Journal of Regulatory Economics; 4:115-138 (1992)

Ex Post vs. Ex Ante Pricing: Optional Calling Plans and Tapered Tariffs

KAREN B. CLAY
Stanford University
Department of Economics, Stanford, CA 94305

DAVID S. SIBLEY
University of Texas at Austin
Department of Economics, Austin, TX 78712

PADMANABHAN SRINAGESH
Bell Communications Research
LCC 3B234, 290 West Mount Pleasant Avenue, Livingston, NJ 07039

Abstract
We study optimal nonuniform pricing in a setting where a customer's demand at the start of a billing period contains a random variable whose realization becomes known by the end of the billing period. In this context, an optional calling plan is a tariff which the consumer must select based on his/her expectations about the random variable, whereas, under a tapered tariff, the consumer's choice of usage charge is made after he/she knows the realization of the random variable. We show that for low to moderate levels of uncertainty about the random variable entering the demand function, the optional calling plan approach to nonuniform pricing yields higher expected profit than does the tapered tariff approach, given risk-neutral consumers. We illustrate this finding with a case study and argue that it is consistent with the historical evolution of tariffs in the interexchange telecommunications market.

1. Introduction

Over the last fifteen years or so, a voluminous literature has grown up on optimal nonlinear prices for a monopolist. Among many others, contributions include Spence (1977), Roberts (1979), Goldman, Leland, and Sibley (1984), and Maskin and Riley (1984); more recent summaries and extensions of this literature are Brown and Sibley (1986) and Wilson (1990). This literature has done much to increase economists' understanding of nonlinear pricing structures, which have been common in regulated industries for decades. However, a review of actual nonlinear prices in telecommunications suggests that there is still something left to be explained. Specifically, nonlinear prices can take one of two distinct forms: optional calling plans (OCPS) and tapers. Optional calling plans are usually sets of two-part tariffs, each consisting of an entry fee and a usage charge, from which a consumer chooses and determines the effective price at the start of a billing period. Tapers are declining block tariffs, for which the effective price is determined at the end of the billing period. For example, under the OCP approach consumers could be given a choice between the two-part tariffs T_1 and T_2, where

T_1: Entry Fee = 50, usage charge = .9

T_2: Entry Fee = 100, usage charge = .1

Under the taper approach, consumers would face a single declining block tariff with

Entry fee	= 50	
Usage Charge	= .9,	Quantity $\le$ 62.5 units
	= .1,	Quantity > 62.5 units.

Nonlinear pricing theory has yet to provide an explanation of the difference between these two forms of nonlinear pricing.

Faulhaber and Panzar (1977) provide a starting point, when they note that any concave nonlinear outlay schedule (including tapers) can be represented as the lower envelope of a set of self-selecting two-part tariffs. Hence, if consumers have no random elements in their demand curves, either implementation would lead to the same result. If the concave outlay schedule is piecewise linear-a declining block schedule-it is equivalent to the set of two-part tariffs whose lower envelope it forms, as long as consumers' demand curves are non-stochastic. This suggests that, to explore the relationship between two part tariffs and tapers, it is necessary to introduce demand uncertainty at the level of the individual customer.

In recent empirical work, Train, Ben-Akiva, and Atherton (1987) report evidence consistent with this view: residential electricity customers rarely buy on the lower envelope of the set of tariff plans available to them. In telecommunications, as well, random shocks appear to be important. Kridel, Lehman, and Weisman (1990) report that most customers on flat-rate telephone service would have had lower bills if they had chosen Local Measured

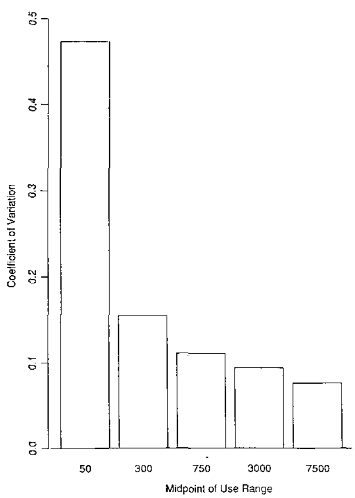

Figure 1

Service. Figure 1 is a histogram showing the coefficients of variation in monthly usage for business customers broken down by usage level into 5 classes: 0-100, 100-500, 500-1,000, 1,000-5,000, and 5,000-10,000 minutes per month. Note that the coefficient of variation decreases with volume of use. The standard deviation of monthly minutes of use exceeds the mean in a large proportion of the observations.

Two aspects of customer demand uncertainty seem relevant. First, with demand uncertainty, different tariff structures may offer different degrees of insurance against the risk of large bills. For example, with a uniform price, the customers' bill varies directly with random shocks in demand. A flat-rate tariff, of course, offers complete insurance against demand fluctuations.

Kridel, Lehman, and Weisman (1990) argue that customers exhibit substantial risk aversion when faced with bill uncertainty. For a data set based on one month's usage, they show that 55% of all customers who choose Flat Rate Service (FRS) would have achieved higher surplus if they had chosen Measured Service (MS) instead. Of those who chose MS, 10% had usage high enough to justify the FRS option. They argued that this observed bias toward FRS could be accounted for by risk aversion toward bill variation. Train (1989) makes a similar argument. A second aspect of demand uncertainty is that customers' tastes change over a billing period. At the start of a billing period, the customer faces a probability distribution of future demands for the duration of the billing period. As time goes by, the consumer begins to know what his demands will be for the remainder of the billing period.

We are inclined to doubt the risk aversion argument. For one thing, fluctuations in telephone bills are so small relative to income, that it is hard to imagine the marginal utility of income being anything but constant over the relevant range. Furthermore, it is not necessary to assume risk aversion to account for the empirical evidence. We base our argument on the observation that the choice of rate plan is made ex ante by the individual. Assume that the choice is based on expected consumer surplus. Observed usage is a realization of a random variable. Let $\bar{q}$ be the breakeven usage level at which MS and FRS are equally expensive. Suppose $\bar{q}$ is located in the left tail of the distribution of mean calls over consumers. There will be a greater mass of customers with means to the right of the breakeven point than there are on the left. As a consequence, the pool of customers who correctly chose FRS but are close enough to $\bar{q}$ to have actual (realized) usage to the left of $\bar{q}$ is greater than the pool of customers correctly on MS who can have a realized usage that is greater than $\bar{q}$. The pattern of actual usage is therefore likely to show the bias documented by Kridel, Lehman, and Weisman (1990). This insight is borne out by the following example.

Suppose customers have uncertain demand curves given by $Q_i = X_i \cdot exp(-\beta_i \cdot p)$, where X_i is distributed according to a gamma distribution with shape parameter 1 and mean μ_i. Assume that μ_i is distributed across the population according to a gamma distribution with shape parameter 1 and mean 120. (Mitchell (1978) obtained this mean in a study of California exchanges.) Next, consider the local service options in the Suburban Essex area of New Jersey. Flat Rate Service is available at \$7.95 per month. Two Message Rate services are available. We consider only the standard message rate service which costs \$6.39 per month, comes with 75 free message units per month and charges 6.5 cents per unit once the allowance is used. (For simplicity, we represent this as a Measured Service with a monthly fee of \$3.075 = (6.39 - 75 x .065), and a usage charge of 6.5 cents.) Based on expected consumer surplus, 46.14% of the customers should purchase MS. Suppose all customers make the rational ex ante choice. Next, we generate a usage for each customer by drawing an observation from the gamma distribution with mean μ_i. Finally, we compute the fraction of those people on MS (FRS) whose actual usage is to the right (left) of $\overline{q}$. For the 15 runs conducted, we found, on average, that 16.2% of the customers on MS had ex post usage that was too high. For consumers on FRS, 42.3% had usage that was too low. These numbers are close to those reported by Kridel, Lehman, and Weisman (1990). Further, our model explains why some people choose MS wrongly. The hypothesis of risk aversion cannot explain this choice. Thus, the bias uncovered by empirical studies does not necessarily imply that risk aversion is an essential feature of customer behavior. All that is needed is that choice of plan be made before usage for the month is known with certainty.

This second aspect of demand uncertainty brings out a difference between optional calling plans and tapers, one that we will focus on in this paper. Imagine that a customer's demand function depends on his underlying tastes, random shocks, and price. If he is presented with a set of two-part tariffs from which to choose at the start of a billing period, his choice will be based on his underlying type and his prior distribution about the future course of demand shocks. Under a tapered tariff, the customers' bill depends on the ex post realization of these demand shocks. Hence, when a firm chooses between offering a set of optional calling plans and a single tapered tariff, it is really deciding whether it is more profitable to offer a set of sales contracts to a population of informed consumers or of uninformed consumers.

In focusing on this aspect of uncertainty at the time of contracting, this paper parallels current developments in the closely-related principal agent literature. Lewis and Sappington (1990) and Crèmer (1990) have examined contracting problems in which the principal has a choice about whether or not to improve the quality of information available to the agent at the time of contracting.

Our theoretical results are as follows:

1. For small levels of uncertainty regarding customer demand, expected profit to the firm is higher with optional two-part tariffs than with a taper.
2. Total surplus and consumer surplus are lower on the profit-maximizing set of two-part tariffs than on the profit-maximizing taper.

