Two-Part Tariffs with Partial Unbundling*

Sissel JENSEN†

Discussion paper 19/2001
Norwegian School of Economics
and Business Administration

October 2001

Abstract

The paper explores second degree price discrimination in a multi-dimensional good context. There are two types of consumers with demand described by a two-dimensional vector, a quantity dimension and a service attribute dimension (mode of usage, usage pattern). The adverse selection parameter determines the consumers' willingness to pay for quantity increments with a certain set of attributes. The multi-dimensionality is exploited by forcing a restriction on the mode of usage towards consumers with low willingness to pay in order to make it less tempting for high types to mimic a low type. We show that the firm introduces distortions in the use of the service against a decrease in the quantity distortions in the low-type's contract.

JEL.CLASS.NO: D42, D82, L96
KEYWORDS: Price discrimination, two part tariffs, quantity discounts, multidimensional goods, 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, and at the 3rd Nordic Workshop in Industrial Organization (NORIO III) in Helsinki for helpful comments. Financial support from Telenor is gratefully acknowledged.

†Norwegian School of Economics and Business Administration, Department of Economics and Telenor Research and Development. Mailing address: Helleveien 30, 5045 Bergen, Norway. Phone: 55 95 92 81, Internet: sissel.jensen@nhh.no.

1 Introduction

Consumers are often heterogeneous along numerous dimensions. In telecommunications, for example, consumers differ with respect to the quantity they purchase (minutes called) as well as in their calling pattern (whom they call, when they call, the duration of each call, etc.). With a few notable exceptions, the literature does not address the question on how a monopoly should price discriminate in a market with multidimensional heterogeneity.[1] The purpose of this paper is to explore how a monopoly might use two instruments to enhance the profitability from second degree price discrimination.

When consumers' willingness to pay is private information and the firm must condition the contract upon observable variables, it is most often assumed that the firm can observe only one variable. It is also common to assume that the observed variable is single dimensional, e.g., quality in Mussa and Rosen (1978) and quantity in Maskin and Riley (1984). Although being welfare improving compared to uniform pricing, the performance of second degree price discrimination relative to first best practice (marginal-cost pricing) depends upon the degree of the demand side heterogeneity.[2] In general therefore, any strategy that increases the observability of consumers' willingness to pay will increase profit and welfare.

In the present paper a monopoly firm sells a single generic good, for instance minutes of network usage. Each quantity increment can be assigned a unique list of observable attributes, such as time-of-day, distance, call termination point, etc, and this describes a consumer's calling pattern. By monitoring consumers' calling patterns, the firm is able to offer a tariff intended for low demand consumers on terms that differ from the terms on which high demand consumers make their purchases. Such practice can potentially improve "the observability" of consumers' characteristics in terms of self-selection, and thus implies less distortions towards low demand consumers.

It is possible to translate the multidimensionality implied by differences in

[1] Work on multi-dimensional screening includes different kinds of problems. One polar case is when consumers are described by several characteristics but the firm has only one instrument at hand, references are Laffont, Maskin and Rochet (1987), Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995), and Jullien (2000). The other polar case is when consumers are described by one characteristic but the firm can use several instruments, as in Matthews and Moore (1987) with risk-aversion, or as in Sappington (1983), and Caillaud, Guesnerie, Rey and Tirole (1988) with several observables and instruments. Rochet and Choné (1998), and Armstrong and Rochet (1999) work within a model with several instruments and several characteristics, also providing an overview of the literature. Wilson (1993) provides definitions and examples of multidimensional goods and multidimensional pricing. Deneckere and McAfee (1996) and Foros, Jensen and Sand (1999) present models similar to the one presented in this paper; Deneckere and McAfee with uniform pricing and Foros et al. with nonlinear pricing, but in both articles the restriction on usage is exogenous.

[2] The price-cost margin depends on the range of the type-space and on the firm's prior beliefs about the distribution of types over this space. If the heterogeneity on the demand side is large then a large fraction of consumers pay a price well above marginal cost.

calling patterns into a multiproduct setting by letting units assigned different sets of attributes be treated as different products. If the firm ignores the heterogeneity in consumers' calling patterns but charge all units the according to the same tariff, it aggregates all taste parameters and practice complete bundling. In the present context we let the firm bundle a subset of the products and charge units within this product bundle according to a different tariff. The firm do not debundle completely, and hence, we refer to this practice as partial bundling.[3]

We hold on to the assumption that consumers differ in their marginal willingness to pay for quantity and say that there are two types of consumers, high demand and low demand consumers. 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. High-dispersion subscribers can be thought of as business consumers while low dispersion subscribers can be thought of as residential consumers. The firm offers a menu of two-part tariffs, each specifying a fixed fee that must be paid up-front, a marginal usage price, and in addition use-of-service restrictions which consumers must obey. Customers choose their preferred tariff scheme and usage is subsequently billed according to this choice.

The predominant method of charging consumers for telecom usage has been to bill for the length of time a connection is used. All multi-dimensionality in the consumers' use of the service was translated into a single-dimensional quantity variable (pulses, and later minutes). The practice of sorting consumers with different willingness to pay for usage was handled by giving high demand consumers quantity discount, in consistence with single-dimensional screening models. Today the multi-dimensionality in usage patterns is to an increasing extent used to achieve separation. Tariff options known as Friends and Family and Best Friend are examples of discounts given on certain calling patterns. Other examples are telecom companies that offer discounts on dial-up internet access, in the form of discounts on standard calling rates or in the form of a monthly fixed fee for a fixed number of hours of usage, (flat rate dial-up internet access).[4]

Firms' use of calling circle tariffs has received some attention in other areas in the economics literature. Wang and Wen (1998) consider a duopoly model with demand side heterogeneity, where such pricing behavior enables a new firm to

[3]In the multiproduct setting it would be the case that although the firm has imperfect knowledge about a given consumer's taste for one product, it knows that it is perfectly correlated with the taste for any other product. Miravete (2001) study multidimensional screening where different type components distinguish quality dimensions of products that can be aggregated. The ability to aggregate type components opens the possibility to reduce the dimensionality of the screening process. 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.

