Proceedings
of the Conference
on

TELECOMMUNICATIONS IN CANADA: Economic Analysis of the Industry

(MONTREAL, MARCH 4, 5, & 6, 1981)

Organized Jointly by
Ecole des Hautes Etudes Commerciales
and the
University of Victoria
in Collaboration with the
Department of Communications, Government of Canada

VOLUME II

Proceedings of the Conference on

TELECOMMUNICATIONS IN CANADA: ECONOMIC ANALYSIS OF THE INDUSTRY

(MONTREAL, MARCH 4, 5, and 6, 1981)

Organized Jointly By
Ecole des Hautes Etudes Commerciales
and the
University of Victoria
in Collaboration with the
Department of Communications, Government of Canada

VOLUME II

LOCAL MEASURED SERVICES

IDENTIFYING TARIFF-INDUCED SHIFTS IN THE SUBSCRIBER DISTRIBUTION OF LOCAL TELEPHONE USAGE

T.F. WONG

Bell Laboratories

ABSTRACT

Most residential and many business telephone subscribers are currently charged a fixed monthly fee, called a flat rate, for their local telephone service. Under flat rate pricing, there is no additional charge for local telephone calls. Conversion of the telephone tariff to measured service, where there are incremental charges for local telephone usage, would be expected to impact the demand for such usage. In this paper, we describe a procedure for quantifying the shift in the subscriber distribution of telephone usage caused by such conversions. This procedure is applied to data on a particular rate conversion. A situation where the usage distribution follows a Weibull function is discussed as an example of this proposed approach.

Identifying Tariff-Induced Shifts in the Subscriber Distribution of Local Telephone Usage

T. F. Wong*

1. Introduction

Most residential and many business subscriber to local telephone service are charged at a flat rate (FR). Within a prescribed local calling area, the customers can make as many calls and talk as long as they like for a fixed monthly fee. Recently, actions by regulatory commissions, pressure from consumer organizations and changes in the economic environment of the telephone business (Garfinkel 1979) have caused consideration of a move to measured telephone service (MS). In the MS environment customers will be billed for their actual local usage over some allowance. Usage charges may depend on the frequency, duration, distance and time of day/day of week of local calls, much like long distance calls are charged for today. When such FR to MS conversion is undertaken, customer usage characteristics are expected to change. These changes will be a crucial input in evaluating the impact of such rate conversion, for the amount of local usage is a major determinant of the resulting telephone company revenues, expenses, and required investments.

Unlike many demand studies, analyses of usage changes at the aggregate or market level may not be sufficient in this case. When the tariff contains a usage allowance, or if the bill rendered to the customer as a function of usage is otherwise nonlinear, aggregate demand is insufficient knowledge to calculate resulting revenues. In addition, there is often interest in such questions as what percent of the customers have higher/lower telephone bills as a result of the conversion. Consequently, past studies (Pavarini 1978) of customer usage change in response to FR to MS tariff changes have modeled the usage under MS of *individual customers* as a function of their previous usage under FR, price per unit usage in the new MS tariff and other economic variables. This disaggregate modeling requires a large customer usage data base

* Bell Laboratories

and is complicated to execute. Analysis at this level of detail, however, may not be necessary for many useful studies. Since often all that is required is knowledge of the subscriber distribution of usage (and changes in it induced by the conversion), analyses aimed specifically at quantifying usage distributions might be considered. Such an approach is described below.

The new (post-conversion) usage density function, $g(x_M)$, is related to the old density, $f(x_F)$, by

$$g(x_M) = \int G(X_M | x_F) f(x_F) dx_F \tag{1}$$

Where the variates x_M and x_F are usages, say in minutes/month, under MS and FR respectively. The conditional probability density G allows for a stochastic transformation, i.e. one where customers previously at the same usage level are allowed to react differently. The integration in x_F offers the prospect for relatively simple relations among $f(x_F)$ and $g(x_M)$. The crucial point is that these simpler relations are sufficient for many of the study purposes discussed before.

In this note, I shall discuss a procedure for direct identification of the shift of the usage distribution, in a situation where the usage distributions before and after the tariff conversion are both assumed to have a Weibull form.

I first explain why I use a Weibull distribution. It was observed (Pavarini 1978) that the local telephone usage distribution under FR can be represented by a truncated powernormal distribution. In section 2, I shall use the result that with the proper selection of parameters, a Weibull distribution can approximate a powernormal distribution. This result makes Weibull a plausible representation of the FR usage distribution. It has the added advantage of being simpler to manipulate analytically than the powernormal.

In section 3, I shall present an exploratory analyses of the subscriber distribution of local telephone usage data from a Denver, Colorado flat-to-measured conversion in 1971 called METROPAC (Metropolitan Preferred Area Calling). I found that a Weibull distribution is adequate to describe the data both under FR and under MS. Details of the statistical analyses and goodness-of-fit test are given in Appendix A.

Assuming that the usage distributions before and after the tariff conversion are represented by a Weibull function, two transformations were found that will conserve the Weibull form. These transformations also imply simple relations among Weibull parameters that can be tested (section 4). The METROPAC data was found to be consistent with this idea (distribution in total connect time seems to follow a power transformation whereas distribution in frequency follows a linear transformation) and I have simple identification of the shift of telephone usage distribution in response to tariff change.

Further discussion of this procedure is presented in section 5.

Although I have limited my discussion so far to the specific case of the change of local telephone usage characteristics in response to an FR to MS conversion, these procedure and results could potentially have application to other fields, e.g. utility services like electricity and water or quality assurance testing. Whenever the before and after distribution of some measure are characterized by a Weibull (or equivalently by a powernormal distribution, see section 2), simple relations could be found which determine the response of the system to the change (tariff rates or the testing procedure in the above examples).

2. Equivalence of Weibull and Powernormal Distribution Functions

A Weibull distribution has the cumulative probability distribution function of the variate x (e.g. a customer's usage in minutes/month)

$$P_W\ (x;\alpha,\beta) = 1 - e^{-(\frac{x}{\alpha})^{\beta}} \tag{2}$$

whereas the powernormal has[1]

$$P_{PN}\ (x;\lambda,\mu,\sigma) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}z^2} dz,\ t = \frac{x^{\lambda}-\mu}{\sigma} \tag{3}$$

α and β in eq. (2), λ, μ and σ in eq. (3) are constant parameters. [When $\lambda=1$, eq. (3) is just the normal distribution P_N]

1. i.e., Y ~ Power-Normal (λ, μ, σ) iff $Y^{1/\lambda}$ ~ Normal (μ, σ). Customers do not have negative usage. A truncated form $\frac{\Phi(x) - \Phi(o)}{1 - \Phi(o)}$ should be used instead of eq. (3). This approximation is justified *a posteriori* by the observation that the origin is 3.24 times σ away from the mean or $\Phi(o) = 0.0006$ which is negligible.

It was observed that the Weibull distribution in (x/α) is similar in shape to a powernormal distribution with $\lambda = \frac{1}{3.60} \beta = 0.28\beta$ (Johnson 1970). Suppose the usage data can be represented by a Weibull distribution, and we have estimated the parameters α and β. The above result shows that the data in the form of $(\frac{x}{\alpha})^{0.28\beta}$ will be equally well represented by a normal distribution, with $\mu = 0.901$ and $\sigma = 0.278$. Comparing with eq. (3)

$$t = \frac{(\frac{x}{\alpha})^{0.28\beta} - 0.901}{0.278}$$

$$= \frac{x^{0.28\beta} - 0.901(\alpha)^{0.28\beta}}{0.278(\alpha)^{0.28\beta}} \quad (4)$$

We conclude that the usage data can be represented by a powernormal distribution with $\lambda = 0.28\beta$, and a fixed mean to standard error ratio of 3.24.

Conversely, if the telephone usage distribution can be represented by a powernormal function with a mean to standard error ratio of 3.24, a Weibull distribution will be an equally good representation (Dubey 1967). In the following section we shall study the hypothesis that the Weibull function is the underlying distribution for usage distribution under both FR and MS tariffs.

3. Exploratory Analyses of METROPAC Usage Data

Metropolitan Preferred Area calling (METROPAC) is an *optional* offering in five exchanges in the state of Colorado. These exchanges are situated just outside the border of the Denver metropolitan local calling area. Under the *standard* FR tariff, residence customers in the five subject exchanges reached only those terminals in their own exchange for fixed monthly charges in the range of $4.05 to $5.05.

In response to demands by subscribers residing outside the metro area to be included in a larger flat rate area, the optional METROPAC offering was created in July 1969. A subscriber opting for the service paid an extra fixed charge ($6.25 to $7.30, depending upon his exchange) and in return was able to reach all exchanges within a thirty mile radius at flat rate. There was no limitation on the number or duration of outgoing calls.