Using some stylized parameter values, we apply the theory to the case of interstate access minutes, with results that suggest that the theoretical conclusion about profits appears likely to hold in practice, even for large levels of uncertainty. A review of interstate tariffs shows that during AT&T's days as a monopolist, its tariff structure consisted of optional calling plans.

The paper is organized as follows. In section 2, we develop the main concepts using simple examples. In section 3, we develop a formal model of optional calling plans and tapers and prove relevant theoretical results concerning tapers. In section 4, we compare optional calling plans and tapers with regard to expected profit, total surplus, and consumer surplus. In section 5, we apply the theory.

2. Basic Concepts

We first specify the timing of the model. At the start of a billing period, each consumer knows his underlying taste for the service in question, but not the realization of a random variable. The realization of the random variable becomes known after the start of the billing period and determines realized demand. The firm has the option of requiring consumers to select from a set of two-part tariffs at the beginning of the billing period or allowing them to select from a set of two part tariffs at the end of the period, when the realization of the random shock is known. Once the shock is known, there is no remaining demand uncertainty, and the lower envelope of the set of ex post two-part tariffs is equivalent to a single tapered tariff. Thus, we can analyze the differences between optional calling plans and a taper simply by analyzing the profit-maximizing set of two-part tariffs from which consumers select based on their priors about the random shock as compared to the profit-maximizing set of two-part tariffs from which consumers can choose when the realized value of the demand shock becomes known.

To begin, consider the case of a single customer with a demand curve $Q(P, \tilde{Z})$, where $\tilde{Z}$ is a random variable; $Q_Z > 0$. Suppose that $\tilde{Z}$ can take on either a low value Z_0 or a high value Z_1, with probabilities γ and $1 - \gamma$ respectively. Under the OCP approach, the profit-maximizing firm gives the consumer the take-it-or-leave-it option of a single two-part tariff, subject only to the individual rationality constraint that expected consumer surplus be nonnegative. Denoting the firm's marginal cost by mc, the profit-maximizing ex ante two-part tariff is

$$P^* = mc \tag{1}$$

$$E^* = \gamma\, CS(P^*, Z_0) + (1 - \gamma)\, CS(P^*, Z_1), \tag{2}$$

where P^* and E^* are the optimal usage charge and entry fee, respectively. Clearly, (P^*, E^*) extracts the entire expected social surplus as profit. To obtain the profit-maximizing taper, we first compute the expected profit-maximizing set of two part tariffs designed for the ex post population of consumers corresponding to the outcomes Z_0 and Z_1 of the random shock. That is, the ex post consumers are of two types, one with demand curve $Q(P, Z_0)$ and the other with demand curve $Q(P, Z_1)$.

The formal pricing problem, in this case, is to maximize expected profits subject to incentive compatibility between types Z_0 and Z_1 and also subject to individual rationality for Z_0:

$$\textit{Maximize}\ \gamma \cdot [(P_0 - mc)\, Q(P_0, Z_0) + E_0] + (1 - \gamma) \cdot [(P_1 - mc)\, Q(P_1, Z_1) + E_1]$$

s.t.

$$CS(P_1, E_1, Z_1) \geq CS(P_0, E_0, Z_1) \tag{IC}$$

$$CS(P_0, E_0, Z_0) \geq 0\,. \tag{IR}$$

The solution to this problem is well known:

$$P_0 > P_1 = mc \tag{3}$$

$$E_0 = \int_{P_0}^{\infty} Q(P, Z_0)\, dP \tag{4}$$

$$E_1 = E_0 + \int_{mc}^{P_0} Q(P, Z_1)\, dP\,. \tag{5}$$

Because $P_0 > mc$, expected total surplus is lower than with ex ante pricing. Hence, expected profit must also be lower. Denoting ex ante and ex post profits by Π^{EA} and Π^{EP}, we have

$$\Pi^{EA} - \Pi^{EP} = \gamma \cdot \left\{ \int_{mc}^{P_0} Q(P, Z_0)\, dP - (P_0 - mc)\, Q(P_0, Z_0) \right\}$$
$$+ (1-\gamma) \int_{P_0}^{\infty} (Q(P, Z_1) - Q(P_1, Z_0))\, dP > 0, \tag{6}$$

because of the noncrossing assumption $Q_Z > 0$. The reason for this result is that ex post pricing must incur incentive compatibility costs from the fact that $P_0 > P_1 = mc$, which are necessary to keep type Z_1 from selecting (P_0, E_0) instead of (P_1, E_1). In addition, expected profit is limited by the Individual Rationality (IR) constraint that $CS(P_0, E_0, Z_0) \geq 0$; under ex ante pricing there is no incentive compatibility constraint in this example, and the IR constraint is that expected consumer surplus over *both* ex post types must be nonnegative, which allows E^* to exceed $CS(P^*, E^*, Z_0)$, the consumer surplus of the lower ex post type.

To extend the analysis, consider a numerical example with two consumer types. Type A has a demand curve given by

$$\begin{aligned} Q_A(p) &= 9.5 - p \quad \text{if } p \leq 7 \\ &= 0, \qquad\quad\ \text{otherwise} \end{aligned} \tag{7}$$

Type B's demand curve is given by $Q_B(P) = \tilde{Z} - p$, where $\tilde{Z} = 10$ and $\tilde{Z} = 8$, each with probability 1/2, if $p \leq 7$ and $Q_B = 0$, otherwise. Thus, if $\tilde{Z} = 10$, type B is the largest consumer, whereas, if $\tilde{Z} = 8$, type A is larger. Type B's expected demand curve is $Q_{\bar{B}} = 9.0 - p$, lower than type A's demand curve.

Under OCP approach, this tariff planner deals with two customer types: A and $\bar{B}$, where $\bar{B}$ represents the expected demand curve of type B. Entry fees E_A and $E_{\bar{B}}$ and usage charges P_A and $P_{\bar{B}}$ are computed so as to maximize expected profit subject to incentive compatibility (IC) between types A and $\bar{B}$ and individual rationality (IR) for type $\bar{B}$. The marginal cost of the firm is 2.

Formally, the optimization problem is

$$\underset{\{E_A, E_{\bar{B}}, P_A, P_{\bar{B}}\}}{Maximize} \left[E_A + (P_A - 2) \cdot (9.5 - P_A) + E_{\bar{B}} + (P_{\bar{B}} - 2) \cdot (q - P_{\bar{B}})\right]$$

subject to:

$$\frac{1}{2} \cdot (9.5 - P_A)^2 - E_A \geq \frac{1}{2} \cdot (9.5 - P_{\bar{B}})^2 - E_{\bar{B}} \tag{IC}$$

$$\frac{1}{2}\cdot\left\{\frac{1}{2}\cdot(10-P_{\overline{B}})^2+\frac{1}{2}\cdot(8-P_{\overline{B}})^2\right\}-E_B-\frac{1}{2}\cdot\left\{\frac{1}{2}\cdot(10-7)^2+\frac{1}{2}\cdot(8-7)^2\right\}\geq 0. \quad \text{(IR)}$$

The solution to this problem is

$$E_A = 22.75 \qquad E_{\overline{B}} = 19.125 \qquad (8)$$

$$P_A = 2 \qquad P_{\overline{B}} = 2.5$$

and profit is equal to 45.125.

Under the taper approach, the first step is to choose two-part tariffs over the ex post population of consumers A, B_L, and B_H, where B_L corresponds to the demand curve $8-p$ and B_H to the demand curve $10-p$. Expected profit is maximized over the set of types to be served, with constraints requiring incentive compatibility between ex post types and individual rationality for the lowest ex post type in the optimal customer set. The largest ex post type is B_H, then type A, and the smallest is type B_L. The optimization problem is choose $m \in \{B_L, A, B_H\}$ so as to

$$\underset{\substack{\{E_m,\ldots,E_{B_H}\} \\ \{P_m,\ldots,P_{B_H}\}}}{Maximize} \sum_{j=m}^{j=B_H} \{E_j + (P_j - 2)\, Q_j(P_j)\} \cdot g_j, \qquad g_{BL} = g_{BH} = 1/2, \quad g_A = 1$$

s.t.

$$\frac{1}{2}(10-P_{BH})^2 - E_{BH} \geq \frac{1}{2}\cdot(10-P_A)^2 - E_A$$

$$\frac{1}{2}\cdot(9.5-P_A)^2 - E_A \geq \frac{1}{2}\cdot(9.5-P_{BL})^2 - E_{BL} \qquad (9)$$

$$\frac{1}{2}\cdot(8-P_{BL})^2 - E_{BL} - \frac{1}{2}\cdot(8-7)^2 \geq 0$$

if $m = B_L$ or s.t.

$$\frac{1}{2}\cdot(10-P_{BH})^2 - E_{BH} \geq \frac{1}{2}\cdot(10-P_A)^2 - E_A$$

$$\frac{1}{2}\cdot(9.5-P_A)^2 - E_A - \frac{1}{2}\cdot(9.5-7)^2 \geq 0, \qquad (10)$$

if $m = A$.