[4]Dial-up internet access is in this way singled out as a separate product.

enter the market despite the presence of consumer switching costs. This result is derived under specific assumptions about consumer calling patterns, specifically that low demand types make calls to other low demand types, whereas high demand types make calls to other high demand types. By relaxing this assumption one might conclude differently (see Klemperer (1995) for a survey on the switching cost literature). Laffont, Rey and Tirole (1998) examine the effects of discriminatory pricing on the negotiated interconnection agreements between rival network operators. When a network operator charges different prices for calls terminating on the subscriber's network and those terminating on a rivals network he can generate network externalities despite network interconnection.

Throughout the paper we shall hold on to a simple example applied to telecommunications and assume that consumers with different willingness to pay for the service have distinctly different calling patterns. Section 2 describes the generic features of telecommunications that are relevant for this paper. The aim in Section 3 is to give a definition of the quantity variable that the usage charge in a two-part tariff applies to. Section 4 presents the demand- and supply side conditions of the market as well as the informational constraints faced by the firm. Finally, in section 5 we draw conclusions from the analysis.

2 Telecommunications services

The telecommunications market has experienced rapid changes during the last decades. A large variety of services are nowadays provided on a common platform and the technology convergence gives rise to significant changes in the demand side of the market as well. New services at reasonable prices and more multifunctional customer premises equipment, for instance the world wide web, personal computers, and all applications on the web, have also led to large increases in the demand for transmission capacity (time length or more bandwidth). Built-in network intelligence and sophisticated monitoring of usage have enabled firms in the market to move from billing customers for single dimensional pulses to billing multidimensional minutes. The method of pricing a call used to be by a conversion from hour-of-day, day-of-week, distance, etc, to pulses by tables in the central office. The firm had no information about a consumer's demand other than the number of pulses consumed at the end of the billing period. Network technology and billing systems now price a call minute according to a detailed call record. Hence, the firm possesses very detailed information about a consumer's usage pattern.

The telecommunications network is a two-way network and a person or a machine that is present at one specific node asks for some type of communication with another specific node at some given hour, weekday, etc.[5] Even though a

[5]A phone call, an e-mail, a web-site etc with an objective to exchange, deliver or gather information). In this respect telecommunications is very different from other network industries,

one minute call within a specific calling area is a perfectly standardized product, its point of destination is of vital importance to the consumer who makes the call. A consumer does not derive any benefit from a call which destines at a B-subscriber he did not intend to call.[6] The same feature also applies to information services generated at a specific network node. These are features of telecommunications that make it a multi-dimensional good. For instance, individual call records usually contain information about at what hour the call is made, who is the B-subscriber, and where the B-subscriber is located (local, long distance, international). Furthermore, subscribers are typically billed according to aggregate minutes (seconds) of peak-time long distance calling, off-peak long distance calling, peak-time international calling, etc. There are examples of service attributes that have an obvious ranking, e.g. if the attributes are different quality levels along a vertical dimension they are ranked the same way by all customers. However, this is not always the case with telecommunications. For instance, not all consumers prefer – at an equal price – to make calls at the same time of day. Service attributes such as the time-of-day or the node where a call terminates are attributes along a horizontal dimension and customers will rank them differently.

A widespread practice is to offer various kinds of calling circle tariffs. Under a calling circle tariff, a subscriber is billed according to aggregate minutes (seconds) of calling to specific B-subscribers (or specific network nodes), and the marginal price varies conditional on the node of termination. The following model aims at explaining and guiding the construction of such tariffs.

3 A two-dimensional good

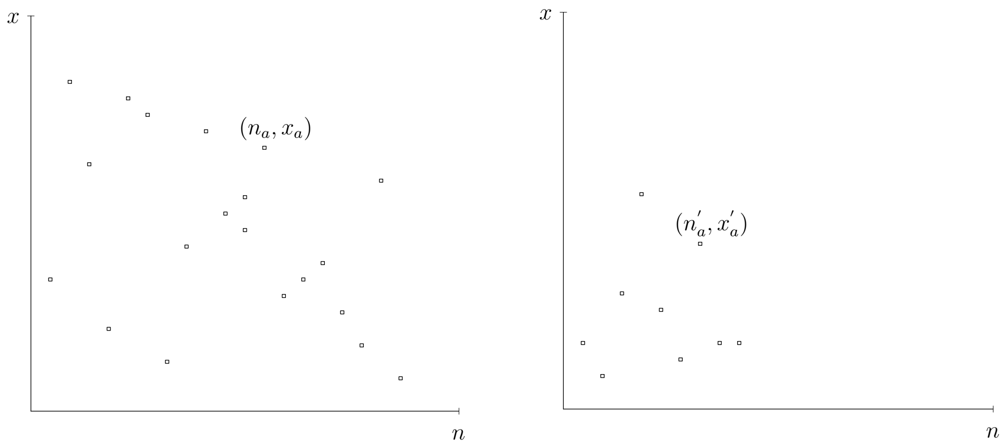
Let q be a two-dimensional good $q = (n, x)$, where x is a quantity variable and n is a service attribute.[7] The vector q tells us how many units (x) with the given service attribute (n) a consumer did buy. When we sum over all possible service attributes (i.e., over every possible n) the sum is equal to a consumer's total demand, i.e., aggregate units of the generic good. This is to collect and sum up

like electricity or water delivery. One kW/h of electricity injected at one point is a perfect substitute for one kW/h injected at any other point of generation.

[6] The B-subscriber is the party being called whereas the A-subscriber is the party making the call.

