p_2) \Big\}\end{aligned} \tag{B.1}$$

subject to $p_i^n \geq 0, \bar{p}_1^n \geq 0, E_i^n \geq 0$ $(i = 1, 2)$, IR_1, and IC_2:

$$
\begin{aligned}
E_1^n &= \int_{p_1^n}^{\theta_1} (\theta_1 - p) F_1(n)\, dp + \int_{\bar{p}_1^n}^{\theta_1} (\theta_1 - p)(1 - F_1)(n)\, dp, && \text{(B.2)} \\
E_2^n &= E_1^n + \int_{p_2^n}^{\theta_2} (\theta_2 - p)\, dp - \int_{p_1^n}^{\theta_2} (\theta_2 - p) F_2(n)\, dp - \\
&\quad \int_{\bar{p}_1^n}^{\theta_2} (\theta_2 - p)(1 - F_2)(n)\, dp. && \text{(B.3)}
\end{aligned}
$$

The unit prices that solve the maximization problem above is the following

$$
\begin{aligned}
p_2^n &= 0 && \text{(B.4)} \\
p_1^n &= \theta_2 - \theta_1 \tfrac{F_1(n)}{F_2(n)} && \text{(B.5)} \\
\bar{p}_1^n &= \theta_2 - \theta_1 \tfrac{1 - F_1(n)}{1 - F_2(n)} && \text{(B.6)}
\end{aligned}
$$

The firms profit under the different pricing strategies are given by the following expressions. If the firm offers a three part tariff to low demand types

$$
\Pi_{3p} = \begin{cases} \frac{1}{2}\theta_2^2 (1 - F_2) + \theta_1 (\theta_1 - \theta_2 (1 - F_1)) + \frac{1}{2}\theta_1^2 \frac{(1-F_1)^2}{(1-F_2)} & \text{if } p_1^n = 0 \\ \frac{1}{2}\theta_2^2 - \theta_1 (\theta_2 - \theta_1) + \frac{1}{2}\theta_1^2 \left(\frac{F_1^2}{F_2} + \frac{(1-F_1)^2}{(1-F_2)} \right) & \text{if } p_1^n > 0 \end{cases} \tag{B.7}
$$

If the firm offer a calling circle tariff to low demand types we have the sam profit as in section 3.

$$
\Pi_{2p} = \begin{cases} \theta_1^2 F_1 + \frac{1}{2}\theta_2^2 (1 - F_2) & \text{if } p_1^n = 0 \\ \frac{1}{2}\theta_2^2 - F_1 \theta_1 (\theta_2 - \theta_1) + \frac{1}{2}\theta_1^2 \frac{F_1^2}{F_2} & \text{if } p_1^n > 0 \end{cases} \tag{B.8}
$$

The firm prefers to serve the low demand consumer only with a calling circle tariff rather than a three part tariff if

$$
\frac{\theta_2}{\theta_1} \geq 1 + \frac{1 - F_1}{2(1 - F_2)} \tag{B.9}
$$

and the unit price in the calling circle is equal to marginal cost if

$$
\frac{\theta_2}{\theta_1} \geq \frac{F_1}{F_2} \tag{B.10}
$$

Figure 2 shows the optimal choice of n for any given heterogeneity in call duration, $\theta_2/\theta_1 = t$. That is, the curves represents the condition $d\Pi/dn = 0$, where the relevant profit function in B.7 and B.8 are chosen according to B.9 and B.10.

References

Armstrong, M. (1999). "Price discrimination by a many-product firm." *Review of Economic Studies*, **66**(1), 151–168.

Bakos, Y. and E. Brynjolfsson (1999). "Bundling information goods: Pricing, profits and efficiency." *Management Science*, **45**(12), 1613–1630.

Caillaud, B., R. Guesnerie, P. Rey and J. Tirole (1988). "Government intervention in production and incentives theory: A review of recent contributions." *RAND Journal of Economics*, **19**(1), 1–26.

Deneckere, R. J. and R. P. McAfee (1996). "Damaged goods." *Journal of Economics and Management Strategy*, **5**(2), 149–174.

Foros, Ø., S. Jensen and J. Y. Sand (1999). "Damaging network subscription." *Communications and Strategies*, **33**, 37–58.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. MIT Press, Cambridge.

Goldman, M. B., H. E. Leland and D. S. Sibley (1984). "Optimal nonuniform prices." *Review of Economic Studies*, **51**, 304–319.

Laffont, J.-J., E. Maskin and J. Rochet (1987). "Optimal nonlinear pricing: The case of buyers with several characteristics." In T. Groves, T. Radner and S. Reiter, eds., *Information, Incentives and Economic Mechanisms: In Honour L. Horwicz*. University of Minnesota Press.

Laffont, J.-J., P. Rey and J. Tirole (1998). "Network competition: II. Price discrimination." *RAND Journal of Economics*, **29**(1), 38–56.

Maskin, E. and J. Riley (1984). "Monopoly with incomplete information." *RAND Journal of Economics*, **15**(2), 171–196.

Miravete, E. J. (2001). "Screening through bundling." CARESS Discussion Paper No. 01-01.

Mirman, L. M. and D. Sibley (1980). "Optimal nonlinear prices for multiproduct monopolies." *The Bell Journal of Economics*, **11**, 659–670.

Mussa, M. and S. Rosen (1978). "Monopoly and product quality." *Journal of Economic Theory*, **18**, 301–317.

Oren, S. S., S. A. Smith and R. B. Wilson (1982). "Nonlinear pricing in markets with interdependent demand." *Marketing Science*, **1**(3), 287–313.

Rochet, J.-C. and L. A. Stole (2003). "The economics of multidimensional screening." Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress.

Sappington, D. (1983). "Optimal regulation of a multiproduct monopoly with unknown technological capabilities." *Bell Journal of Economics*, **14**, 453–463.

Sharkey, W. W. and D. S. Sibley (1993). "Optimal nonlinear pricing with regulatory preference over customer type." *Journal of Public Economics*, **50**(2), 197–229.

Shi, M. (2003). "Social network-based discriminatory pricing strategy." *Marketing Letters*, **14**(4), 239–256.

Sibley, D. S. and P. Srinagesh (1997). "Multiproduct nonlinear pricing with multiple taste characteristics." *RAND Journal of Economics*, **28**(4), 684–707.

Stole, L. A. (2001). "Price Discrimination in Competitive Environments." Forthcoming in *The Handbook of Industrial Organization.*

Tirole, J. (1988). *The Theory of Industrial Organization.* MIT Press, Cambridge.

Wang, R. and Q. Wen (1998). "Strategic invasion in markets with switching costs." *Journal of Economics and Management Strategy*, **7**(4), 521–549.

Wilson, R. (1993). *Nonlinear Pricing.* Oxford University Press, Oxford.