At the optimum, type B_L is priced out of the market, so $m = A$. The optimal tariffs are

$$E_{B_H} = 25.13 \qquad P_{B_H} = 2$$

$$E_A = 23.16 \qquad P_A = 2.25. \qquad (11)$$

Because $\tilde{Z}$ is realized ex post, this pair of two-part tariffs is equivalent to the following taper:

$$E = 23.16$$

$$P_1 = 2.25 \quad 0 \le Q \le 7.88 \tag{12}$$
$$P_2 = 2.0, \; Q > 7.88 \, .$$

This taper yields a profit of 37.53, less than profit under the OCP approach. The reason, again, is that the firm charges an entry fee $E_{\bar{B}}$ that exceeds type B's ex post willingness to pay in the event $\tilde{Z} = Z_0 = 8$.

This outcome changes, however, if $Z_0 = 5$, instead of 8. Performing the same two optimizations, the OCP achieves a profit of 32, whereas profit under taper is still 37.53. The advantage of the taper in this case is that it can exclude type B_L from the market and extract high profit from the large demand types B_H and A. The failure of the OCP approach in this example is that it cannot separate ex post types B_L and B_H. The IR constraint that the expected consumer surplus of type B be nonnegative must be observed, and with the lower value of Z_0, this expected value is pulled down for any two-part tariff. This constrains the entry fee $E_{\bar{B}}$ and also constrains E_A, because of the IC constraints. The taper, being the lower envelope of two-part tariffs tailored to the ex post types, can simply exclude B_L and extract surplus from the two higher types.

From these examples, we see the outline of a theory of OCPs vs. tapers. The advantage of the OCP approach is that is can force consumers to choose two-part tariffs that violate both IR and IC ex post. The disadvantage of OCPs is that, because they are constrained by expected values of consumer surplus, profitability is dragged down by low realizations of $\tilde{Z}$, which lower expected surplus. The taper, however, can sort between different outcomes of $\tilde{Z}$ and induce low ex post types to consume nothing. The net outcome can go either way. We now proceed to a more formal analysis of ex ante and ex post two-part tariffs.

3. Ex Ante Pricing and Ex Post Pricing

3.1 Tastes, Random Error, and Demand

In line with the standard literature on nonuniform pricing, we assume that there is a set of N consumer types, with taste parameters $T_1, T_2, \ldots, T_N$ with $T_i < T_{i+1}$, and that demand is increasing in taste; we denote the set of tastes by $I \equiv \{1, 2, \ldots, N\}$. In addition, realized demand depends on the outcome of the random error term $\tilde{Z}_i$, which is unknown at the beginning of a billing period, but becomes known after the start of the billing period.[1] In general, then, demand for type i is given by $Q_i = Q(T_i, \tilde{Z}_i, P)$, where P is the usage charge. We will make two extreme (but useful) assumptions about the form of the demand curve: (1) that T_i and $\tilde{Z}_i$ enter additively, as $T_i + \tilde{Z} \equiv \tilde{T}_i$, an "ex post" taste parameter; (2) that demand follows the form

$$Q_i = \tilde{T}_i \, q(p) + f(p), \tag{13}$$

where $q(p)$ and $f(p)$ are nonnegative functions such that for all $\tilde{Z}$ and i

$$\frac{\partial Q_i}{\partial p} = \tilde{T}_i \, q'(p) + f'(p) < 0 \, . \tag{14}$$

For simplicity, we assume that $\tilde{Z}$ can take on these realizations: $\tilde{Z} = Z_{0i} < 0$,

$\tilde{Z} = 0$, $\tilde{Z} = Z_{1i} > 0$, with the following probabilities:

$$\begin{aligned} Prob\{\tilde{Z} = Z_{0i} < 0\} &= \gamma_i \\ Prob\{\tilde{Z} = Z_{1i} > 0\} &= \gamma_i \\ Prob\{\tilde{Z} = 0\} &= 1 - 2\gamma_i . \end{aligned} \tag{15}$$

Intuitively, we can think of $\tilde{Z}$ as indicating some random event that influences demand over the course of a billing period. We will further assume that the $\tilde{T}_i$ are ordered in the same way as are the T_i, so that if $T_i < T_{i+1}$, then

$$T_i + Z_{0i} < T_i < T_i + Z_{1i} < T_{i+1} + Z_{0,\,i+1} < T_{i+1} < T_{i+1} + Z_{1,\,i+1} < \ldots < T_N < T_N + Z_{1N} . \tag{16}$$

To provide a convenient comparison below between OCPs and tapers, we will assume for simplicity that the mean of $\tilde{Z}$ is zero:

$$\gamma_i Z_{0i} + \gamma_i Z_{1i} = 0, \tag{17}$$

implying that $Z_{0i} = -Z_{1i}$.

Given these assumptions, expected demand and expected consumer surplus will be independent of Z_i and γ_i. We use this property repeatedly in what follows. In the ex ante pricing problem, the firm designs a set of two-part tariffs so as to maximize expected profit when consumer choice among these options takes place at the start of the billing period, before $\tilde{Z}_i$ is known. Because the expected value of $\tilde{T}_i$ is T_i, the maximization problem is written

$$\underset{\substack{\{P_m, P_{m+1},\ldots,P_N\} \\ \{E_m, E_{m+1},\ldots,E_N\} \\ m}}{Maximize} \sum_{i=m}^{N} [(P_i - mc)\, Q(T_i, P_i) + E_i] \tag{OCP}$$

subject to incentive compatibility and individual rationality constraints which hold in expectation:

$$\sum_{k=0}^{2} \gamma_{ki} \cdot \left\{ \int_{P_i}^{P_{i-1}} Q(T_i, \tilde{Z}, p)\, dp \right\} = \int_{P_i}^{P_{i-1}} \{T_i\, q(p) + f(p)\}\, dp \geq E_i - E_{i-1} \quad i = 1, 2, \ldots, N \tag{IC}$$

$$\sum_{k=0}^{2} \gamma_{km} \cdot \left\{ \int_{P_m}^{\infty} Q(T_m, \tilde{Z}, p)\, dp \right\} = \int_{P_m}^{\infty} (T_m\, q(p) + f(p)]\, dp - E_m \geq 0 \tag{IR}$$

where $k = 0, 1, 2$, $Z_{2i} = 0$, and $\gamma_{2i} = 1 - 2\gamma_i$. Because we assume risk-neutrality in $\tilde{Z}_i$ and because the mean of $\tilde{Z}_i$ is zero, neither $\tilde{Z}_i$ nor γ_i appear in the optimization problem.

In addition, we require that the set of tariffs not contain any that are dominated. We do this by requiring that P_i and P_{i+1} conform to self-selection constraints: $P_i \leq P_{i+1}$. The separating solution to the OCP problem is standard. Given the optimal value of m:

$$P_i = mc + (N-i)\cdot\left[\frac{Q_{i+1}(P_i) - Q_i(P_i)}{-Q_i'(P_i)}\right], i = m, m+1,...,N-1 \tag{18}$$

$$P_N = mc \tag{19}$$

$$E_m = \int_{P_m}^{\infty} (T_m \cdot q(p) + f(p))\, dp \tag{20}$$

$$E_i = E_{i-1} + \int_{P_i}^{P_{i-1}} (T_i\, q(p) + f(p))dp\,,\ i = m+1,...,N. \tag{21}$$

(To provide a convenient benchmark below for comparing OCPs and tapers, we will assume that the optimal set of OCPs is completely separating.)

In the case of the taper, we first compute the profit-maximizing set of two-part tariffs given the ex post taste distribution $\{\tilde{T}_i\}$; the taper is then obtained as the lower envelope of these two-part tariffs. We index ex post types by l, where $l_1 = T_1 - Z_{01}, l_2 = T_1$, $l_3 = T_1 + Z_{11}, l_4 = T_2 - Z_{02},..., l_{3N-1} = T_N, l_{3N} = T_N + Z_{1N}$. To arrive at the taper, we first compute the set of optimal two-part tariffs given the taste parameters $\{\tilde{T}_l\}$, $l = 1,...,3N$.

$$\underset{\substack{\{E_m,...,E_{3N}\} \\ \{P_m,...,P_{3N}\} \\ m}}{Maximize} \sum_{l=m}^{l=3N} [E_l + (P_l - mc)\, Q(P_l, \tilde{T}_l)]\, \gamma_l$$

subject to

$$\int_{P_l}^{P_{l-1}} Q(\tilde{T}_l, P)\, dp \geq E_l - E_{l-1}\,,\ l = m, \ldots, 3N \tag{IC}$$

$$\int_{P_l}^{\infty} Q(\tilde{T}_m, P)\, dp \geq E_m \tag{IR}$$

$$P_l \geq P_{l+1}\,. \tag{SS}$$

Note the important contrasts between this ex post optimization problem and the ex ante optimization used in the OCP approach. First, the objective function depends on the possible ex post realizations of $\tilde{Z}_i$ and on their probabilities, γ_{0i} and γ_{1i}. Also, the IC and IR constraints hold between the various possible realizations of $\tilde{T}_l$, and not simply between the expected values T_i.