[7] A context with multidimensional products is similar to a multiproduct context since units assigned different sets of attributes can be treated as different products. The distinction between a multiproduct context and a multidimensional product context is that the consumer is allowed to custom design the service attributes by assigning each item his preferred attributes (termination node/B-subscriber, time of day, day of week etc) instead of choosing between a fixed and more constrained number of products. See Wilson (1993, part 3) for a description of multidimensional products and multidimensional pricing and for instance Armstrong (1996) on multiproduct pricing.

the x's at every point in figure 1(a) (or 1(b)). Figure 1 gives two examples on representation of q. In the figure, x_a is the number of minutes a high demand consumer called network node n_a, and x'_a is the number of minutes a low demand consumer called network node n'_a. Note that the n-axis merely gives the identity of the party called (phone numbers) and is not ordered in any sense.



(a) High demand consumer

(b) Low demand consumer

Figure 1: *Demand bundles. High demand consumers make calls of longer duration and to a larger number of network subscribers*

We represent a consumer's purchase set $\mathbb{Q}$ by sorting along the attribute dimension, and describe this set with a "continuous boundary" $x(n)$. Using the telephone example again, and saying that the attribute assignment is network node (B-subscriber), sorting along the attribute dimension gives a presentation of the most called number, the second most called number, and so on.[8] We introduce heterogeneity on the demand side by assuming that a consumer's willingness to pay for the good is characterized by a privately known parameter θ, measuring the intensity of a consumer's valuation of quantity. A consumer θ includes in his purchase set $\mathbb{Q}$ all increments for which his valuation $v(q, \theta)$ exceeds the marginal price p charged

$$\begin{aligned} \mathbb{Q}(\theta; p) &= \{q : v(q, \theta) \geq p\} \\ &= \{(n, x) : v((n, x), \theta) \geq p\}. \end{aligned} \tag{1}$$

The boundary around the set given by (1) is those points where the marginal

[8]Such a presentation is only possible if the service attributes can be interpreted as cardinal levels and if n is continuous. The various attributes can not be along dimensions such as for instance color.

valuation equals the marginal price. Marginal valuation is given by

$$\begin{aligned} v\left(\left(n,x\right),\theta\right) &= \frac{\partial^2 U\left(.\right)}{\partial n \partial x}, & (2)\\ v\left(\left(n,x\right),\theta_1\right) &\leq v\left(\left(n,x\right),\theta_2\right), \text{when } \theta_1 < \theta_2. \end{aligned}$$

Type θ gains gross utility from consuming the purchase set $\mathbb{Q}\left(\theta;p\right)$ given by the double integral

$$U\left(\mathbb{Q}\left(\theta;p\right)\right) = \iint_{\mathbb{Q}} v\left(\left(n,x\right),\theta\right) dndx. \tag{3}$$

By saying that the boundary around the set can be represented by a monotonic function $x(n;\theta,p)$, which is continuous and everywhere differentiable, we can derive the demand from consumer type θ by solving a single integral over the attribute dimension. Aggregate demand for the generic good over all possible attribute levels is given by

$$Q\left(p,\theta\right) = \int_0^\infty x\left(s;\theta,p\right) ds, \tag{4}$$

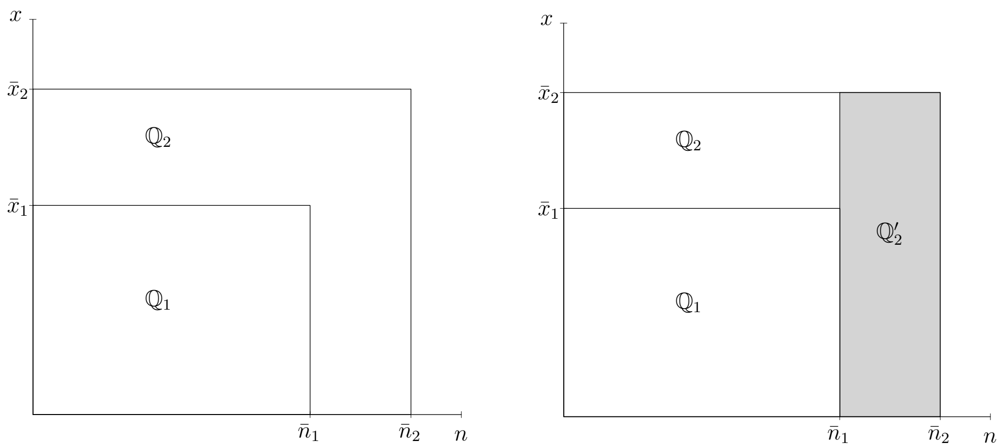
and we define $Q_i(p) \equiv Q(p,\theta_i)$, $i = 1,2$. Aggregate demand for the generic good with attribute level $\bar{n}$ or lower, i.e., aggregate calls to the $\bar{n}$ most called network nodes, is given by

$$\bar{Q}\left(p,\theta,\bar{n}\right) = \int_0^{\bar{n}} x\left(s;\theta,p\right) ds, \tag{5}$$

and similarly we define $\bar{Q}_i(p,\bar{n}) \equiv Q(p,\theta_i,\bar{n})$, $i = 1,2$. If a consumer can customize demand freely, demand is given by (4). If he is to choose attribute levels within the interval $[0,\bar{n}]$, demand is given by (5), and $Q_i\left(p\right) \geq \bar{Q}_i\left(p,\bar{n}\right)$, $i = 1,2$. In the following we assume that $x\left(\bar{n};\theta_1,p\right) < x\left(\bar{n};\theta_2,p\right)$, $\forall p,\bar{n}$ for $\theta_1 < \theta_2$ and also that $x\left(\bar{n};\theta,p\right)$ is monotonic. Hence $Q_2\left(p\right) > Q_1\left(p\right)$ and $\bar{Q}_2\left(p,\bar{n}\right) > \bar{Q}_1\left(p,\bar{n}\right)$, $\forall p,\bar{n}$. Further, $\bar{Q}_i\left(p,\bar{n}\right)$ is nonincreasing in p and nondecreasing in $\bar{n}$, and $Q_i\left(p\right)$ is also nonincreasing in p, $i = 1,2$.