The high market penetration coupled with the high usage of the METROPAC subscribers placed considerable burden on the network. In June of 1970, Mountain Bell of Colorado filed to convert METROPAC to a measured offering. Under the new tariff, subscribers would pay a fixed charge for METROPAC ($5) for which they would receive an allowance measured in total connect time of 60 minutes. Each minute over the allowance was charged at $0.08. Measured METROPAC became effective in January of 1971.

Individual subscribers' monthly usages in minutes per month were available over two four-month periods, April through July in 1970 under FR and the same months in 1971 under MS. To smooth out the month-to-month variation, each subscribers' usage had been averaged over these four months. In other words, I have two samples (FR and MS) of subscriber distribution of usage. A total of 383 permanent residence customers, who subscribed to both FR and MS METROPAC, were tracked.

Assuming that the underlying distribution is a Weibull function, I used the Maximum Likelihood Method (MLE) (Leone 1960) to estimate the scale parameter α and the shape parameter β (eq. 2), separately for the before and after usage distribution. The results and the estimated standard errors are summarized in table I of Appendix A. I used S^* statistics (Dubey 1966) to test the goodness of fit and the hypothesis that α and β equal the MLE estimated values. I used Chi-square statistics (Hahn 1967) to test the goodness-of-fit and the hypothesis that Weibull is the underlying distribution. Details of these statistical analyses are contained in Appendix A.

The probability density (p.d.) of the subscriber distribution data, in the form of a histogram, and the fitted Weibull distribution are shown in Fig. 1 (Appendix A). The Weibull function provides an adequate description, except for two bins near the peak of the density, of the subscriber distribution of usage both before and after MS conversion.

Individual subscribers' usages can be measured in frequency (calls per month), although the total connect time is a more natural choice for the METROPAC tariff. I have another two independent samples (FR and MS) of the distribution of subscriber usage measured in calls/month. I repeated the statistical analyses in Appendix A. Corresponding results are summarized in table II and Figure 2.

4. Shift of Usage Distribution

Suppose the usage distribution before and after a tariff change (e.g. FR to MS) can be represented by a Weibull function, with parameters $\alpha_F,\beta_F;\alpha_M,\beta_M$ respectively. That is (Burlington 1970),

$$P_W(x_F;\alpha_F,\beta_F) = 1 - e^{-\left(\frac{x_F}{\alpha_F}\right)^{\beta_F}} \tag{5}$$

$$P_W(x_M;\alpha_M,\beta_M) = 1 - e^{-\left(\frac{x_M}{\alpha_M}\right)^{\beta_M}} \tag{6}$$

where P_W is the cumulative probability and x is customer usage, for example in minutes/month.

Given this hypothetical situation, we search for possible relations among the flat and measured Weibull parameters. The first step is to look for classes of variable transformation that will conserve the Weibull form.

4.1 Power Transformation

Let us rewrite eq. (5) by raising x_F and α_F to a constant power δ. To maintain the equality[2] we divide β_F by the same content.

$$P_W(x_F;\alpha_F,\beta_F) = 1 - e^{-\left(\frac{(x_F)^\delta}{(\alpha_F)^\delta}\right)^{\frac{\beta_F}{\delta}}}$$

$$= P_W\left[(x_F)^\delta;(\alpha_F)^\delta,(\frac{\beta_F}{\delta})\right] \tag{7}$$

2. Eq. (7) requires the *equality* of the cumulative probabilities at variates x_F and x_M respectively. If this requirement is relaxed, then eqs. (9) and (10) need not be true.

where δ is a constant (independent of usage) not yet specified. We get another Weibull distribution with different variate and parameters.

From eq. (7) if we assume a "power" transformation (which is non-trivial) and identify

$$x_M = (x_F)^{\delta}; \quad \delta \neq 1 \tag{8}$$

and if the usage distribution under FR can be fitted by a Weibull function (section 3), then the usage distribution under MS will also follow a Weibull function, with new parameters

$$\alpha_M = (\alpha_F)^{\delta} \tag{9}$$

$$\beta_M = \frac{1}{\delta}\beta_F \tag{10}$$

These equations can be rewritten

$$\delta = \frac{\ln\alpha_M}{\ln\alpha_F} \tag{11}$$

$$\delta = \frac{\beta_F}{\beta_M} \tag{12}$$

and should be tested: The RHS of eq. (11) and (12) can be calculated from the estimated parameters and their values should be compared for equality. If the data is consistent with eqs. (11) and (12), we have a simple way to identify and forecast the shift of usage distribution in terms of the change in Weibull parameters [eqs (9) and (10)].

The constant δ would be expected to depend on the tariff parameters, income, the level of other prices and perhaps other factors at the time of the tariff change and in a particular location. In principle if we have several tariff change experiments, the structure of δ can be mapped out in detail.

For a Weibull distribution (Burlington 1970)

$$\mu = \alpha\,\Gamma\,(1 + \frac{1}{\beta}) \tag{13}$$

For usage distributions similar to those reported by Pavarini (1978), reasonable values of β are in the range 0.9 to 2.0. For these β, the gamma function attains values around unity (1.05 to

0.89),

$$\mu_M \doteq \alpha_M, \ \mu_F \doteq \alpha_F \tag{14}$$

and

$$\frac{\ln\mu_M}{\ln\mu_F} \doteq \frac{\ln\alpha_M}{\ln\alpha_F} \tag{15}$$

From the results of the exploratory analyses of the METROPAC data (Table I of Appendix A), we find that

$$\frac{\ln\alpha_M}{\ln\alpha_F} = \underset{(0.01)}{0.77} \tag{16}$$

$$\frac{\beta_F}{\beta_M} = \underset{(0.05)}{0.89} \tag{17}$$

$$\frac{\ln\mu_M}{\ln\mu_F} = 0.76 \tag{18}$$

where α and μ are measured in minutes/month and the values in brackets are estimated standard errors. These values are in rough agreement with eqs. (11), (12) and (15). Comparing with eqs. (11) and (12), the weighed (by inverse variance) estimate of δ is

$$\hat{\delta} = \underset{(0.01)}{0.77} \tag{19}$$

4.2 Linear Transformation

There is another simple transformation that will preserve the Weibull function. If we multiply x_F and α_F by the same constant ϵ, eq. (5) becomes

$$P_W\ (x_F;\alpha_F,\beta_F) = 1 - e^{-(\frac{\epsilon x_F}{\epsilon\alpha_F})^{\beta_F}}$$

$$= P_W[(\epsilon x_F);(\epsilon\alpha_F),\beta_F] \tag{20}$$

The linear transformation

$$x_M = \epsilon x_F; \ \ \epsilon \neq 1 \tag{21}$$

implies

$$\alpha_M = \epsilon \alpha_F \tag{22}$$

$$\beta_M = \beta_F \tag{23}$$

Eqs. (13) and (23) give

$$\frac{\mu_M}{\alpha_M} = \frac{\mu_F}{\alpha_F} \tag{24}$$

Finally, eqs. (23) and (24) can be rearranged

$$\frac{\beta_F}{\beta_M} = 1 \tag{25}$$

$$\frac{\alpha_M}{\alpha_F} = \frac{\mu_M}{\mu_F} \tag{26}$$

and should be tested. The LHS can be calculated from the estimated parameters and the RHS of eq. (26) can be calculated from the mean usage data.

From the results of the analyses of the METROPAC data in table II of Appendix A, we find that

$$\frac{\alpha_M}{\alpha_F} = \begin{matrix} 0.37 \\ (0.02) \end{matrix} \tag{27}$$

$$\frac{\beta_F}{\beta_M} = \begin{matrix} 1.01 \\ (0.06) \end{matrix} \tag{28}$$

$$\frac{\mu_M}{\mu_F} = 0.36 \tag{29}$$

where α and μ here are measured in calls/month. These results are in good agreement with eqs. (25) and (26). The estimated ϵ [eq. (22)] is

$$\hat{\epsilon} = \begin{matrix} 0.37 \\ (0.02) \end{matrix} \tag{30}$$

Distributions in total connect time are related to distributions in frequency in a non-trivial way [the individual subscriber's total connect time is the product of his call frequency and his average holding time in that month]. I can not offer any explanation of why these distributions follow different transformations. The observation that their pre- and post-conversion parameters are so simply related is just amusing.