We start with the separating solution, as a prelude to considering pooling solutions. Denote by $\tilde{P}_i$ the usage charge selected by customer type i in the event that $\tilde{Z} = 0$ and by

$\tilde{P}_{iL}$ and $\tilde{P}_{iH}$ the usage charges on the two-part tariffs selected by $T_i + Z_{0i}$ and $T_i + Z_{1i}$, respectively.

$$\tilde{P}_{iL} = mc + \left(\frac{N+1-i-\gamma_i}{\gamma_i}\right) \cdot \left(\frac{Z_{0i}\, q(\tilde{P}_{IL})}{(T_i + Z_{0i})\, q'(\tilde{P}_{iL}) + f'(\tilde{P}_{iL})}\right) \tag{22}$$

$$\tilde{P}_i = mc + \left(\frac{N-i+\gamma_i}{1-2\gamma_i}\right) \cdot \left(\frac{Z_{1i}\, q(\tilde{P}_i)}{-T_i\, q'(\tilde{P}_i) - f'(\tilde{P}_i)}\right) \tag{23}$$

$$\tilde{P}_{iH} = mc + \left(\frac{N-i}{\gamma_i}\right) \cdot \left(\frac{T_{i+1} - T_i - (Z_{0i} - Z_{1i})}{-[(T_i + Z_{1i})\, q'(\tilde{P}_{iH}) + f'(\tilde{P}_{iH})]}\right) \cdot q(\tilde{P}_{IH}). \tag{24}$$

Scrutinizing the solution, we see that for a separating solution both $\tilde{Z}_i$ and γ must exceed a certain threshold level. For Z_{0i} and Z_{1i} small enough, $\tilde{P}_{iL}$ and $\tilde{P}_i$ are both less than $\tilde{P}_{iH}$, which violates self-selection, requiring pooling. Similarly, as $\gamma_i \to 0$, $\tilde{P}_{iH}$ tends to infinity, again violating self-selection. Therefore, $Z_{0i} << 0$, $Z_{1i} >> 0$, and $\gamma_i >> 0$ are needed for it to be optimal to separate the types *iL*, *i*, and *iH*. Intuitively, this comes about because, to maintain separation between these adjacent types, incentive compatibility costs must be incurred. For example, if type *iH* has selected a tariff with a low usage charge and a high entry fee, then in order to keep type *iH* from selecting the two-part tariff selected by type *i*, which would have a lower entry fee, the usage charge $\tilde{P}_i$ must be set enough higher than $\tilde{P}_{iH}$ to make $(\tilde{E}_i, \tilde{P}_i)$ unattractive to type *iH*. This, however, limits the surplus that can be extracted from type *i*. If Z_{1i} were small, so that demand curves of *iH* and *i* were close together, $\tilde{P}_i$ would have to be set very far indeed above $\tilde{P}_{iH}$ to induce separation, with a correspondingly larger efficiency loss and reduction in the profit that can be earned from type *i*. Hence, with a small $\tilde{Z}_i$, separation is not profitable. A similar intuition applies when γ_i is small.

This suggests that we examine pooling when γ and/or Z are small. We can do so using a graphical approach. In figure 2, we depict a separating solution at some $\gamma_i = \gamma^0$, with $\tilde{P}_{i+1,L} < \tilde{P}_{iH} < \tilde{P}_i < \tilde{P}_{iL} < \tilde{P}_{i-1,H} < \tilde{P}_{i-1} < \tilde{P}_{i-1,L} < \tilde{P}_{i-2,H} < \tilde{P}_{i-2}$. Now reduce γ_i towards zero, and, based on (22)-(24), the following pattern emerges:

$$\begin{cases} \tilde{P}_{i+1,L} & \text{rises} \\ \tilde{P}_{iH} & \text{rises} \\ \tilde{P}_i & \text{falls} \end{cases} \tag{25}$$

$$\begin{cases} \tilde{P}_{iL} & \text{rises} \\ \tilde{P}_{i-1,H} & \text{rises} \\ \tilde{P}_{i-1} & \text{falls} \end{cases} \tag{26}$$

$$\begin{cases} \tilde{P}_{i-1,L} & \text{rises} \\ \tilde{P}_{i-2,H} & \text{rises} \\ \tilde{P}_{i-2} & \text{falls.} \end{cases} \tag{27}$$

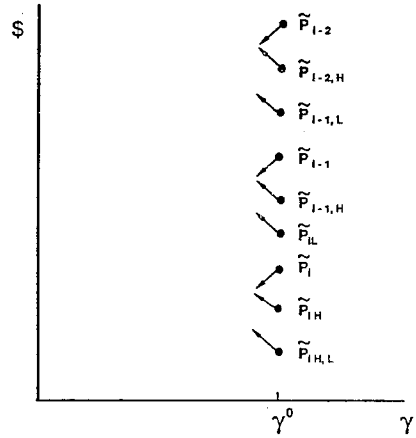

Figure 2

This process is depicted in figure 2. For the self-selection constraint $\tilde{P}_{i+1,L} \le \tilde{P}_{iH} \le \tilde{P}_i$ to be maintained as $\gamma_i \to 0$, pooling of types i, iH, and $i+1, L$ is necessary for some sufficiently small γ_i. Similarly, for some suitably small γ_i types iL, $i-1$, and $i-1, H$ become pooled, and so on. Denote the lowest index in market by m; depending on whether type mL is priced out of the market or not, there are either $N-m+1$ or $N-m+2$ optional two-part tariffs offered when γ is small enough so that all ex post types are pooled in this way. The expression for the optimal usage charge for the two-part tariff selected by types i, iH, and $i+1, L$ is denoted

$$\tilde{P}(i; iH; i+1, L) = mc + \frac{q(P_i)\cdot\{Z_{i+1}\gamma_{i+1} - Z_i\gamma_i + (N-i)\cdot(T_i - T_{i+1})\}}{q'(\tilde{P}_i)\cdot\{T_i\cdot(1-\gamma_i)+T_{i+1}\gamma_{i+1}+Z_i\gamma_i - Z_{i+1}\gamma_{i+1}\}+f'(\tilde{P}_i)\cdot(1-\gamma_i+\gamma_{i+1})}. \quad (28)$$

This resembles the marginal price selected by type i under the ex ante OCP approach, except that the denominator of the right hand term is larger.

In figure 3, we depict pooling behavior as Z goes to zero. $\tilde{P}_{iH}$ rises, whereas $\tilde{P}_i$ and $\tilde{P}_{iL}$ both fall. The only way to observe the self-selection constraint is for types iL, i, and iH to be pooled on the same two-part tariff. The same conclusion holds for all consumer types, except possibly for type m. The optimal ex post tariffs now are given by

$$\tilde{P}(iL, i, iH) = mc + \frac{q(\tilde{P}_i)\,[(N-i)\cdot(T_{i+1} - T_i + Z_i - Z_{i+1}) + Z_i]}{-(T_i\, q'(\tilde{P}(\cdot)) + f'(\tilde{P}(\cdot)))}. \quad (29)$$

Next, we examine pooling as $\gamma_i \to 1/2$ and $\tilde{Z}_i \to (T_{i+1} - T_i)/2$. These turn out to be the same thing because both lead to the eradication of type i, leaving only iL and iH. In figure 4, beginning with a separating solution at γ^0, as $\gamma \to 1/2$, $\tilde{P}_{i+1}$ rises and $\tilde{P}_{i+1,L}$ and $\tilde{P}_{iH}$ both fall. For self-selection, types iH, $i+1, L$, and $i+1$ must be pooled; this common usage charge is

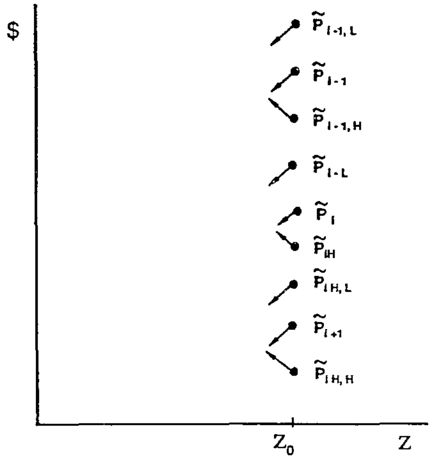

Figure 3

$$\tilde{P}(iH;i+1,L;i+1)=mc+\frac{\{(N-i)\cdot(T_i+Z_i-T_{i+1}-Z_{i+1})+T_i+Z_i-T_{i+1}\}\,q(\tilde{P}_i)}{q'(\tilde{P}_i)\cdot[(T_i+Z_i)\cdot\gamma_i+T_{i+1}\cdot(1-\gamma_{i+1})-Z_{i+1}\gamma_{i+1}]+f'(\tilde{P}_i)\cdot(\gamma_i-\gamma_{i+1})}. \tag{30}$$

Despite the fact that different patterns of pooling take place as $\gamma_i \rightarrow 0$ and $\tilde{Z}_i \rightarrow 0$, it should be clear that in the limit they both converge to the optimal OCP two-part tariff given in equations (18)-(21) above. However, when either $\gamma_i \rightarrow ½$ or $\tilde{Z}_i \rightarrow (T_{i+1} - T_i)/2$, type i is, in effect, vanishing, and the limiting usage charge is not equivalent to the OCP usage charge.