Using telephony as an example, heterogeneity in consumer demand is given by differences in call duration and call dispersion. We define call dispersion according to a cumulative distribution $F_1\left(n\right) \geq F_2\left(n\right)$ with a probability density function $f_i\left(n\right)$, $i = 1,2$.[9] Hence, we make the assumption that call dispersion is independent of the price per call minute and that type 2 has a more dispersed calling pattern compared to type 1. Since we are only interested in the 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.

[9] The distribution of n conditional on θ_2 first-order stochastically dominates the distribution of n conditional on θ_1, if $\theta_2 \geq \theta_1$. For notation we use $f_i\left(n\right) \equiv f\left(n;\theta_i\right)$, $F_i\left(n\right) \equiv F\left(n;\theta_i\right)$



(a) Purchase set (b) Relinquish set

Figure 2: *Rectangular purchase set (a) and the effect on type 2's purchase set of a restriction in call dispersion (b)*

Figure 2 gives an illustration in the case of a rectangular purchase set. With a rectangular purchase set we have implicitly assumed that $f_i(n)$ is uniform on $[0, \bar{n}_i]$, $F_i(n) = n/\bar{n}_i$ ($\bar{n}_i = \{\bar{n}_1, 1\}$). The height of the rectangle measures the number of call minutes x to network node n. Type 1 makes calls to $\bar{n}_1$ different network nodes, whereas type 2 makes calls to $\bar{n}_2$ different network nodes, i.e., $\bar{n}_i$ is a measure of call dispersion. Figure 2 above reflects that there is heterogeneity in both call duration and call dispersion. If $\bar{x}_1 = \bar{x}_2$ all heterogeneity would be in call dispersion, whereas $\bar{n}_1 = \bar{n}_2$ would describe a situation with all heterogeneity in call duration. The shaded area $\mathbb{Q}_2'$ in figure 2 represents the part of type 2's ideal purchase set that he has to give up if he chooses a tariff with a restriction in call dispersion $\bar{n}_1$.

4 The model

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, type 1 with low willingness to pay and type 2 with high willingness to pay. A consumer's type is unobservable to the firm but each type's preferred calling pattern is known. We assume that type 2 has a more dispersed calling pattern than type 1. The types' call dispersion $f_i(n)$ is exogenous. The reservation utility is assumed to be equal for the two consumers and normalized to zero.

Because call dispersion is independent of the marginal price of a call minute, we can also write consumers' utility as a function independent of call dispersion. We use the following utility function that is quasilinear and quadratic in x[10]

$$U_i = \begin{cases} \theta_i x - \frac{1}{2}x^2 - T & \text{if they pay } T \text{ for } x \text{ minutes of calling,} \\ 0 & \text{if they do not buy.} \end{cases} \tag{6}$$

T is an increasing and continuous price schedule with a constant unit price $p = \{p_1, p_2\}$. Given information about each type's call dispersion, we derive expected call length to a network node n as $(\theta_i - p) f_i(n)$. Consumers' demand is thus given by

$$Q_i(p) = \int_0^1 (\theta_i - p) f_i(n)\, dn = (\theta_i - p), \tag{7}$$

$$\bar{Q}_i(p, n) = \int_0^n (\theta_i - p) f_i(n)\, dn = (\theta_i - p) F_i(n). \tag{8}$$

The density function f_i is positive and integrable on the support $n \in [0, 1]$ with a distribution function $F_i(n)$ with $F_1(n) > F_2(n)$, and $f_1 F_2 \leq f_2 F_1$. Aggregate demand for call minutes to the entire network is given by (7) and aggregate demand for call minutes to the n most frequently called nodes is given by (8). The latter case resembles the first, 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).[11]

Consumer surplus under a two-part tariff $T_i = \{p_i, E_i\}$ for some given $n \leq 1$ is given by

$$CS_i(p_i, E_i, n) = \int_{p_i}^{\theta_i} (\theta_i - p) F_i(n)\, dp - E_i, \quad i = 1, 2, \tag{9}$$

$$CS_2(p, E, n) > CS_1(p, E, n). \tag{10}$$

When both types choose consumption subject to the same tariff, type 2 obtains a larger surplus given that $F_1(n)/F_2(n) \leq \theta_2/\theta_1$. Under this condition the demand

[10]We abstract from the fact that some consumers may have positive utility even in the case when consumption is zero. A subscriber may want a network connection in order to receive calls only, or to be able to make emergency calls. Our assumption in this model is that if the expected net utility from making calls weakly exceeds a consumer's reservation utility he will find it beneficial to subscribe to the network. By assuming quasilinear utility we also ignore income effects.

[11]Laffont et al. (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.

curves of the two types never cross for any price. Since the demand curves are linear and $\theta_2 \geq \theta_1$, it is sufficient to evaluate the condition $(\theta_2 - p)F_2(n) \geq (\theta_1 - p)F_1(n)$ as p approaches zero.

When we solve the model we 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 restriction in call dispersion in the two-part tariffs T_1 and T_2.

5 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 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 $\{p_i, E_i; n_i\}$, p_i is the marginal price, E_i is a fixed fee and $n_i \leq 1$ 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 $\{p, E; n_i\}$ 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.[12]

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 the restriction on n is fixed we have the following maximization problem

$$\Pi = \max_{p_1, p_2, E_1, E_2} \left\{ E_1 + p_1\left(\theta_1 - p_1\right) F_1\left(n\right) + E_2 + p_2\left(\theta_2 - p_2\right) \right\} \tag{11}$$

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

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

$$E_1 = \int_{p_1}^{\theta_1} (\theta_1 - p) F_1(n)\, dp, \tag{12}$$

$$E_2 = E_1 + \int_{p_2}^{\theta_2} (\theta_2 - p)\, dp - \int_{p_1}^{\theta_2} (\theta_2 - p) F_2(n)\, dp. \tag{13}$$

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 (13) 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 can 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 restriction in call dispersion, Q is treated as a single dimensional good, $n_1 = n_2 = 1$. 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 too 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

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

(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 (11)–(13). 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).