4.3 Comparison with Pavarini's Repression Model

Pavarini (1978) studied individual customer usage in response to FR to MS tariff change. He attempted to construct a demand function of a household, and to identify the relevant explanatory variables. In his preliminary study, he concluded that

$$\hat{p} = \frac{p}{(p_e)^{0.5} y^{0.5}} \tag{31}$$

is the appropriate variable, where p is the unit charge in the MS tariff, p_e is a price index for other goods, and y is the household disposable income. The form of (31) was motivated by the budget constraint and it captures the effect of both inflation and regional difference in income. [Effect of other demographic factors are presumably reflected in the FR usage]. We could use consumer price index (CPI) to measure p_e and medium household income (MHI) to measure income in a relatively homogeneous neighbopod.

The repression indexes, δ and ε defined above, are subjected to the boundary condition

$$\begin{cases} \delta = 1 \\ \epsilon = 1 \end{cases} \quad when \quad p = o \tag{32}$$

as they, by definition, measure the *change* of usage in response to FR ($p=o$) to MS conversion.

In the absence of other data, I *assume* a linear dependency of δ and ϵ in $\hat{p}$. Results in eqs. (19) and (30) then give

$$\delta = 1 - \frac{(3166 \pm 138)}{(CPI)^{0.5}(MHI)^{0.5}} p \tag{33}$$

$$\epsilon = 1 - \frac{8671 \pm 275}{(CPI)^{0.5}(MHI)^{0.5}} p \tag{34}$$

where I have used CPI = 122 [Labor 1978] and MHI = $9938 [Ziprofile 1980] for the Denver suburb (Zipcode 80501) in 1971.

These empirical results, eqs. (33) and (34), provide a complete quantification of the shift of subscriber distribution. They should be tested against data from other conversion experiments when available.

5. Discussion

I have proposed a procedure to directly study the shift of the subscriber usage distribution in response to a tariff change. In the case of a Weibull-to-Weibull shift, simple relations among the parameters could be found. The METROPAC data was found to be consistent roughly with these ideas. Eqs. (11) and (12) or (25) and (26) should be tested in detail with data from local telephone tariff change experiments.

The transformations in eqs. (8) and (21) may not be the only ones[3] that conserve the Weibull distribution. It will be useful to conduct a systematic search for a complete set of possible transformations that preserves the distribution form. This however is beyond the scope of this short note.

The approach presented here is partially supported by empirical observation. It would be interesting to construct a model, with economic theory and consumer behavior built in, from which quantities such as distribution functions can be generated directly.

Alternatively, results obtained here could be used to guide an ambitious model development. One could construct an individual demand function under a general tariff. The model, when partially aggregated, should reproduce the shape of subscriber distribution that I observed. In this approach distribution parameters would be explicit functions of the tariff and other economic variables. Then the model should be estimated on data and its performance in predicting usage change verified.

I have presented a specific situation and found simple relations to hold between post and pre conversion distributions. This result indicates that a distribution-to-distribution study is a potentially rewarding approach. There are many systems which can be described by a Weibull, or powernormal distribution function, and our simple relations, eqs. (9) and (10) or (22) and (23) may be relevant in identifying such systems' response to external stimuli.

3. The discrete transformation, minimum value of a Weibull, gives rise to a Weibull distribution. It is not helpful in the problems that we are concerned.

Appendix A

I shall summarize the statistical analyses of the subscriber distributions of usage, before and after MS conversion, of the 383 permanent subscribers in METROPAC. The background of this MS conversion was discussed in section 3.

a) Assuming that the underlying distribution is a Weibull function, with parameters α and β (eq. 2), I used the Maximum Likelihood Method (Leone 1960) to estimate the unknown paramters. They are the solutions to

$$\frac{n}{\beta} - n\frac{\sum (x_i)^\beta \ln x_i}{\sum (x_i)^\beta} + \sum \ln x_i = 0 \tag{A1}$$

$$\alpha^\beta = \frac{\sum (x_i)^\beta}{n} \tag{A2}$$

where n, the sample size, equals 383.

The values of (α_F, β_F), (α_M, β_M) and their estimated standard errors (in brackets), for the FR and MS distribution respectively, are listed in table I. The associated variance was estimated by (Engelhardt 1977)

$$nVar(\alpha) \doteq 1.1624 \left(\frac{\alpha}{\beta}\right)^2$$

$$nVar(\beta) \doteq 0.6482\beta^2 \tag{A3}$$

b) To test the goodness of fit and the hypothesis that α and β equal the MLE estimated values, Dubey (1966) suggested a test function which involves all n observations

$$S^* = 2\sum_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^\beta \quad , n = 383 \tag{A4}$$

Under the null hypothesis this function obeys the Chi-square distribution with 2n ($n_D = 766$) degrees of freedom. For such a large n_D, the x^2 distribution is well approximated by a normal distribution with variate (Abramovitz 1964).

$$\frac{x^2 - n_D}{\sqrt{2n_D}} \tag{A5}$$

I followed this procedure and calculated S^* in table I. The estimated values cannot be rejected at 57% confidence level.

c) To test the goodness-of-fit and the hypothesis that Weibull is the underlying distribution, I calculated the Chi-square statistics after dividing the data into 41 cells. The boundary of these cells were determined such that every one of them have the expected probability of 1/41 or 9.34 observations in our sample. (Hahn 1967).

The x^2 values, together with the corresponding confidence level (Harter 1964) are listed in table I. The low confidence level is perhaps not suprising. It is well known (Hahn 1967) that the more data there are (here n=383) the better are the chance of rejecting any model.

I plot the probability density of the usage data (histogram) and the fitted Weibull distribution in Fig. 1. The Weibull functions fail to reproduce the magnitude of the bins near the peak. They do provide an adequate description of the data before and after MS conversion.

d) The above analyses were repeated for the FR and MS subscriber distribution in frequency (calls per month). Results are summarized in table II and Fig. 2.

Table I

Estimated Parameters and Test Functions of the
Assumed Weibull Distribution of Usages in Total Connect Time

	μ (MINS/MONTH)	α (MINS/MONTH)	β	$x^2(n_D=38)$	C.L.	$S^*(n_D=766)$	C.L.
FR	484.13	541.97 (20.04)	1.49 (0.06)	51.09	7.50%	758.86	57%
MS	112.75	127.79 (4.22)	1.67 (0.07)	78.28	0.02%	759.51	57%

FR = Individual customer usage averaged over 4 months (April to July 1970) under Flat Rate.

MS = Individual customer usage averaged over 4 months (April to July 1971) under Measured Service.

C.L. = Confidence level

Table II

Estimated Parameters and Test Functions of the Assumed Weibull Distribution of Usages in Frequency

	μ (CALLS/MONTH)	α (CALLS/MONTH)	β	$x^2(n_D=38)$	C.L.	$S^*(n_D=766)$	C.L.
FR	70.95	79.73 (2.48)	1.77 (0.07)	33.74	66.0%	772.00	44%
MS	25.88	29.33 (0.92)	1.75 (0.07)	53.66	4.5%	761.60	54%

FR = Individual customer usage averaged over 4 months (April to July 1970) under Flat Rate

MS = Individual customer usage averaged over 4 months (April to July 1971) under Measured Service