4. Comparisons Between OCPs and Tapers

Having derived the pricing rules and pooling patterns for all possible values of γ_i and $\tilde{Z}_i$, we can now compare expected profitability, consumer surplus, and total surplus to the firm

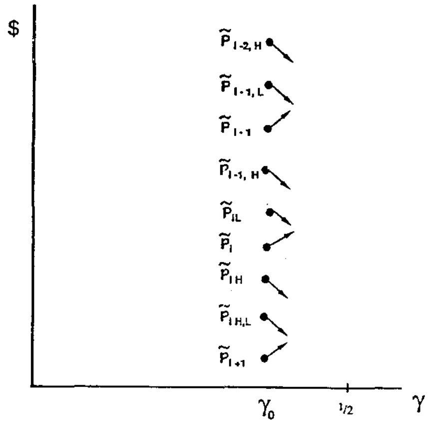

Figure 4

under the profit-maximizing ex ante set of two-part tariffs (i.e., the OCP approach) with those quantities, given ex post choice among two-part tariffs, the lower envelope of which is a single tapered tariff. We are particularly concerned with small values of γ_i and $\tilde{Z}_i$, because we know that the two approaches are equivalent and yield the same results when $\gamma_i \cdot \tilde{Z}_i = 0$.

4.1 Expected Profit

Expected profit under the OCP approach, Π^{OCP}, is (trivially) unaffected as γ_i and $\tilde{Z}_i$ vary because of our assumptions that $\tilde{Z}_i$ enters linearly in the demand and that the expected value of $\tilde{Z}_i$ is zero. The behavior of expected profit under the taper approach, Π^T, is fairly complicated. The analysis must proceed in two stages. First, given m^T, the marginal consumer type, what is the behavior of Π^T as $\gamma_i \to 0$ and $\tilde{Z}_i \to 0$? Second, how does m^T change and what is the effect of this change on Π^T?

4.1.1. Behavior of Π^T for Fixed m^T. We first consider the case $\gamma_i \to 0$. As we saw above, for γ_i sufficiently small, types i, iH, and $i+1, L$ are pooled onto the same two-part tariff. For subsequent notational convenience, we denote $\tilde{P}(i; iH; i+1, L)$ by $\tilde{P}_{i\gamma}$, the marginal price when γ_i is small and types i, iH, and $i+1, L$ are pooled onto the same two part tariff. The entry fee $\tilde{E}(i; iH; i+1, L)$ is similarly denoted $\tilde{E}_{i\gamma}$. Also, we write

$$\Pi^T_{i\gamma}(\tilde{P}_{i\gamma}, \tilde{E}_{i\gamma}) \equiv (\tilde{P}_{i\gamma} - mc) \cdot Q_i(\tilde{P}_{i\gamma}) + \tilde{E}_{i\gamma}, \tag{31}$$

as the profit earned on sales to type i when γ is small enough so that i, iH, and $i+1, L$ have been pooled onto the same tariff, for all i. Finally, we can write expected total profit under the taper as

$$\Pi^T = \sum_{i=m}^{N-1} [\gamma_i \, \Pi^T_{iH\gamma} + (1 - 2\gamma_i) \, \Pi^T_{i\gamma} + \gamma_{i+1} \, \Pi^T_{i+1, L\gamma}]$$
$$+ \Pi^T_{N\gamma} \cdot (1 - 2\gamma_N) + \gamma_N \, \Pi^T_{NH\gamma} \tag{32}$$

where types N and NH are pooled. Holding m^T fixed,

$$\frac{\partial \Pi^T}{\partial \gamma_i} = \Pi^T_{iH\gamma}(\tilde{P}_{i\gamma}, \tilde{E}_{i\gamma}) - 2\Pi^T_{i\gamma}(\tilde{P}_{i\gamma}, \tilde{E}_{i\gamma}) + \Pi^T_{iL}(\tilde{P}_{i-1,\gamma}, \tilde{E}_{i-1,\gamma}) \tag{33}$$
$$= (\tilde{P}_{i\gamma} - mc) \cdot [Q_{iH}(\tilde{P}_{i\gamma}) - 2\, Q_i(\tilde{P}_{i\gamma})] - \tilde{E}_{i\gamma} + \tilde{E}_{i-1,\gamma} + (\tilde{P}_{i-1,\gamma} - mc) \cdot Q_{iL}(\tilde{P}_{i-1,\gamma})$$
$$= (T_i - Z_i) \cdot [- (\tilde{P}_{i\gamma} - mc) \cdot q(\tilde{P}_{i\gamma}) + (\tilde{P}_{i-1,\gamma} - mc) \cdot q(\tilde{P}_{i-1,\gamma})] - \tilde{E}_{i\gamma} + \tilde{E}_{i-1,\gamma}$$
$$\leq (T_i - Z_i) \cdot q(\tilde{P}_{i-1,\gamma}) \cdot (\tilde{P}_{i-1,\gamma} - \tilde{P}_{i\gamma}) - T_i \cdot \int_{\tilde{P}_{i\gamma}}^{\tilde{P}_{i-1,\gamma}} q(p)\, dp < 0 .$$

To calculate the effect of $\tilde{Z}_i \to 0$ on Π^T, we require a similar series of definitions for two-part tariffs when $\tilde{Z}_i$ is small enough so that all types (iL, i, iH) are pooled:

$$\begin{aligned} \tilde{P}_{iZ} &= \tilde{P}(iL; i; iH) \\ \tilde{E}_{iZ} &= \tilde{E}(iL; i; iH) \\ \Pi^T_{iZ} &\equiv (\tilde{P}_{iZ} - mc) \cdot Q_i(\tilde{P}_{iZ}) + E_{iZ}\,. \end{aligned} \tag{34}$$

Then, we write total profit, Π^T, as follows

$$\Pi^T = \sum_{i=m^T}^{N} \left[\gamma_i \, \Pi_{iL}(\tilde{P}_{iZ}, \tilde{E}_{iZ}) + (1 - 2\gamma_i)\, \Pi_i(\tilde{P}_{iZ}, \tilde{E}_{iZ}) + \gamma_i \, \Pi_{iM}(\tilde{P}_{iZ}, \tilde{E}_{iZ})\right]. \tag{35}$$

To calculate the effect of $\tilde{Z}_i$ on profits, let $T_{iL} \equiv T_i - Z_i$ and $T_{iH} \equiv T_i + Z_i$

$$\frac{\partial \Pi^T}{\partial \tilde{Z}_i} = -\gamma_i \cdot (\tilde{P}_{iZ} - mc)\frac{\partial Q_{iL}}{\partial \tilde{T}_{iL}} + \gamma_i \cdot (\tilde{P}_{iZ} - mc) \cdot \frac{\partial Q_{iH}}{\partial \tilde{T}_{iH}} - (N + 1 - i) \cdot \int_{\tilde{P}_{iZ}}^{\tilde{P}_{i-1,Z}} \frac{\partial Q_{iL}}{\partial \tilde{T}_{iL}}\, dp \tag{36}$$

$$= \gamma_i \cdot [q(\tilde{P}_{iZ}) - q(\tilde{P}_{iZ})] \cdot (\tilde{P}_{iZ} - mc) - (N + 1 - i) \cdot \int_{\tilde{P}_{iZ}}^{\tilde{P}_{i-1,Z}} \frac{\partial Q_{iL}}{\partial \tilde{T}_{iL}}$$

$$= -(N + 1 - i) \int_{P_{iZ}}^{\tilde{P}_{i-1,Z}} q(p)\, dp < 0\,.$$

Thus, where γ_i, Z_i (or both) are "small" enough, an increase in uncertainty, however measured, leads to a reduction in expected profit. As suggested in section 2 above, this effect comes from the fact that an increase in the likelihood of low ex post types raises the self-selection costs of extracting surplus from high ex post **types.**

4.1.2. Changes in m^T. Thus far, we have examined the behavior of the optimal level of expected profit, Π^T, as γ_i and $\tilde{Z}_i$ change, holding m^T fixed. We must now take into account changes in the marginal consumer type as well. Define $\Pi^T(\gamma_i, \tilde{Z}_i; \mu)$ by

$$\Pi^T(\gamma_i, \tilde{Z}_i; \mu) = \max_{\substack{\{E_\mu,\ldots,E_{nH}\} \\ \{P_\mu,\ldots,P_{NH}\}}} \sum_{i=\mu}^{NH} [\Pi_{iL}\,\gamma_i + \Pi_i \cdot (1 - 2\gamma_i) + \Pi_{iH} \cdot \gamma_i]$$

s.t.

$$CS_i(P_i, E_i) \geq CS_i(P_{iL}, E_{iL}) \tag{IC}$$

$$CS_{iL}(P_{iL}, E_{iL}) \geq CS_{iL}(P_{i-1,H}, E_{i-1,H})$$

$$\vdots$$

$$CS_{\mu}(P_{\mu}, E_{\mu}) \geq 0 \qquad \text{(IR)}$$

$$P_m \geq \ldots \geq P_{iL} \geq P_i \geq P_{iH} \geq \ldots P_{NH} \qquad \text{(SS)}$$

Thus, $\Pi^T(\gamma_i, \tilde{Z}_i; \mu)$ is maximum expected profit given that the lowest participating customer type is type μ. From above, $\dfrac{\partial \Pi^T(\gamma_i, \tilde{Z}_i; \mu)}{\partial \gamma_i} < 0$ and $\dfrac{\partial \Pi^T(\gamma_i, \tilde{Z}_i; \mu)}{\partial \tilde{Z}_i} < 0$.