Now we turn to the case of a wider strategy set, i.e., where the tariff intended for type 1 may have a restriction in call dispersion. Type 1 can only reach a limited number of call termination points (a fraction n of the full network). According to (7) and (8) 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. Type 2 is less tempted by type 1's tariff if he cannot reach the entire network and he is less tempted when the unit price in type 1' s tariff is high. Although type 1 also suffers under such distortions, he is not as seriously affected as type 2. 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).

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. For a given restriction n, type 2 is offered a cost-plus-fixed-fee tariff $\{0, E_2^n\}$ and type 1 is offered a two-part tariff $\{p_1^n, E_1^n, n\}$, $n \leq 1$*

$$p_1^n = \theta_2 - \theta_1 \frac{F_1(n)}{F_2(n)},$$

and where the fixed fees E_1^n and E_2^n are determined by (12) and (13).

Under our assumptions on F_1 and F_2, p_1^n is nondecreasing in n, continuous, and differentiable whenever $p_1^n > 0$. 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).

On the other hand, both instruments are costly to use in the sense that type 1's consumption is de facto restricted (whereas type 2's consumption is restricted only if he opts for type 1's tariff). 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. 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.

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

$$\Pi(n) = \begin{cases} \frac{1}{2}\theta_2^2 + \frac{1}{2}\theta_1^2 \frac{F_1(n)^2}{F_2(n)} - F_1(n)\,\theta_1(\theta_2 - \theta_1) & \text{if} \quad p_1^n > 0, \\ \\ \theta_1^2 F_1(n) + \frac{1}{2}\theta_2^2(1 - F_2(n)) & \text{if} \quad p_1^n = 0. \end{cases} \tag{14}$$

The firm maximizes profit with respect to n and the tariffs are determined by Lemma 2. If the heterogeneity in quantity type is large relative to the heterogeneity in call dispersion, the firm will offer type 1 consumers a two-part tariff with a restriction in call dispersion together with a distorted unit price. In the opposite case the firm will offer type 1 consumers flat-rate pricing with restriction in call dispersion. Whenever there is heterogeneity in the types' calling pattern the firm will restrict type 1's calling.

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*

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

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 first inequality in part (i) of Lemma 3 states the condition for $p_1^n > 0$, whereas the last inequality in part (ii) of Lemma 3 states the condition for $p_1^n = 0$. In the latter 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 cost-plus-fixed-fee tariff, i.e., when θ_2/θ_1 is small and/or F_1/F_2 is large. 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. We do this in the following two sections. For simplicity we assume that type 2 makes calls of equal length to all nodes, i.e., $f_2(n)$ is uniformly distributed on the interval $[0, 1]$. Regarding type 1's call dispersion we assume two different cases, call dispersion is described either by the uniform distribution or by a Beta distribution.

5.1 Uniform distribution

Call dispersion for type 1 is uniformly distributed on the interval $[0, \bar{n}_1]$, and call dispersion for type 2 is uniformly distributed on the interval $[0, 1]$, $0 < \bar{n}_1 < 1$.

The marginal unit price is $p_1^n = \theta_2 - \theta_1 \frac{1}{\bar{n}_1}$. For $p_1^n > 0$, the derivative of the firm's profit with respect to n can be written as

$$\frac{d\Pi}{dn} = \begin{cases} -\dfrac{\theta_1^2}{2n^2} & \text{if} \quad \bar{n}_1 \leq n \leq 1, \\ -\dfrac{\theta_1 (\theta_2 - \theta_1)}{\bar{n}_1} + \dfrac{\theta_1^2}{2\bar{n}_1^2} & \text{if} \quad 0 < n < \bar{n}_1. \end{cases} \tag{15}$$

And if $p_1^n = 0$ we have $\Pi = E_1^n + E_2^n$, and the derivative with respect to n is

$$\frac{d\Pi}{dn} = \begin{cases} -\frac{1}{2}\theta_2^2 & \text{if} \quad \bar{n}_1 \leq n \leq 1, \\ \dfrac{1}{\bar{n}_1}\theta_1^2 - \frac{1}{2}\theta_2^2 & \text{if} \quad 0 < n < \bar{n}_1. \end{cases} \tag{16}$$

The profit function is linear for $n \in [0, \bar{n}_1)$ but the sign of the derivative is ambiguous, for $n \in (\bar{n}_1, 1]$ profit decreases in n. The optimal restriction in call dispersion will be one of the extremes $n^* = 0$ or $n^* = \bar{n}_1$. In the first case type 1 is *de facto* excluded. Henceforth, we define a variable $t \equiv \frac{\theta_2}{\theta_1}$. The propositions that follow describe the monopoly's pricing strategy.

Proposition 1 *If heterogeneity in call dispersion is sufficiently large relative to heterogeneity in call duration, $n^* = \bar{n}_1$ and type 1 is served with a cost-plus-fixed-fee tariff $\{0, E_1^n, \bar{n}_1\}$. For $t \in [1, 2]$ this occurs for $\bar{n}_1 \leq \frac{1}{t}$, for $t > 2$ it occurs for $\bar{n}_1 \leq \frac{2}{t^2}$.*

Proposition 1 shows that a restriction in call dispersion in type 1's tariff may be sufficient to separate the types. Consumers with different willingness to pay are charged identical unit price, but type 2 pays a larger fixed fee. In terms of pricing, this looks like first degree price discrimination. For $t \leq \frac{3}{2}$, type 1 is served with a restriction in call dispersion instead of with a distortion in the unit price, for $t > \frac{3}{2}$, type 1 is served with a restriction in call dispersion instead of being excluded.