C.L. = Confidence level

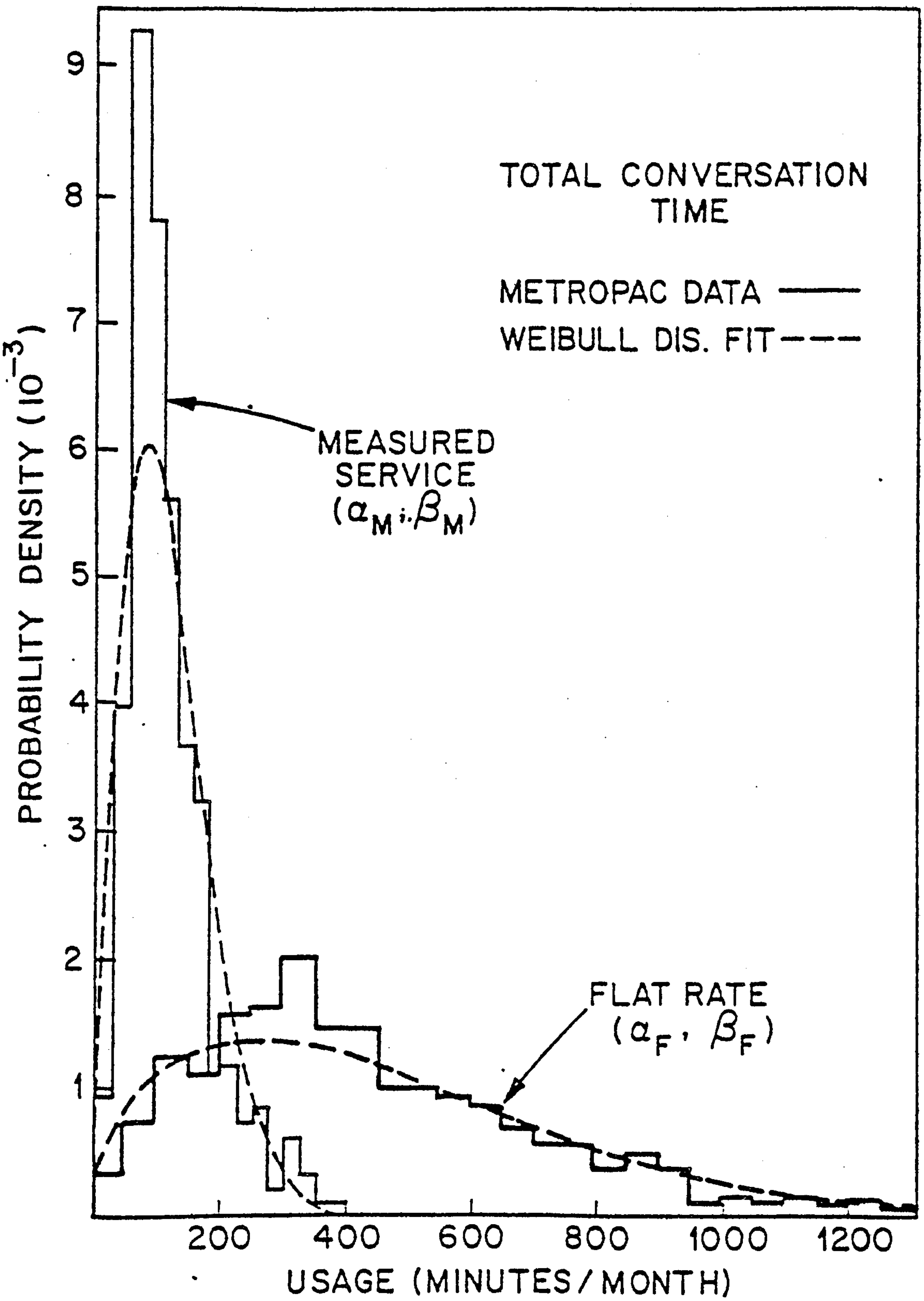


Fig.1 COMPARISON OF THE METROPAC USAGE DISTRIBUTION DATA AND THE MAXIMUM LIKELIHOOD FIT BY A WEIBULL FUNCTION.

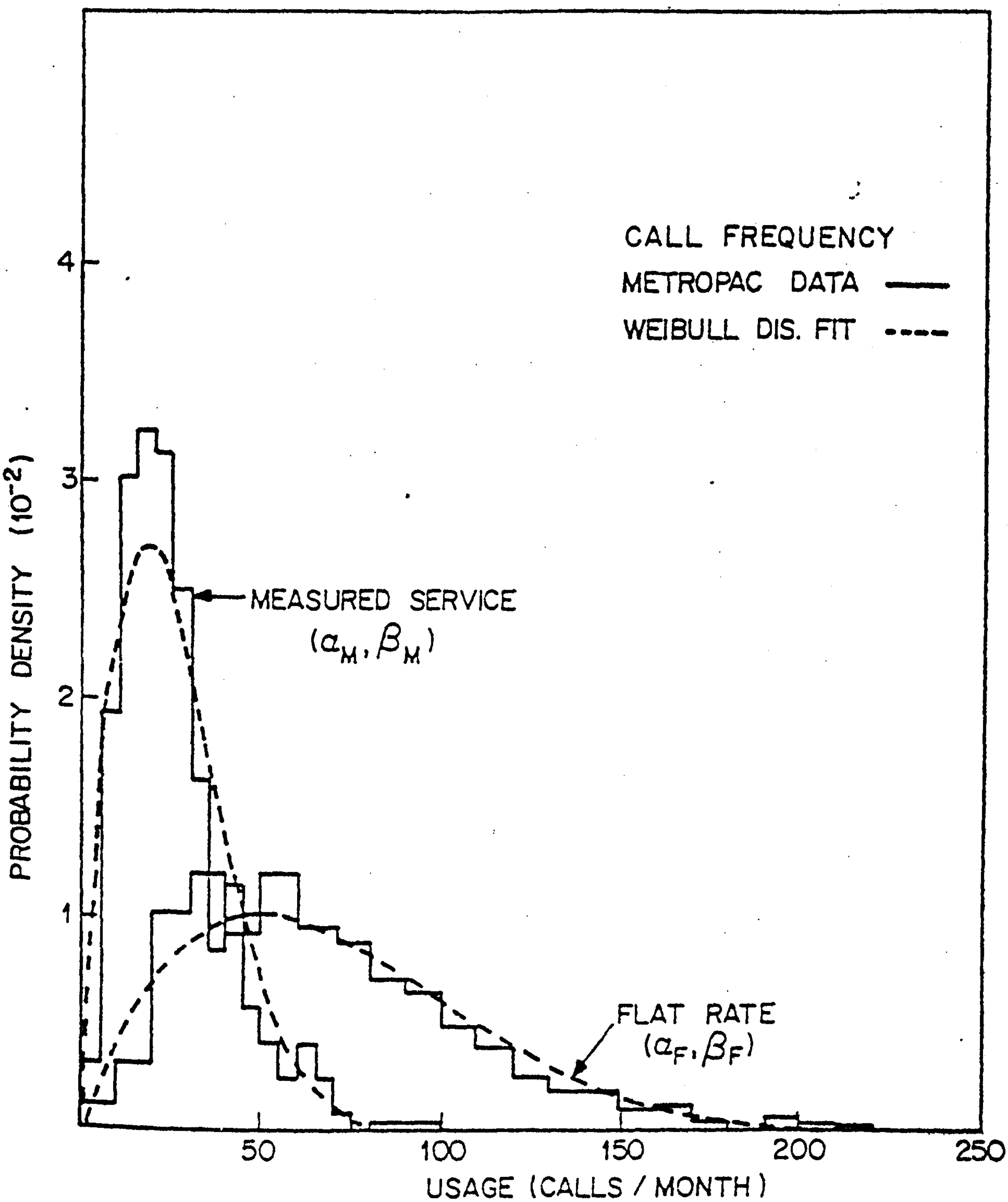


FIG. 2 COMPARISON OF THE METROPAC USAGE DISTRIBUTION DATA AND THE MAXIMUM LIKELIHOOD FIT BY A WEIBULL FUNCTION

Note Added in Proof

Although the use of a Weibull function as the representation of usage distribution was motivated by Pavarini's result and the "equivalence" of Weibull and powernormal, the significance of such assumption was *directly* tested by the goodness-of-fit tests as discussed in the paper. The reader may notice that the confidence level (according to chi-square statistic) ranges from a low of 0.02% to a high of 66% (tables I and II).

After the submission of this paper, I have reanalyzed the same METROPAC data with a lognormal function. The new confidence levels are very high and the results and MLE fits are summarized in a figure.

It can be shown that transformations in egs. (8) and (21) also conserve the lognormal functional form, with the (transformed) mean and standard deviation (μ and σ) related in simple ways.

power transformation (total connect time data)

$$\mu_M = \delta \mu_F \qquad \sigma_M = \delta \sigma_F$$

and a consistency condition

$$\underset{(0.005)}{\overset{\text{(Data)}}{0.134}} = \frac{\sigma_M}{\mu_M} = \frac{\sigma_F}{\mu_F} = \underset{(0.005)}{\overset{\text{(Data)}}{0.127}}$$

linear transformation (call frequency data)

$$\mu_M = \mu_F + \ln \epsilon$$

and a consistency condition

$$\underset{(0.022)}{\overset{\text{(Data)}}{0.614}} = \sigma_M = \sigma_F = \underset{(0.023)}{\overset{\text{(Data)}}{0.631}}$$