Next, define m^T by

$$\Pi^T(\gamma, Z; m^T) = \max_{\mu} \Pi^T(\gamma, Z; \mu), \tag{37}$$

which defines $m^T = m^T(\gamma, Z)$ as a function of γ and Z. Therefore, the reduced-form profit function is written

$$\Pi^T(\gamma, Z; m^T(\gamma, Z)) \equiv \Pi^*(\gamma, Z) \geq \max_{\mu \neq m^T} \Pi^T(\gamma, Z, \mu). \tag{38}$$

Furthermore, because of the definition of m^T as the marginal consumer type that maximizes expected profit, $\Pi^*(\gamma_i, \tilde{Z}_i)$ is a continuous function in γ_i and $\tilde{Z}_i$. The derivatives of m^T are zero almost everywhere; hence, $\partial\Pi^*/\partial\tilde{Z}_i < 0$ and $\partial\Pi^*/\partial\gamma_i < 0$ almost everywhere, and Π^T is a decreasing function of γ_i and $\tilde{Z}_i$ for $(\gamma_i, \tilde{Z}_i)$ close enough to (0, 0) so that the pooling solutions hold.

4.1.3. Profit Comparison Between OCPs and Tapers. The foregoing discussion has established three points:

1. When $\gamma_i \cdot \tilde{Z}_i = 0$. $\Pi^{OCP} = \Pi^*(\gamma_i, \tilde{Z}_i)$
2. Π^{OCP} is independent of γ_i and $\tilde{Z}_i$.
3. Π^* is decreasing in both γ_i and $\tilde{Z}_i$ for sufficiently small values of γ_i and $\tilde{Z}_i$.

Thus, there exists a neighborhood of $(\gamma_i, \tilde{Z}_i)$ around (0, 0) within which $\Pi^*(\gamma, Z)$ is less than Π^{OCP}. We state this formally.

Theorem 1. There exists a neighborhood $(\gamma_i, \tilde{Z}_i)$ around (0, 0) within which the optimal set of OCPs generate higher expected profit than does the optimal taper.

This result is illustrated in figure 5. As the example in Section 2 above illustrates, for sufficiently large uncertainty, the result of Theorem 1 can be reversed. The practical effect of Theorem 1 depends on how large the neighborhood around (0, 0) must be for the result to hold. We can gain some insight into this question by doing an extended numerical example.

Numerical Example. We can adapt a model used by Heyman, Lazorchak, Sibley, and Taylor (HLST) to study nonlinear pricing of interstate access minutes for a 1986 sample of business customers of New York Telephone Company. See HLST (1987). We assume isoelastic demand functions of the form

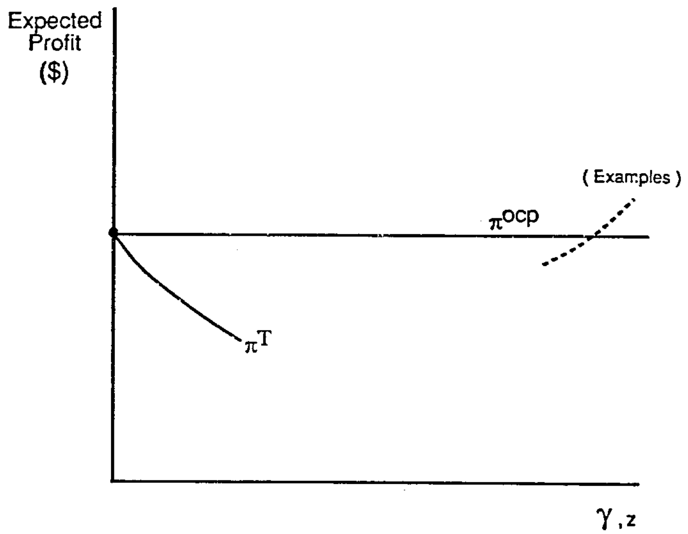

Figure 5

$$Q(\tilde{T}_i, P_i) = \tilde{T}_i P_i^{-e_i}, \ i = 1, ..., N. \tag{39}$$

We calibrate the model with data on the usage distribution presented in table 1.

Table 1. Distribution of NY TEL MTS Usage

Usage Band	Percentage of Accounts	Monthly MTS Minutes of Use (mou)
0-60	74.03	14.55
61-1,000	25.47	160.21
1,001-2,000	0.26	1,364.46
2,001-7,000	0.17	3,547.77
7,001-20,000	0.05	11,026.07
20,000 +	0.02	67,425.60

Souce: NYTel Co., Tariff FCC No. 41, Transmittal No. 775, June 26, 1986.

At the time this usage was generated, the usage-sensitive access price was 7.56¢/min. Table 1 suggests a model with six consumer types with underlying taste parameters $T_1, T_2, ..., T_6$. By assuming values for the price elasticities $e_1, e_2, ..., e_6$, we can solve for the set $\{T_i\}$ implied by the usage distribution in table 1; these elasticities are contained in table 1, too. The marginal cost of an access minute assumed by HLST was 1.3¢, which we use here, as well. We specify the random component of demand as a proportion of T_i. Thus, $\tilde{T}_i = \tilde{Z}_i T_i$ where the expected value of $\tilde{Z}_i$ is unity. Demand is

$$Q(T_i, \tilde{Z}, P) = \tilde{Z}_i T_i P^{-e_i}, \ i = 1, ..., 6. \tag{40}$$

We analyze pricing under a number of scenarios concerning γ and $\tilde{Z}$. (For simplicity, let $Z_{0i} = Z_{0,i+1}$ and $Z_{1i} = Z_{i,i+1}$.)

We follow HLST in maximizing profit subject to the constraint that the initial price 7.56¢/min be available to consumers as an option, in addition to the two-parts tariff which maximize expected profit. Under the OCP approach, we do this given the six expected consumer types $T_1, T_2,...,T_6$. Specifically,

$$\underset{\substack{\{P_1, P_2,...,P_6\} \\ \{E_1, E_2,...,E_6\}}}{Maximize} \sum_{i=1}^{6} \Pi_i \, g_i \,,$$

where g_i is the proportion of all customers who are of type i, subject to

$$CS_i(P_i, E_i) \geq CS_i(P_{i-1}, E_{i-1}), \quad i = 2,...,6 \qquad \text{(IC)}$$

$$CS_i(P_i, E_i) \geq CS_1(P_0) \qquad \text{(IR)}$$

$$P_i \geq P_{i+1}. \qquad \text{(SS)}$$

Under the taper approach, we first design two-part tariffs for the 18 ex post types:

$$Z_0\, T_1, T_1, Z_1\, T_1, Z_0\, T_2, T_2, Z_1\, T_2,...,Z_1\, T_6\,.$$

We then take the lower envelope of the set to arrive at the taper. When $\tilde{Z} = Z_i$ for ex ante type i, we denote the resulting ex post type by $iH(iL)$. The maximization problem is

$$\underset{\substack{\{P_{iL}, P_1, P_{iH},...,P_{6H}\} \\ \{E_{iL}, E_i, E_{iH},...,E_{6H}\}}}{Maximize} \sum_{j=1L}^{j=6H} \Pi_j(P_j, E_j)\, g_j$$

s.t.

$$CS_j(P_j, E_j) \geq CS_j(P_{j-1}, E_{j-1})$$

$$CS_{1L}(P_{1L}, E_{1L}) \geq CS_{1L}(P_0)$$

$$P_j \geq P_{j+1}.$$

We should note that because the tariffs must Pareto dominate the initial price of P_0, all size types will participate in the market under the optimal set of two-part tariffs, both with OCPs and the taper.

The optimal solution under the OCP approach is to pool types 1-5 on the initial price of 7.56¢ per minute and to offer a single optional two part tariff which will be selected only by type 6. The entry fee is $8818 per month and the usage charge equal to 1.3¢, the firm's marginal cost. Expected profit rises by 18.36 percent compared to the initial situation, in which the single price 7.56¢ was offered.