Proposition 2 *If demand side heterogeneity is more moderate and balanced, type 1 is served with a two-part tariff $\{p_1^n, E_1^n, \bar{n}_1\}$, and $p_1^n > 0$. This occurs for $\bar{n}_1 > 0.5$ and $t < 2$, and t such that $\frac{1}{\bar{n}_1} \leq t \leq 1 + \frac{1}{2\bar{n}_1}$.*

A restriction in call dispersion will always be used, either alone ($p_1^n = 0$) or in combination with a restriction on usage via distortionary pricing.

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* $t > \sqrt{2/\bar{n}_1}$ *for* $\bar{n}_1 \in [0, 0.5)$*, or for* $t > 1 + \frac{1}{2\bar{n}_1}$ *for* $\bar{n}_1 \in [0.5, 1]$*. Type 2 is served with a cost-plus-fixed-fee tariff,* $\{0, E_2, 1\}$*, which extracts the entire social surplus. Type 1 is served in more cases relative to the single-dimensional case.*

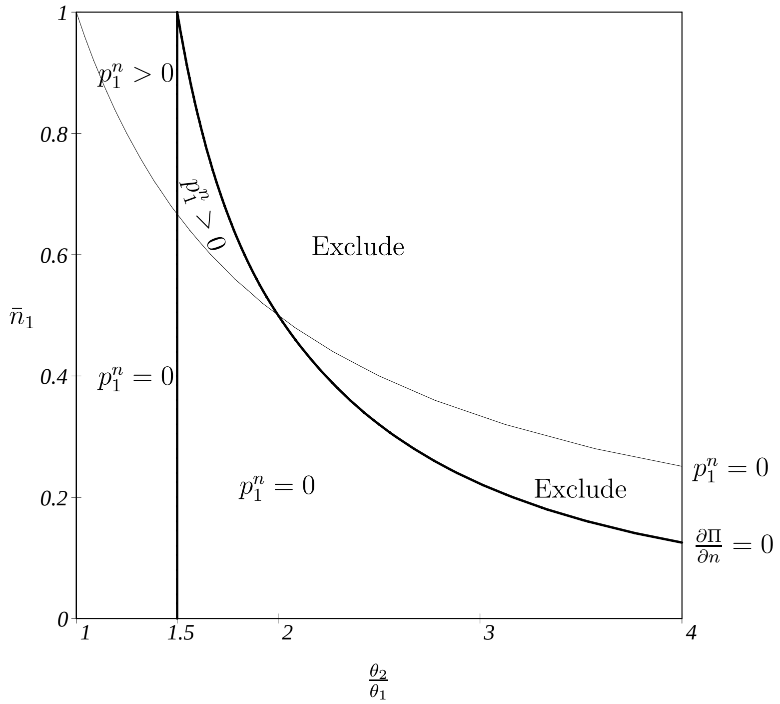


Figure 3: *Pricing policy towards type 1 depending on the heterogeneity along the two dimensions. The larger the heterogeneity in call dispersion (low* $\bar{n}_1$*) the larger is the possibility that type 1 is served and that he is served with an efficient tariff, i.e., a cost-plus-fixed-fee tariff.*

Although increased heterogeneity in call dispersion reduces the incentive to exclude type 1, proposition 3 states that this incentive still exists.[13] The proofs of the propositions are given by simple calculations that are shown in the appendix. Figure 3 illustrates the results.

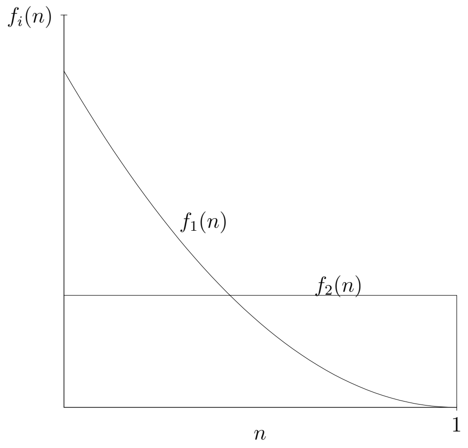
The effect of a reduction in call dispersion is that the firm can give informational rent to type 2 as if the types were described according to the demand curves $\bar{Q}_i(p, \bar{n}_1)$, but extract gross surplus from type 2 according to the demand curve $Q_2(p)$. Typically, the possibility of type 1 being served increases as $\bar{n}_1$ decreases

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

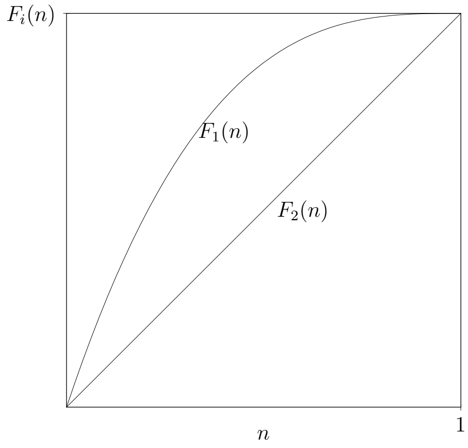
because this increases the 'observability' of the two types. The generalization of this is the fact that the firm is always better, or at least equally well, off with an additional observable and instrument at hand.[14]

5.2 Beta distribution

The Beta-distribution allows for the possibility that the call length may vary over n, i.e., over points of call termination. That is, call termination points are ordered according to the most called number, the second most called number etc. We keep the simplification that type 2's calling is uniformly distributed on $[0, 1]$ but say that type 1 has a more concentrated calling pattern by using the Beta distribution and placing more probability weight to the left tail of the distribution. Figure 4 illustrates this difference between the types.