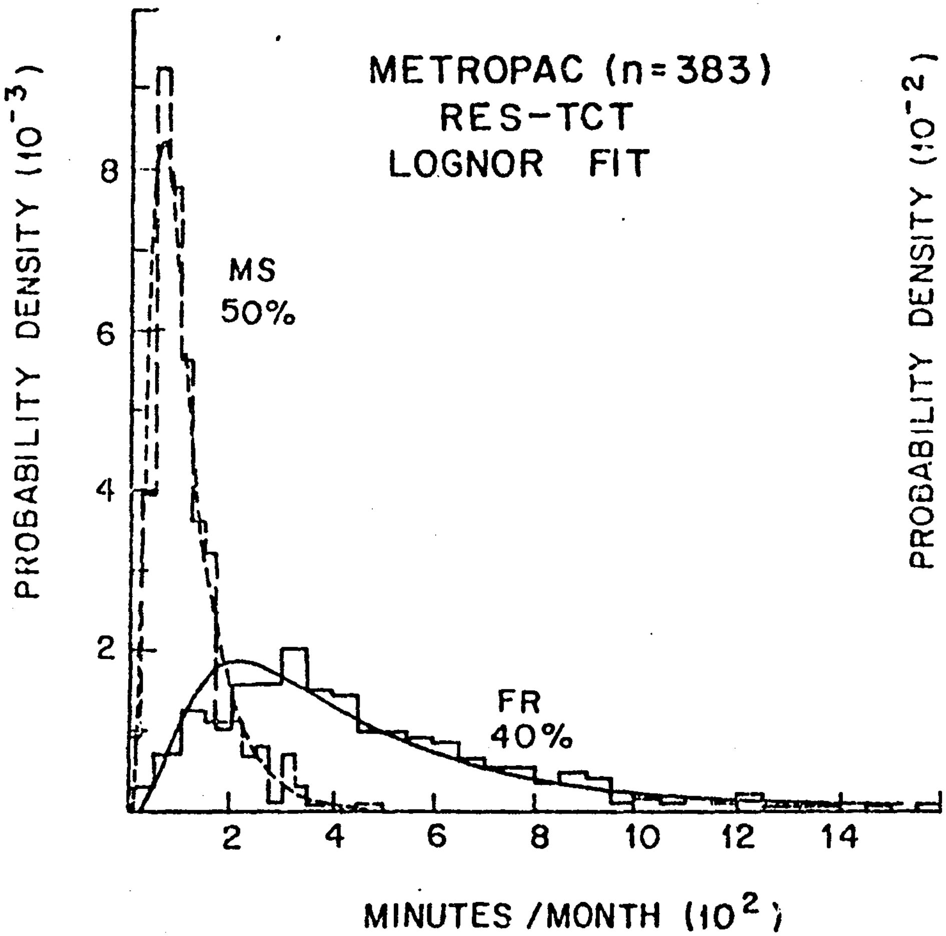
In both cases, the consistency conditions are well satisfied. Furthermore, the estimated repression indices are "robust" in the sense that the new values (based on lognormal fit)

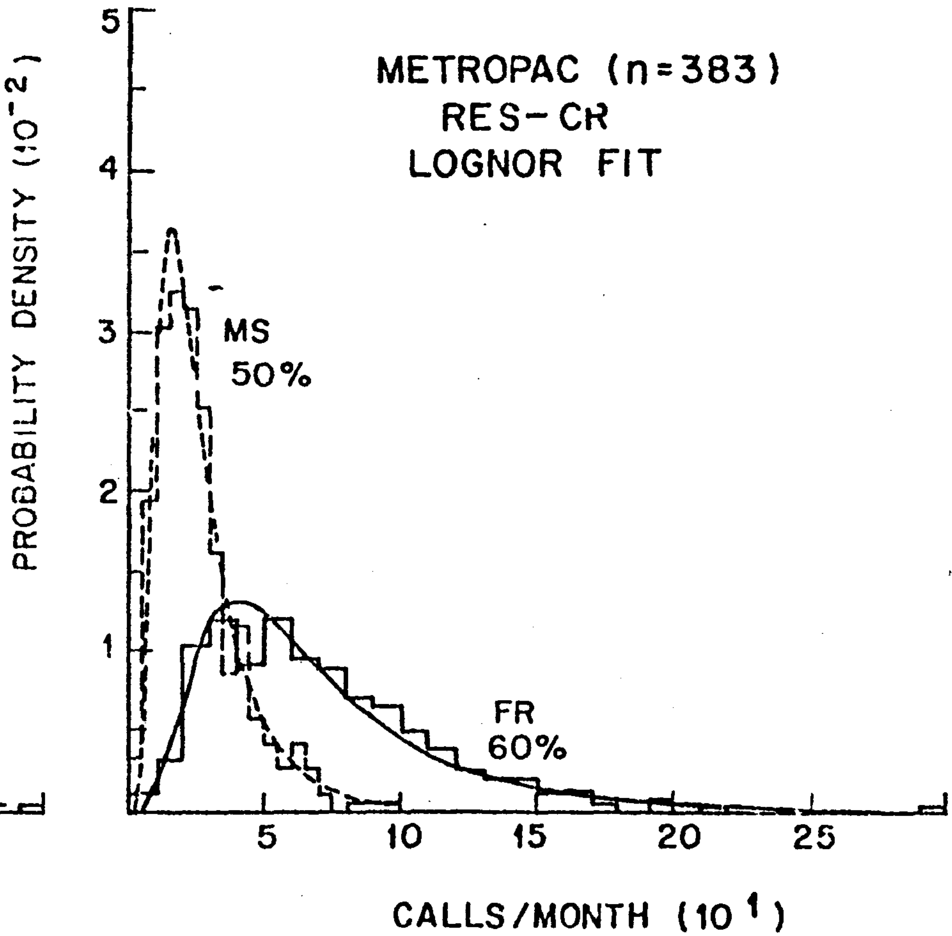
$$\hat{\delta} = \underset{(0.007)}{0.767}$$

$$\hat{\epsilon} = \underset{(0.096)}{0.366}$$

are almost identical to the old results (egs. 19 and 30).

Finally, procedures discussed in this paper have been applied successfully to a mandatory MS conversion implemented in Ohio in 1978. This result and other studies will be presented elsewhere.


METROPAC (n=383)
RES-TCT
LOGNOR FIT
MS
50%
FR
40%
PROBABILITY DENSITY (10⁻³)
2
4
6
8
2
4
6
8
10
12
14
MINUTES /MONTH (10²)


METROPAC (n=383)
RES-CR
LOGNOR FIT
MS
50%
FR
60%
PROBABILITY DENSITY (10⁻²)
1
2
3
4
5
5
10
15
20
25
CALLS/MONTH (10¹)

LOCAL TELEPHONE COSTS AND THE DESIGN OF RATE STRUCTURES

B. MITCHELL

Rand Corporation

I. INTRODUCTION

A well-developed body of economic theory is available to guide the setting of prices for the multi-product regulated firm. Economic efficiency can be increased by designing rate structures that incorporate the basic principles developed from this theory. These principles call for provisionally pricing each of the firm's outputs at its marginal cost, testing to determine whether such rates satisfy a specified budget constraint (e.g., revenues = costs), and then suitably modifying the marginal-cost rates in order to satisfy the constraint. Most commonly, the trial rates produce insufficient revenue, and then rates must be raised according to the Ramsey rule--prices are increased above marginal costs in inverse proportion to the individual price elasticities of demand. This paper applies ratemaking theory to the design of rate structures for local telephone calls that efficiently reflect the costs of the local network.

The principal costs of supplying local telephone calls are embodied in the switching capacity of a local central office (exchange) and the trunking capacity that connects local offices together. (The dedicated local loop connecting the subscriber to the office is, of course, essential but its cost is independent of telephone usage). Several operating companies which are proposing to introduce local measured service are conducting special studies that will gather information about these costs, and how they vary with maximum loads. How should these costs, which relate to specific items of network equipment, be used to develop prices for telephone calls, which are commonly classified in terms of the hour at which they are placed and the distance between subscribers?

The markets for telephone services are characterized by high capacity-related equipment costs and very low variable (traffic) costs, the joint use of some equipment by several outputs, and the grouping together of several outputs that are charged a common price. The following sections develop a series of simple models that successively incorporate these basic elements. Throughout the paper I make several simplifying assumptions:

- the unit of output is a "call" of fixed duration, and there is a uniform rate of demand during any given period
- demand for a given output depends only on its own price, so that there are no temporal or spatial cross-elasticities
- all costs are due to providing capacity to meet the maximum rates of output and capacity can be constructed at constant returns to scale
- the costs of connecting subscribers to the telephone network via local loops are recovered in fixed monthly charges

II. A SINGLE EXCHANGE

All telephone calls originate and terminate in one exchange and this output is produced using a single component of switching capacity. There are only two commodities, x^1 and x^2, the number of calls made in two equal-length periods ("day" and "night"). This situation is a version of the well-known Boiteux-Steiner peak-load pricing problem in which a homogeneous resource with a maximum capacity is available to produce output in each period.

The economic structure of the one-exchange telephone call market is given by the rate of demand functions for the two periods

(1) $x^1 = x^1(p^1), \quad x^2 = x^2(p^2),$

the required capacity

(2) $K = \max_t \{x^t\},$

and total cost

(3) $C = \beta K$

where β is the per-call unit capacity cost.