The taper approach requires a specification of γ and $\tilde{Z}$. We assume $Z_0 = .95$, $Z_1 = 1.05$, and $\gamma = 0.5$, so that there is a five percent chance of a five percent variation in demand. The

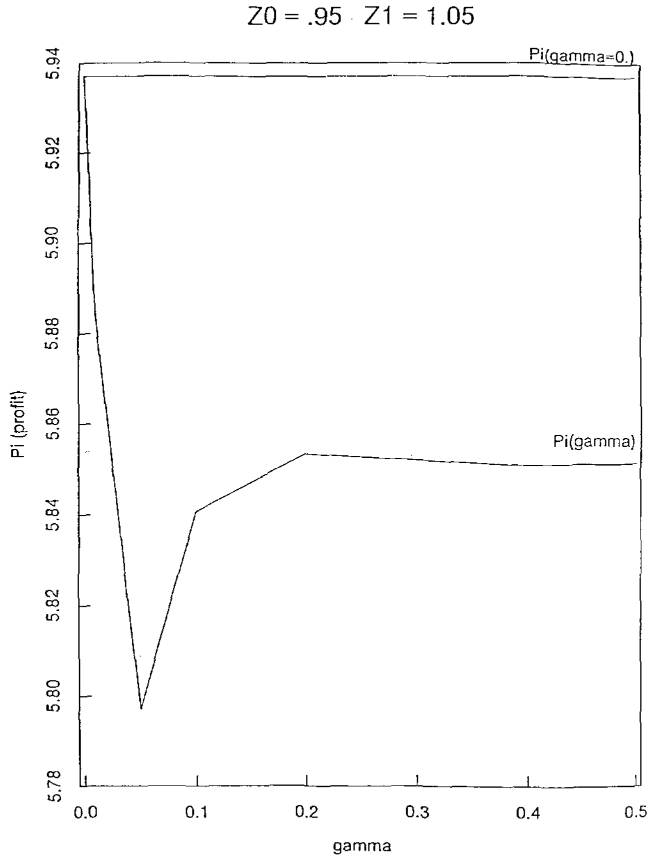

Figure 6

set of two-part tariffs resulting yields, as a lower envelope, the taper:

$$\begin{aligned} P &= 7.56¢,\ 0 \le Q \le 88{,}000 \text{ minutes per month} \\ &= 3.8¢,\ 88{,}000 < Q \le 171{,}500 \text{ minutes per month} \\ &= 2.1¢,\ 171{,}500 < Q \le 308{,}125 \text{ minutes per month} \\ &= 1.3¢,\ Q > 308{,}125\,. \end{aligned}$$

Expected profit rises by 15.77% as compared to the original situation, less than the profit increase brought about by OCPs. In figures 6 and 7, we depict the relative performance of OCPs and tapers for ($Z_0 = .95, Z_1 = 1.05$) and for ($Z_0 = .8, Z_1 = 1.2$) as a function of γ. In both figures, the taper is less profitable than the OCP. This suggests that the result of Theorem 1 holds for large enough neighborhoods of (0, 0) to be of some interest in a practical sense.

Interstate Telecommunications Pricing. The result of Theorem 1 seems consistent with

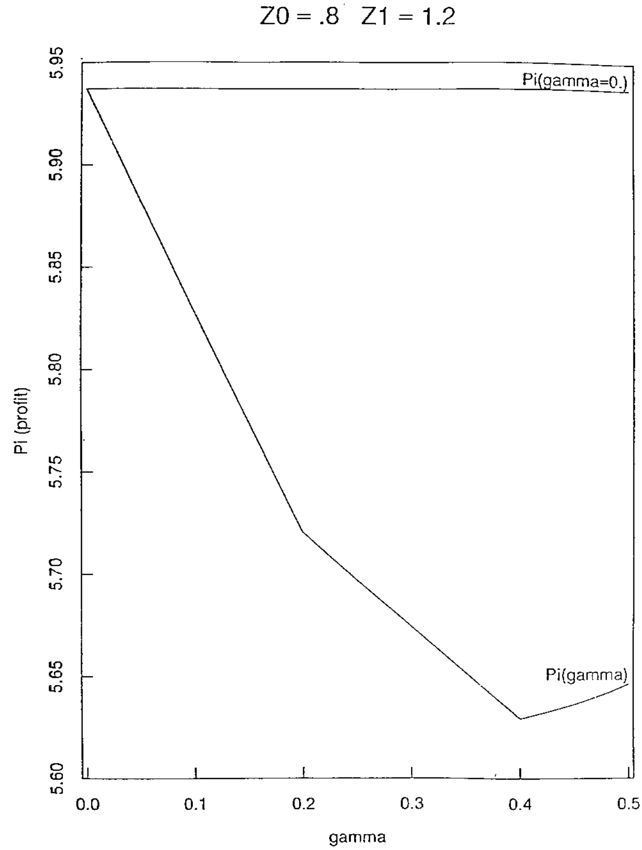

Figure 7

tariff history in telecommunications for ordinary voice-grade services. In the 1950s, AT&T was a virtual monopoly in a rapidly growing market. The vast majority of customers used ordinary long distance Message Toll Service (MTS), which required no presubscription and was essentially a uniform price which varied by length of haul and by time of day. The only alternatives to MTS were private networks and leasing private lines from AT&T. AT&T's private line tariff was a flat rate per line, with no usage charges, and was intended for business users with large volumes of point-to-point traffic. The choice between MTS and private line was simply a choice between OCPs for point-to-point tariffs. WATS was introduced in 1962 to offer customers volume discounts for a service very like MTS. The customer's WATS line terminated at the local switch, allowing the customer to complete calls in a broad area, rather than be limited to point-to-point traffic. WATS, too, required presubscription and had a zero usage charge.

Apart from MTS, AT&T's other voice-grade services were OCPs aimed at large business users. Large business users tend to have very stable demands, because their traffic volumes represent the summation of usage levels over many individual callers. This claim is borne out by figure 1, which suggests that the coefficient of variation of demand for a sample of business users is a decreasing function of mean demand. This suggests that the premise of Theorem 1 that demand uncertainty be "small" is met. Given that, AT&T's interstate tariff structure up until the late 1970s is consistent with Theorem 1: OCPs were chosen instead of tapers. See Brock (1981) and Faulhaber (1987) for more details on these developments.

4.2. Total Surplus and Consumer Surplus

In this section, we compare the levels of total surplus (TS) and consumer surplus (CS) achieved by the profit-maximizing OCPs and taper. There is a technical complication here that did not exist in the analysis of expected profits under the two regimes: TS^T and CS^T, total and consumer surplus under the taper, will be discontinuous functions of γ and z, even in the region around (0, 0) where pooling takes place. This has to do with choice of m^T, the marginal consumer type on the taper. Because we are only considering OCPs and tapers which maximize expected profit, m^T is defined by

$$\Pi^T(\gamma_i, \tilde{Z}_i; m^T) \geq \Pi^T(\gamma_i, \tilde{Z}_i; k), \quad \underset{k \in I.}{k \neq m^T} \tag{41}$$

Thus, even though $m^T(\gamma_i, \tilde{Z}_i)$ is integer-valued, $\Pi^*(\gamma_i, \tilde{Z}_i)$ is continuous, as argued above. However, when m^T changes with γ_i and/or $\tilde{Z}_i$, TS^T and CS^T will change discontinuously. Hence, we can only examine the effects of γ_i and $\tilde{Z}_i$ in a region close enough to (0, 0) that m^T is unaffected. (This requires the additional assumption that (41) holds with strict inequality.)

With this in mind, we can state and prove the following theorem.

Theorem 2. Assume that $\Pi^T(\gamma_i, \tilde{Z}_i; m^T) > \max_{\substack{k \neq m^T \\ k \in I}} \Pi^T(\gamma_i, \tilde{Z}_i; k)$. Then, for a suitably small region around the origin (0, 0) (a) for $\gamma_i > 0$, TS^T is smaller under the taper; (b) given $\gamma_i > 0$, in this region, increases in $\tilde{Z}_i$ increase CS^T but leave TS^T unchanged.

Proof of (a):

As shown above, if γ_i is small enough, types i, iH, and $i+1$, L are pooled together. And from (28), $\tilde{P}_i(i; iH: i{+}1, L)$ does not depend on γ or Z. Thus,

$$TS^T = \sum_{i=m^T}^{N-1} \Bigg[\int_{\tilde{P}_i(\cdot)}^{\infty} \{q(p)\cdot(2T_i{+}Z_1{+}T_{i+1}{-}Z_0){+}3f(p)\}\, dp$$

$$- mc\cdot[q(\tilde{P}_i(\cdot))\cdot(2T_i{+}Z_1{+}T_{i+1}{-}Z_0){+}3f(\tilde{P}_i(\cdot))\,]\Bigg]$$

$$+ \int_{\tilde{P}_N(\cdot)}^{\infty} [q(p)\cdot(2T_N + Z_1) + f(p)\,]\, dp - mc \cdot [\,(2T_N + Z_1)\, q\,(\tilde{P}_N(\cdot)) + 2f\,(\tilde{P}_N(\cdot))\,] \quad \text{and}$$

$$\frac{\partial TS^T}{\partial \gamma} = (\tilde{P}_i(i; iH; i+1, L) - mc)\cdot(Q_{iH}(\tilde{P}_i(\cdot)) - 2Q_i\,\tilde{P}_i(\cdot))$$

$$+ (\tilde{P}_{i-1}(i-1; i-1, H; iL) - mc) \cdot Q_{iL}(\tilde{P}_{i-1}(\cdot))$$

$$+ \int_{\tilde{P}_{i-1}(\cdot)}^{\infty} Q_{iL}\, dp - 2\int_{\tilde{P}_i(\cdot)}^{\infty} Q_i\, dp + \int_{\tilde{P}_i(\cdot)}^{\infty} Q_{iH}\, dp$$

$$\leq (T_i - Z_i)\, q(\tilde{P}_{i-1}(\cdot)) \cdot (\tilde{P}_{i-1}(\cdot) - \tilde{P}_i(\cdot)) + \int_{\tilde{P}_i(\cdot)}^{\tilde{P}_{i-1}(\cdot)} q(p)\ dp < 0\,.$$

Because $\frac{\partial \Pi^T}{\partial \gamma} < 0$ from Theorem 1, the effect of γ on consumer surplus is ambiguous.