(a) Probability density



(b) Cumulative density

Figure 4: *Probability distribution over n, the uniform distribution and the Beta distribution with* $v = 1$

The probability density function for the beta distribution is

$$f(n, v, w) = \begin{cases} \dfrac{n^{v-1}(1-n)^{w-1}}{B(v, w)} & \text{if } 0 \leq n \leq 1 \\ \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

[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 et al. (1988) generalize the case with several observable variables.

where the shape parameters v and w are positive numbers. The denominator $B(v,w)$ is the Beta function. With $v = 1$, the shape of the distribution is determined by w, the higher is w the larger is the mass for low n. We can redefine the distributions for type 1 by fixing v to be 1 and letting w vary ($w = 1$ is the uniform distribution on $[0,1]$). The p.d.f and the c.d.f. are defined by

$$f_1(n,w) = \begin{cases} w(1-n)^{w-1} & \text{if } 0 \leq n \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

$$F_1(n,w) = \begin{cases} 1-(1-n)^{w} & \text{if } 0 \leq n \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

The probability density and cumulative density functions $f_2(n)$ and $F_2(n)$ are the same as before. The firm seeks to maximize profit with respect to n according to the optimality condition in Lemma 3. The monopoly's pricing strategy is given in the following propositions.

Proposition 4 *If heterogeneity in call dispersion is sufficiently large relative to the heterogeneity in call duration, both types are served with a cost-plus-fixed-fee tariff* $\{0, E_1^n, n^*\}$, $n^* \in [n', n'')$. *This occurs for* $t \leq t' \leq t''$. n' *and* n'' *decrease whereas* t' *and* t'' *increase as the heterogeneity in call dispersion increases (*w *increases).*

Proposition 4 is a replication of proposition 1, the larger the heterogeneity in call dispersion, the more powerful is a restriction in call dispersion 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 (reflecting that t'' increases as w increases).

Proposition 5 *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* $t' \leq t \leq t''$, *and* $w \leq 2$.

Proposition 5 is a replication of 2. When the heterogeneity in call duration increases, it is necessary to increase the restriction in call dispersion in order to restore incentive compatibility.

Proposition 6 *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. If* $w = 2$, *then* $t' = t'' = 2$ *and all types that are served are served with a cost-plus-fixed-fee tariff.*

Finally, proposition 6 is a replication of proposition 3. The incentive to exclude low demand consumers still exists when the heterogeneity in call duration is sufficiently large. The propositions 4, 5, and 6 are proved in the Appendix. Figure 5 illustrates the results.

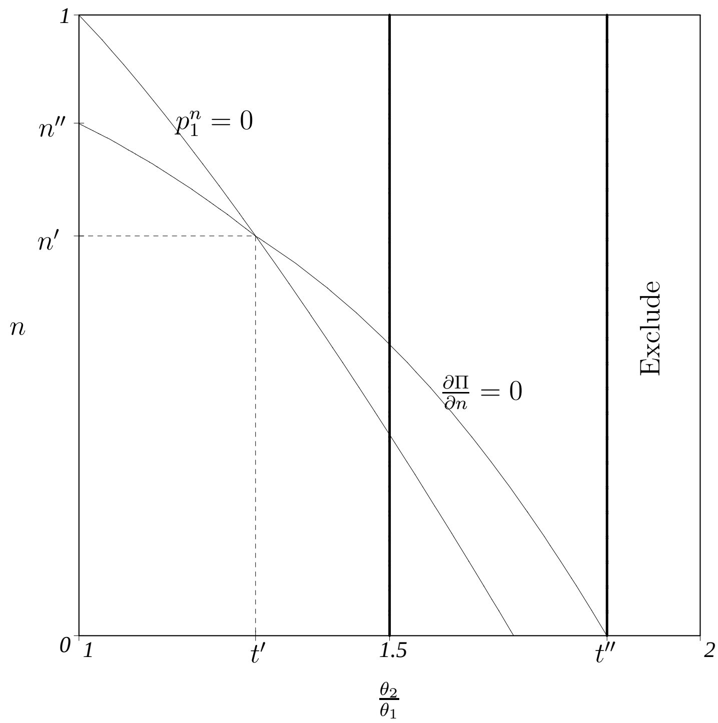


Figure 5: *Pricing policy towards type 1, $w = 1.7$. The larger the heterogeneity in call dispersion (high w) the larger is the possibility that type 1 is served and that he is served with an efficient tariff, i.e., a cost-plus-fixed-fee tariff.*

6 Concluding remarks

In the model presented in this paper, we have assumed that a monopoly firm can use two instruments to achieve second-degree price discrimination. 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 in such a way that high demand consumers are punished more than low demand consumers. 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 above marginal cost, together with a restriction on usage. However, the restriction on usage 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 restriction on usage in type 1's contract. Sometimes, imposing a restriction on usage is sufficient to achieve separation. We also show that the results are qualitatively the same in the case when calls are distributed according to the uniform distribution and the Beta-distribution.

The theoretical model contributes to explain the practice of optional tariffs such as calling circle tariffs, in which the restriction is really severe. However, it should be remarked that promotion of calling circle tariffs might also serve as a strategy to create lock-in effects in duopolistic competition.

Further, 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. Although this paper applies the model to a very simple example within telecoms, the pricing principles derived are of general validity.