The marginal cost of a commodity is the change in total cost that results from a one-unit increase or decrease in the production of that commodity. In the long run, capacity can be adjusted to meet maximum demand. Therefore if $x^1 > x^2$ at the observed prices, the marginal costs are

$$mc^1 = \partial C/\partial x^1 = \beta$$

(4) $mc^2 = \partial C/\partial x^2 = 0.$

This basic model can be used to illustrate several approaches to pricing telephone service. Each rate structure can be evaluated in terms of its effect on economic welfare, measured by the sum of consumers' and producer's surplus

$$(5) \qquad W = CS + PS = \sum_t \int_{p_t}^{\infty} x^t(\xi^t)d\xi^t + \sum_t p^t x^t - \beta \max_t\{x^t\} .$$

1. Flat-rate pricing

Set $p^1 = p^2 = 0$. The total (usage) cost of local telephone service is recovered by increasing the fixed monthly charge per subscriber. Such flat-rate pricing seems inefficient. But because prices are zero, equipment to measure the number of calls is not needed, and the resulting saving in resources can outweigh the gains of per-call charges. Nevertheless, in order to focus on the design of usage-sensitive rate structures I will neglect measurement costs in this paper.[1]

2. Average-cost pricing

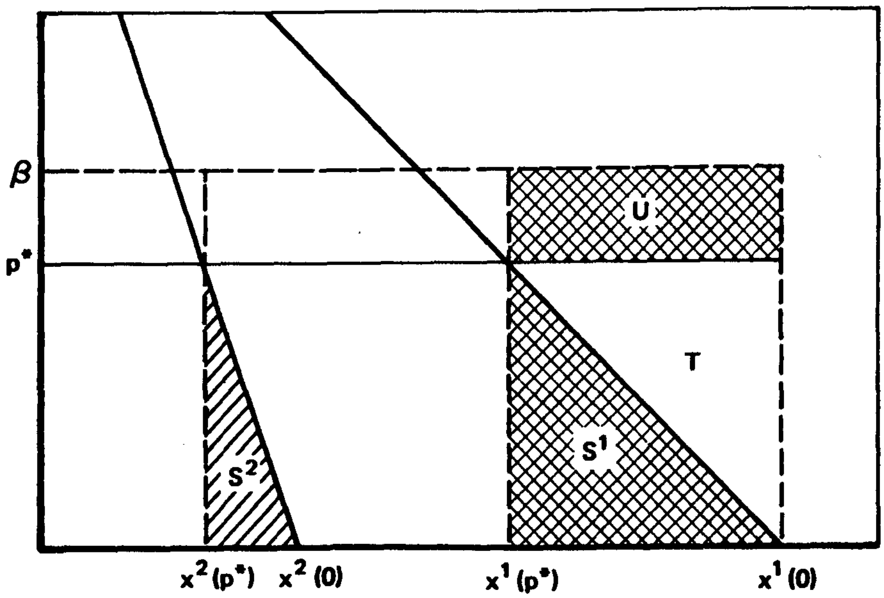
Set $p^1 = p^2 = p^* =$ average cost $= C/(x^1+x^2)$. Charging a positive price p^* per call is seemingly more efficient than flat-rate pricing. In period 1 capacity is a scarce resource; the reduced demand due to the positive price will reduce calling and therefore capacity and total costs. But calling will also be reduced in period 2, even though excess capacity is available. Compared to flat-rate pricing the net result can be either a gain or loss in welfare.

[1]For a comparison of benefits and costs under flat versus measured rates in a simple case, see Mitchell, 1980.

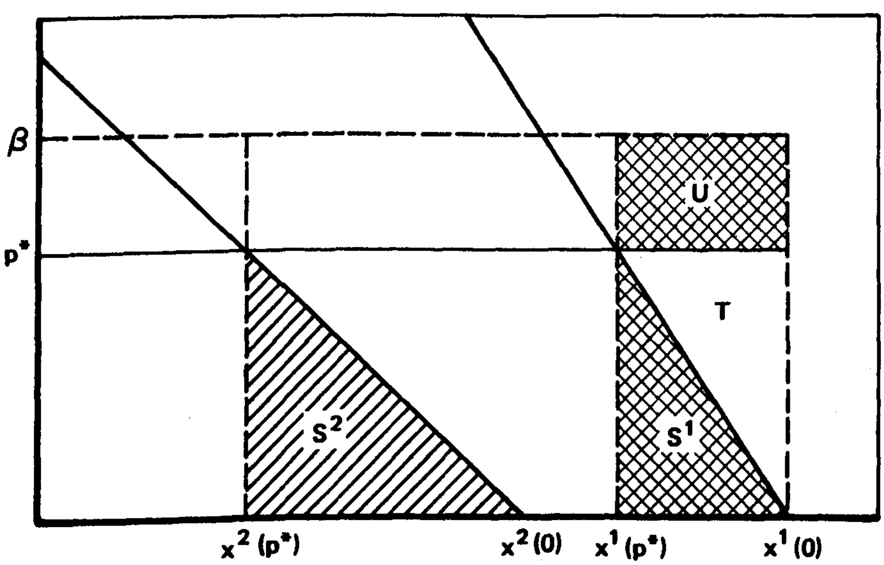
In Fig. 1 the reduction in calling from flat-rate levels, $x^1(0)$ and $x^2(0)$, to average-price levels, $x^1(p^*)$ and $x^2(p^*)$ reduces capacity costs by $\beta[x^1(0) - x^1(p^*)]$, shown by areas $S^1 + T + U$. At the same time consumer surplus is reduced by S^1 in period 1 and S^2 in period 2. Thus, as illustrated in Fig. 1a, welfare is increased if $S^1 + T + U > S^1 + S^2$. If the demand curve is linear, $S^1 = T$, and a welfare gain occurs when $S^1 + U > S^2$. In contrast, Fig. 1b shows a relatively more elastic off-peak demand. In this case $S^1 + U < S^2$; average-cost pricing imposes greater welfare losses in the off-peak market than it achieves in net savings in the peak period.

3. Peak-load pricing with a firm peak

Set $p^1 = \beta$, $p^2 = 0$. In period 1, the marginal cost of an additional call is the marginal cost of increasing capacity, β. So long as demand in period 2, at a zero price, is less than period 1 output, the marginal cost of an additional call in period 2 is zero. These rates are optimal. Moreover, because capacity is produced at constant returns to scale, average cost and marginal cost are equal per unit of peak output. Therefore, these marginal-cost prices exactly recover total costs.



(a) Increased welfare: $S^1 + U > S^2$



(b) Reduced welfare: $S^1 + U < S^2$

Fig. 1—Welfare effects of average-cost pricing

4. Peak-load pricing with a shifting peak

Set $p^1 > 0$, $p^2 > 0$, $p^1 + p^2 = \beta$. If the previous rate structure, with $p^2 = 0$, would cause the period 2 demand to exceed period 1 demand the result is a "shifting peak." In this case a positive period 2 price is necessary to equalize demands ($x^1 = x^2$) in both periods. The optimal rates are those that simultaneously (a) bring about this joint peak, and (b) sum to the marginal costs of capacity.

In the joint peak case the marginal cost of a commodity depends on whether its output is increased or decreased. An increase of 1 unit of either x^1 or x^2 requires adding a full unit of capacity and therefore has a marginal cost of β; but a decrease in either output permits no saving in capacity and has a zero marginal cost. However, the optimal prices of the joint peak case may be interpreted as the marginal opportunity cost of output in each period when capacity is fixed.[2] The opportunity cost of supplying a marginal call in period 1 is the value of the most valuable alternative that must be foregone--the withdrawal of one period-1 call worth p^1 from some other subscriber. Similarly, the opportunity cost in period 2 is p^2. And the sum of subscribers' marginal valuations of capacity, $p^1 + p^2$, must equal the marginal cost of expanding capacity, β . Except where explicitly noted below, I assume hereafter that a "firm peak" exists at the rate structures under consideration.

[2]See Hirshleifer, 1958.

This one-exchange model of the peak-load pricing problem yields clear-cut guidelines for ratesetting:

- price should be highest in the period with the maximum demand;
- price should exclude capacity costs in a period that has excess capacity
- optimal prices are equal to marginal costs.

This conventional economic wisdom is an extensive abstraction from the complexities of actual regulated industries. When expanded to include fuel costs, it is perhaps most nearly applicable to the pricing of electricity, an industry in which the bulk of the fixed resources take the form of central generating and transmission capacity which is needed by all consumers.

However, the technology of the telephone industry corresponds less accurately to this paradigm. Instead, capacity is distributed throughout the network in a large number of separate facilities, each of which is available to serve only certain types of calls. To better characterize these aspects of telephone technology, I examine successively more detailed models.