Proof of (b):

As $Z_i \to 0$, types iL, i, and iH pool on $\tilde{P}_i(iL, i, iH)$. Therefore,

$$TS^T = \sum_{i=m^T}^{N} \left[\int_{\tilde{P}_i(\cdot)}^{\infty} [q(p) \cdot \{3T_i - Z_0 + Z_1\} + 3f(p)]\, dp \right.$$

$$\left. - mc \cdot (q(\tilde{P}_i(\cdot)) \cdot \left[3T_i - Z_0 + Z_1\right] + f(\tilde{P}_i(\cdot))) \right]$$

$$\frac{\partial TS^T}{\partial \tilde{Z}_i} = \cdot \left(\frac{\partial Q_{iH}}{\partial \tilde{T}_{iH}} - \frac{\partial Q_{iL}}{\partial \tilde{T}_{iL}} \right) \cdot (\tilde{P}_i(\cdot) - mc) \cdot \gamma_i$$

$$+ \int_{\tilde{P}_i(\cdot)}^{\tilde{P}_{i-1}(\cdot)} \left[\frac{\partial Q_{iH}}{\partial \tilde{T}_{iH}} - \frac{\partial Q_{iL}}{\partial \tilde{T}_{iL}} \right] dp = 0,$$

because $\frac{\partial Q_{iH}}{\partial \tilde{T}_{iH}} = \frac{\partial Q_{iL}}{\partial \tilde{T}_{iL}} = q(p)$. Because, from Theorem 1, $\frac{\partial \Pi^T}{\partial \tilde{Z}_i} < 0$, this means that CS^T is locally increasing in $\tilde{Z}_i$.

End of Proof

Thus, from the standpoint of economic efficiency, if uncertainty is small enough, the profit-maximizing set of OCPs is preferable to the profit-maximizing taper. The issue is ambiguous as far as the effect of γ on consumer surplus is concerned, although increases in Z around (0, 0) do unambiguously raise consumer surplus.

5. Conclusion

This paper is a beginning effort at accounting for the fact that nonuniform prices are implemented in two quite distinct ways: as optional calling plans and as tapered tariffs. The presence of customer-level demand uncertainty gives the firm a choice: would it prefer to have consumer choice take place at the start of the billing period, when demand uncertainty is present, or at the end of the billing period, when it is absent? The OCP approach is to force customers to choose at the start, from a set of two-part tariffs designed to price discriminate between the ex ante types $\{T_i\}$. The taper approach is to choose a set of two-part tariffs designed to price discriminate between the ex post types $\{\tilde{T}_i\}$ and to compute the lower envelope of this set of two-part tariffs as a single tapered tariff.

On an expected profit basis, the comparison between OCPs and tapers can go either way. The virtue of the OCP approach is that it can induce the consumer to select a tariff that is not incentive-compatible or even individually rational in an ex post sense. The advantage of the taper is that it can do a better job of excluding low ex post types that would otherwise limit the profit to be had from higher types. For small to moderate levels of uncertainty, the OCP approach is more profitable. Numerical simulations and historical evidence suggest that this result is fairly robust in practice.

The limitations of our results are severe, but may suggest issues for further work. First, we have assumed risk neutrality on the part of consumers, which rules out potentially interesting issues involving the insurance properties of tapers and OCPs. A flat-rate tariff for example, provides complete insurance against the effects of demand fluctuations on a consumer's bill, whereas a uniform price provides none. A taper has the property that, because large outcomes of demand will be on rate steps with low usage charges, the impact of incrementally large demand on expenditure is muted as compared to its impact under an OCP arrangement, in which usage charge does not decline with consumption level.

Second, we assumed that the ex ante taste parameters $\{T_i\}$ and the ex post taste parameters $\{\tilde{T}_i\}$ are ordered in the same way, so that if $T_i < T_{i+1}$, then $T_i + Z_1 < T_{i+1} + Z_0$, etc. Given that our interest is in cases where Z_0 and Z_1 are small, this may not matter too much, but this assumption may not be met in practice.

Third, we have assumed that $\tilde{Z}_i$ has a single realization soon after the start of a billing period. In practice, a billing period may contain numerous random shocks over time, each of which affects demand. As time goes on through a billing period, demands in remaining days are doubtless known with more and more certainty; but our assumption is still an extreme one. How much this matters depends partly on the question being asked. Analytically, thinking of $\tilde{Z}$ as a random process depending on time means that the "final" usage charge on a taper is a random variable. Thus, for the "small" uncertainty assumed for Theorems 1 and 2, this difficulty is probably unimportant. However, the nature of the dynamics of customer demand along a taper during a billing period is an interesting question for future study.

Finally, competitive issues are ignored. This is a particularly interesting area to explore further, because tariff history in telecommunication since in early 1980s suggests an interesting trend regarding OCPs and tapers in a competitive environment. In 1980, AT&T faced minimal competition and offered only two switched, voice-grade servies: MTS and WATS. The latter was sold on a per-line basis at a monthly flat rate of $3,000. Thus, consistent with Theorem 1, the market was characterized by OCPs. By the late 1980s, AT&T faced mounting competition from MCI, Sprint, and a variety of resellers. Although OCPs are still offered, each carrier now also offers tapers aimed at each market segment. AT&T's WATS, for example, now has a tapered usage schedule with a monthly flat charge as low as roughly $50. Its ProWATS tariff is a taper aimed at midsized users. This suggests that the present analysis could usefully be extended to study competition.

Notes

This research was substantially completed while Karen Clay was at Bellcore. We would like to acknowledge the helpful comments of Bridger Mitchell. The views represented herein represent those of the authors and do not represent those of Bellcore or its owners.

1. Formally, we have a two dimensional model of consumers. The choice of ex post tariffs by a monopolist facing heterogeneous customers with uncertain demand was examined by Srinagesh (1985) in a model with two dimensional consumers. Other authors who have considered models of multidimensional consumers include Seade (1979), Wilson (1990), and Laffont, Maskin and Rochet (1987). While all these papers (and ours) represent generalizations of the one parameter customer heterogeneity/demand uncertainty models that are now standard, none of them focuses on the issues of timing we raise in this paper.

References

Brock, G.W. 1981. *The Telecommunications Industry: The Dynamics of Market Structure.* Cambridge, Masachussets: Harvard University Press.

Brown, S.J. and D.S. Sibley. 1986. *The Theory of Public Utility Pricing.* Cambridge: Cambridge University Press.

Crémer, J. 1990. "Precontractual Gathering of Information." Paper presented at *The International Conference on Information, Incentives and Regulations.* University of Toulouse, September 5, 1990.

Faulhaber, G.R., and J.C. Panzar. 1977. "Optional Self-Selecting Two Part Tariffs." *Bell Laboratories Discussion Paper No. 74.*

Faulhaber, G.R. 1987. *Telecommunications in Turmoil.* Cambridge: Ballinger Press.

Goldman, M.B., H.L. Leland, and D.S. Sibley. 1984. "Optimal Nonuniform Pricing." *Review of Economics Studies* 51:305-19.

Heyman, D.P., J.M. Lazorchak, D.S. Sibley, and W.E. Taylor. 1987. "An Analysis of Tapered Access Charges for End Users." In *New Regulatory and Management Strategies in a Changing Environment,* 191-212. East Lansing, MI: Michigan State University.

Kridel, D., D. Lehman, and D. Weisman. 1990. "Option Value, Telecommunications Demand, and Policy." *mimeo.*

Laffont, J-J., E. Maskin, and J-C. Rochet. 1987. "Optimal Nonlinear Pricing with Two-Dimensional Characteristics." In *Information, Incentives and Economic Mechanisms: Essays in honor of Leonid Hurwicz",* edited by T. Groves, R. Radner, and S. Reiter. Minneapolis: University of Minnesota.

Lewis, Tracy and D.E.M. Sappington. 1990. "Selecting an Agent's Ability." Paper presented at *The International Conference on Information, Incentives and Regulations.* University of Toulouse, September 5, 1990.

Maskin, E., and J. Riley. 1984. "Monopoly with Incomplete Information", *Rand Journal of Economics,* 15: 171-96. Mitchell, Bridger M. 1978. "Optimal Pricing of Local Telephone Service" *American Economic Review,* 68: 517-537.

Mitchell, Bridger M. 1978. "Optimal Pricing of Local Telephone Service." *American Economic Review* 68:517-537.

Roberts, Kevin. 1979. "Welfare Considerations of Nonlinear Pricing." *Economic Journal* 89: 66-83.

Seade, J. 1979. "The Optimal Taxation of Multidimensional Consumers" *CEPREMAP,* mimeo.

Spence, A. Michael. 1977. "Nonlinear Prices and Welfare." *Journal of Public Economics* 8: 1-18.

Srinagesh, P. 1985. "Nonlinear Pricing with Heterogeneous Consumers and Demand Uncertainty" *Indian Economic Review* 20(2): 299-315.

Train, Kenneth E. 1989. "Optional versus Mandatory Measured Rates for Local Telephone Service", *Cambridge Systematics Inc.,* mimeo.

Train, Kenneth E., M. Ben-Akiva and D.L. McFadden. 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *Rand Journal of Economics* 18: 109-123.

Wilson, R. B. 1991. *Nonlinear Pricing,* manuscript.