Appendix

A.1 Proof of Propositions 1–3

For $p_1^n(\bar{n}_1) \geq 0$ the profit function in (15) is increasing in n if

$$\frac{\theta_2}{\theta_1} \leq 1 + \frac{1}{2\bar{n}_1}. \tag{20}$$

For $p_1^n(\bar{n}_1) = 0$ the profit function in (16) is increasing in n if

$$\frac{\theta_2}{\theta_1} \leq \sqrt{\frac{2}{\bar{n}_1}}. \tag{21}$$

The unit price is positive if

$$\frac{\theta_2}{\theta_1} \geq \frac{1}{\bar{n}_1}. \tag{22}$$

Both conditions (20) and (21) define a curve that is steeper in the $(\theta_2/\theta_1, \bar{n}_1)$ space than does the condition $p_1^n = 0$. Also, $p_1^n = 0$ and $d\Pi/dn = 0$ are binding jointly for $(\theta_2/\theta_1, \bar{n}_1) = (2, \frac{1}{2})$.

$$\left.\frac{d\bar{n}_1}{d(\theta_2/\theta_1)}\right|_{p_1^n=0} = -\bar{n}_1^2 \tag{23}$$

$$\left.\frac{d\bar{n}_1}{d(\theta_2/\theta_1)}\right|_{\Pi_n=0} = \begin{cases} -2\bar{n}_1^2 & \text{for} \quad \theta_2/\theta_1 < 2 \\ \\ -\left(\sqrt{\frac{2}{\bar{n}_1}}\right)\bar{n}_1^2 & \text{for} \quad \theta_2/\theta_1 > 2 \end{cases} \tag{24}$$

Proposition 1 is derived by solving for $\bar{n}_1$ in (22) (or (21)) respectively for $t < (>)2$. Proposition 2 is simply given by (20) and (22). Proposition 3 is derived by turning the inequality in (20) for $\bar{n}_1 \in [0, 0.5)$, and by turning the inequality in (21) for $\bar{n}_1 \in [0.5, 1)$. Since $lim_{\bar{n}_1 \to 1}(1 + \frac{1}{2\bar{n}_1}) = \frac{3}{2}$, type 1 is served in more cases relative to the single-dimensional case.

A.2 Proof of Propositions 4–6

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 5. The slopes of these are given by

$$\left.\frac{dn}{dt}\right|_{p_1^n=0} = \frac{n^2}{nf_1 - F_1} \leq 0 \tag{25}$$

$$\left.\frac{dn}{dt}\right|_{\Pi'_n=0} = \begin{cases} \dfrac{\sqrt{2f_1}}{f_{1n}} \leq 0 & \text{if} \quad p_1^n = 0 \\ \\ \dfrac{2n^3 f_1^2}{2f_1(nf_1 - F_1)^2 + nF_1^2 f_{1n}} \leq 0 & \text{if} \quad p_1^n > 0 \end{cases} \tag{26}$$

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 $p_1^n = 0$.

$$\left.\frac{dt}{dw}\right|_{p_1^n=0} = \frac{F_{1w}}{n} \geq 0 \tag{27}$$

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

$$\left.\frac{dt}{dw}\right|_{\Pi'_n=0} = \begin{cases} -\dfrac{f_{1w}}{f_{1n}} & \text{if} \quad p_1^n = 0 \\ \\ \dfrac{F_{1w}}{n} - \dfrac{1}{2n^2}\dfrac{F_1(2F_{1w}f_1 - f_{1w})}{f_1^2} & \text{if} \quad p_1^n > 0 \end{cases} \tag{28}$$

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

$$lim_{n\to 0^+}\left[\left.\frac{dt}{dw}\right|_{\Pi'_n=0}\right] = \begin{cases} \frac{1}{\sqrt{2w}} & \text{if} \quad p_1^n = 0 \\ \\ \frac{1}{2} & \text{if} \quad p_1^n > 0 \end{cases} \tag{29}$$

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

$$\frac{F_1(n',w)}{n'} = \sqrt{2f_1(n',w)} \tag{30}$$

which gives us

$$\frac{dn'}{dw} = -\frac{n^2 f_{1w} - nF_{1w}\sqrt{2f_1}}{n^2 f_{1n} - \sqrt{2f_1}(nf_1 - F_1)} \leq 0 \tag{31}$$

Since we have $t' = \frac{F_1(n',w)}{n'}$, which is monotonic with $dt'/dn' < 0$ (by 25), 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 4 and 5

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

$$lim_{n\to 0^+}\left[\frac{F_1}{n}\right] = w \tag{32}$$

$$lim_{n\to 0^+}\left[\sqrt{2f_1}\right] = \sqrt{2w} \tag{33}$$

$$lim_{n\to 0^+}\left[1+\frac{F_1}{n}(1-\frac{1}{2}\frac{F_1}{f_1})\right] = 1+\frac{1}{2w} \tag{34}$$

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

$$lim_{n\to 0^+}\left[\left.\frac{dt}{dw}\right|_{p_1^n=0}\right] = 1 \tag{35}$$

The shift in (35) is larger than (29). Since $w > 1$ type 1 is for certain served when $t < 3/2$. Together with the preceding statements this completes the proof of Proposition 6.

References

Armstrong, M. (1996). "Multiproduct nonlinear pricing." *Econometrica*, **64**(1), 51–75.

Armstrong, M. and J.-C. Rochet (1999). "Multi-dimensional screening: A user's guide." *European Economic Review*, **43**(4-6), 959–979.

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.

Jullien, B. (2000). "Participation constraints in adverse selection models." *Journal of Economic Theory*, **93**, 1–47.

Klemperer, P. (1995). "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade." *Review of Economic Studies*, **62**(4), 515–539.

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.

Lewis, T. and D. Sappington (1989). "Countervailing incentives in agency problems." *Journal of Economic Theory*, **49**, 294–313.

Maggi, G. and A. Rodriguez-Clare (1995). "On countervailing incentives." *Journal of Economic Theory*, **66**, 238–263.

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

Matthews, S. and J. Moore (1987). "Monopoly provision of quality and warranties: An exploration in the theory of multidimensional screening." *Econometrica*, **55**(2), 441–467.

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

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

Rochet, J.-C. and P. Choné (1998). "Ironing, sweeping, and multidimensional screening." *Econometrica*, **66**(4), 783–826.

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**, 197–229.

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

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.