III. SEVERAL ISOLATED EXCHANGES

In each exchange, subscribers place and receive calls only within the exchange. If each exchange has its own rate structure, the pricing problem is that of the previous model. But in practice, a single rate structure must be designed for an entire group of exchanges--for example, all exchanges within one state.[1]

To illustrate this case, it is sufficient to consider just two exchanges, A and B, with demands

$$(6) \qquad x^t_A(p^t_A), \quad x_B(p^t_B) \qquad t = 1, 2$$

and capacities

$$(7) \qquad K_A = \max_t \{x^t_A\}, \quad K_B = \max_t \{x^t_B\} .$$

The total cost of local telephone usage is

$$(8) \qquad C = \beta_A K_A + \beta_B K_B .$$

Restricted Rate Structures

Because the exchanges are grouped the rates must satisfy the restrictions

$$(9) \qquad p^1_A = p^1_B , \quad p^2_A = p^2_B .$$

[1]In discussing this paper William Vickrey points out that such restrictions on the rate structure would be avoided if the telephone company could signal the price to the subscriber at the time he placed his call. Indeed, such dynamic pricing, when combined with equipment to automatically forward one-way messages, promises substantial improvements over static time-of-day rate structures.

Of course, a common rate structure for both exchanges that is based either on flat rates or on an average price that applies in both periods will satisfy these restrictions. These cases are much like those considered for a single exchange.

Optimal restricted peak-load rate structures can be determined by maximizing the welfare function (5) subject to the pricing restrictions (9). In general, the optimal prices are weighted averages of the marginal costs of the individual commodities

(10) $$p^t = [b^t_A/(b^t_A+b^t_B)]\ mc^t_A \ + \ [b^t_B/(b^t_A+b^t_B)]\ mc^t_B$$

where

(11) $$b^t_j = \partial x^t_j/\partial p^t \qquad j = A,\ B$$

It is important to note that the weights for the marginal costs are composed of the slopes of the demand curves, not the number of calls. When a common price must be charged for two commodities with differing marginal costs, some loss of efficiency must result. For example, suppose the common price were set equal to the marginal cost in market A. Then the gap between this price and marginal cost in market B would cause a distortion given by the familiar welfare triangle with area proportional to the slope of the demand curve in that market. Bringing the price closer to mc_B will reduce that loss but create one in market A. The best balance of gain and loss depends on the demand changes in each market, as shown by equation (10).

A key result of restricting the admissible set of rate structures is that the optimal pricing rules can no longer be stated solely in terms of cost data, i.e., set price equal to marginal cost. Instead, as shown in equation (10), demand data, in the form of slopes or elasticities, are commingled into the pricing rule. Two types of peak-load cases need to be considered.

1. Same peak period in each exchange

With maximum demand in period 1 in both exchanges, marginal costs are

$$(12)\qquad \begin{aligned} &mc^1_A = \beta_A, \qquad mc^1_B = \beta_B \\ &mc^2_A = mc^2_B = 0\ . \end{aligned}$$

Thus, the optimal rates are

$$(13)\qquad \begin{aligned} &p^1 = [b^1_A/(b^1_A+b^1_B)]\ \beta_A \ + \ [b^1_B/(b^1_A+b^1_B)]\ \beta_B \\ &p^2 = 0\ . \end{aligned}$$

The requirement that the exchanges be grouped for ratemaking imposes a particular type of data aggregation. Although there are four separate commodities, the admissible rate structure distinguishes only two types of output--total period-1 demand and total period-2 demand (the number of daytime calls and the number of nighttime calls throughout the state). For this case the optimal price for period-1 calls is a weighted average of the per-unit capacity costs in each exchange; in period 2 each exchange has idle capacity and the price is therefore zero.

Because the weights for the capacity costs are the slopes of the demand curves in each period, not the number of calls, this rate structure will not (except by chance) exactly recover total costs when capacity costs vary by exchange. To satisfy the revenue constraint (without resorting to a fixed charge), one or both prices must be adjusted. The best feasible rate structure would modify these prices, taking the demand elasticities in each market at each period into account. As a result, a positive off-peak price could be efficient if demand is relatively inelastic in that period.

2. Peak periods vary by exchange

Suppose that in exchange A the maximum demand occurs in the first period, whereas in exchange B demand is maximal in period 2. In this case marginal costs are

$$mc^1_A = \beta_A, \quad mc^1_B = 0 \tag{14}$$
$$mc^2_A = 0, \quad mc^2_B = \beta_B$$

and the optimal rates are

$$p^1 = [b^1_A/(b^1_A+b^1_B)]\ \beta_A, \qquad p^2 = [b^2_B/(b^2_A+b^2_B)]\ \beta_B\ . \tag{15}$$

In period 1, the price is a fraction of the marginal capacity cost in exchange A. The relationship for period 2 price is similar. In each case the proportions depend on the demand slopes of the commodities in each exchange. Again, these optimal prices will not generally satisfy the budget constraint and the best feasible prices would modify these rates on the basis of demand elasticities.

Quantity-weighted marginal costs

A feasible method of meeting the budget constraint is to construct the prices using quantity weights in place of slope weights in the previous formulas. Let

$$\theta^t_A = x^t_A/(x^t_A+x^t_B), \qquad \theta^t_B = x^t_B/(x^t_A+x^t_B) = 1 - \theta^t_A \qquad t = 1, 2 \tag{16}$$

be the proportions of the grouped outputs that occur in each exchange in period t. For case 1 (same peak period) set

$$(17)\qquad p^1 = \theta^1{}_A\beta_A + \theta^1{}_B\beta_B, \quad p^2 = 0 .$$

For case 2 (different peak periods) set

$$(18)\qquad p^1 = \theta^1{}_A\beta_A, \qquad p^2 = \theta^2{}_B\beta_B .$$

These rates, based only on quantity information, can be given an informative interpretation in terms of suitably defined marginal costs.

Marginal cost of a group of commodities

When commodities are grouped it is not immediately apparent just what the "marginal cost" of the aggregate is. To define its marginal cost we must specify how each of the components of the group changes when the group itself changes by one "unit."

One plausible definition is to specify that the quantities of each commodity in the group vary proportionately. Thus for a change dx^t in the group quantity let the components change by

$$(19)\qquad dx^t{}_A \equiv \theta^t{}_A dx^t, \qquad dx^t{}_B \equiv \theta^t{}_B dx^t .$$

The marginal cost of the grouped output in period t is then

$$(20)\qquad mc^t = \partial C/\partial x \equiv \theta^t{}_A mc^t{}_A + \theta^t{}_B mc^t{}_B \qquad t = 1, 2 .$$

Thus, the group marginal cost is defined as the quantity-weighted average of the individual commodity marginal costs.

With constant returns to scale, the rates (equation (17) or (18)) based on this measure of marginal costs are feasible. And in one special case they will be optimal--when the individual commodities that make up a group have the same elasticities of demand. To see this, write the equation (15) for the optimal restricted rates in terms of elasticities

$$(21) \qquad p^t = \frac{\eta^t_A x^t_A}{\eta^t_A x^t_A + \eta^t_B x^t_B} \, mc^t_A + \frac{\eta^t_B x^t_B}{\eta^t_A x^t_A + \eta^t_B x^t_B} \, mc^t_B$$

When $\eta_A = \eta_B$, the weights for the terms mc^t_A and mc^t_B are just the quantity weights θ^t_A, θ^t_B. In this case, commodities are homogeneous in terms of demand, and the optimal pricing rule requires only cost data.

IV. A NETWORK OF EXCHANGES

Each exchange has intra-exchange calling as in the previous model. In addition, there are inter-exchange (AB) calls which make use of capacity in the originating and terminating exchanges and also require a third capacity component--trunking facilities that connect the exchanges. The key feature of this model is the introduction of joint production, which occurs when local exchange switching capacity is shared by two different commodities.

Demands are

$$x^t_A = x^t_A(p^t_A)$$

$$\text{(22)} \qquad x^t_B = x^t_B(p^t_B)$$

$$x^t_{AB} = x^t_{AB}(p^t_{AB}) .$$

The capacity constraints are

$$K_A = \max \{x^t_A + x^t_{AB}\}$$

$$\text{(23)} \qquad K_B = \max \{x^t_B + x^t_{AB}\}$$

$$K_{AB} = \max \{x^t_{AB}\}$$

where I assume each inter-exchange call requires the switching capacity of an intra-exchange call in each of the two exchanges as well as inter-exchange trunking. Total costs are then

$$\text{(24)} \qquad C = \beta_A K_A + \beta_B K_B + \beta_{AB} K_{AB} .$$

The optimal prices are obtained from the Kuhn-Tucker conditions of the mathematical program.[1] The prices are

$$p^1_A = \mu^1_A, \qquad p^2_A = \mu^2_A$$

$$(25) \qquad p^1_B = \mu^1_B, \qquad p^2_B = \mu^2_B$$

$$p^1_{AB} = \mu^1_A + \mu^1_{AB}, \qquad p^2_{AB} = \mu^2_B + \mu^2_A$$

where μ^t_j is the dual variable in period t for capacity of type j. The central result is that even when there are as many prices as commodities, the technological interdependence of the separate markets destroys the simple correspondence between maximum demands and maximum prices. However, with firm peaks, the optimal prices are equal to the marginal costs of the individual commodities.

For example, suppose that exchange A and inter-exchange calls are day peaking ($x^1_A > x^2_A$ and $x^1_{AB} > x^2_{AB}$) and exchange B is night peaking ($x^1_B < x^2_B$). The optimal prices will be

$$p^1_A = \beta_A, \qquad p^2_A = 0$$

$$(26) \qquad p^1_B = 0, \qquad p^2_B = \beta_B$$

$$p^1_{AB} = \beta_A + \beta_{AB}, \qquad p^2_{AB} = \beta_B \;.$$

Thus the inter-exchange calls should pay positive prices in both periods, not only in their peak (t=1) period. Moreover, despite the

[1]See, e.g. Littlechild, 1970.

fact that AB calling is highest in period 1, p^2_{AB} could mathematically exceed p^1_{AB}, although this is unlikely in the particular example of local and inter-exchange calls.

Restricted rate structures

Here we reach the "realistic" case for telephone ratemaking. In practice, rates might well be restricted to be the same for all intra-exchange calls at a given time of day throughout a region or state, with separate rates applying for inter-exchange calls. Frequently, however, the same "off-peak" percentage discount is applied to both types of calls. In this case the restrictions are

$$\text{(27)} \quad \begin{aligned} & p^t_A = p^t_B, \quad t = 1, 2 \\ & p^2_{AB} = \lambda p^1_{AB} \quad \text{where} \quad \lambda = p^2_A/p^1_A = p^2_B/p^1_B \ . \end{aligned}$$

Effectively, there are three rate parameters--the mean levels of the intra-exchange and inter-exchange rates and the percentage discount in the off-peak period.

In principle, the mathematical program for the welfare-maximizing network prices can be solved for any specified constraints on the rate structure. For a small problem--such as this example--this is quite feasible. But for realistic situations, the dimensions of the problem are substantially greater. M, rather than two exchanges, must be considered. Because the rate of demand varies over both the daily and weekly cycle, N distinct periods must be analyzed. And there are several levels of inter-exchange calls, conventionally grouped according to distance bands.

V. EVALUATING THE EFFICIENCY OF TELEPHONE RATE STRUCTURES

A practical approach is to use the structure of the demand and cost model to evaluate the welfare effects of alternative rate structures without attempting to achieve a global optimum. This approach should be undertaken at two levels.

1. A Given Rate Structure

A particular rate structure specifies a grouping of commodities into time periods, distance bands, and perhaps geographic areas. The quantity-weighted marginal costs of each grouped output can be calculated by proportionately incrementing demands of each commodity in the group. (For example, if the peak period is 8 a.m. - 5 p.m., weekdays, the traffic load curve at those hours can be increased by a constant percentage). By calculating the "Ramsey number" of each group k at current prices and output levels

$$R_k = \eta_k(p_k - mc_k)/p_k \qquad (28)$$

the group marginal costs can be compared with prices.[1] If rates are optimal (given the rate structure), all of the Ramsey numbers will be equal. If not, welfare can be increased by raising rates for groups with low Ramsey numbers and reducing rates for high R_k values.

2. Alternative Rate Structures

Some redesigning of the rate structure may yield welfare gains at least as large as those achievable by adjusting rate levels. Two closely related questions must be investigated--the number of different

[1]See Willig and Bailey, 1977.

prices to charge, and the particular commodities to be included in each group. For example, local telephone rates might be limited to two price levels throughout the week, with the particular hours that peak prices apply determining which telephone calls are grouped together.

Guidance for grouping commodities is provided by two results from the earlier analysis:

1. If, within each group, all commodities have the same marginal cost, then a group price equal to the common marginal cost will be (first-best) optimal. This will be true even if the commodities have different elasticities of demand.
2. If, within each group, all commodities have the same elasticity of demand, then price should be equal to the quantity-weighted average of the commodity marginal costs.

For the telephone network, the marginal costs of several commodities will be similar when they (a) have the same peak period, and (b) use equipment that has similar unit capacity costs. As for demand elasticities, they will perhaps be similar when exchanges are grouped by type of customer.

These general considerations suggest that efficient grouping will be promoted by combining commodities according to similarities in both marginal costs and demand elasticities. For example, alternatives to a proposed 8 a.m. - 5 p.m. peak period could be considered by comparing both the demand elasticities and marginal costs at, say, 6, 7, and 8 p.m. with those in earlier hours. Hours that clearly follow the earlier elasticity and cost pattern readily suggest an extended period for the time-of-day rate structure. A mixed pattern of elasticities and costs, however, would require evaluating different combinations of grouped hours.

To evaluate a change in the number of prices in the rate structure additional data are needed. Practical restrictions on the number of separate rates are presumably due to the "transactions costs" the subscriber must bear to cope with an increasingly detailed structure of rates, and the additional administrative complexity for the telephone company of calculating, defending and revising such rates. Measurements of transaction costs are not readily available. However, one can demonstrate the size of the efficiency gain that could be realized by adding an additional rate, or on the other hand, the efficiency cost of simplifying the rate structure in a specified manner. These values can then be compared to subjective assessments of the hassle of coping with rate structures of differing complexity.

VI. SUMMARY

The design of appropriate rate structures for local telephone calls should be determined by the technology and cost characteristics of the local network. Apart from the equipment dedicated exclusively to serve each subscriber, nearly all of the costs of local telephone service are due to providing capacity sufficient to meet maximum demands. Thus, some form of peak-load pricing is desirable. A uniform average-cost price at all hours may be less efficient than a flat-rate tariff which charges nothing per call, even if metering were costless.

Switching and trunking capacity is distributed throughout the network and jointly used by different types of calls. As a result, optimal prices may be positive when demand is below the maximum level, and the highest rate need not occur at the hour of peak demand.

Realistic rate structures can have only a limited number of separate prices, requiring that individual commodities be aggregated into groups. An efficient rate structures will combine hours and routes that have similar marginal costs and demand elasticities.

THE ESTIMATION OF USAGE REPRESSION
UNDER LOCAL MEASURED SERVICE
EMPIRICAL EVIDENCE FROM THE GTE EXPERIMENT

G.F. WILKINSON

G.T.E.

INTRODUCTION

One of the major changes forthcoming in the pricing of local telephone service in the United States is the conversion to some form of usage sensitive pricing. Presently, most consumers within the U.S. pay a single flat rate charge each month for which they can make an unlimited number of calls within a specified local calling area. Currently, the Bell System and the major independent telephone companies are pursuing plans to convert, at least portions of their serving territory, to measured service.[1/]

In order to learn about the impact of converting to measured service both upon the consumer and company operations, the GTE system initiated a measured service experiment in three exchanges in central Illinois in 1975. Data have been gathered since May, 1975 on consumer telephone usage, and these data have been used as the basis for a number of papers on measured service. In particular, Jensik (1979) and Park (1981) had similar research objectives and each used time-series data from the same experimental exchanges.[2/] This paper uses models of the same type used by Jensik to elaborate further on the changes in consumer usage under measured service.

THE MEASURED SERVICE EXPERIMENT AND THE TARIFFS

The GTE measured service experiment is being conducted in the Jacksonville, Clinton, and Tuscola exchanges in Illinois. 3/ These exchanges were chosen as the study area since they are reasonably representative of a large portion of GTE serving territories. Since May, 1975, data have been gathered on calls and minutes of usage by class of service (residence one-party, residence multi-party, business one-party, key trunks, and PBX trunks). From May, 1975 until August, 1977, flat rate tariffs were in effect in these exchanges.

In September, 1977, residence one-party, business one-party, and key and PBX trunks were converted to measured tariffs. In the Jacksonville Exchange, the initial rates were structured to include a fixed charge each month (approximately 40% of the former flat rate) and a charge for each call and minute of usage. In the Clinton and Tuscola exchanges the initial measured service tariff included only a charge per minute and a fixed monthly charge. These usage tariffs were designed to charge the same amount for a four minute call in all of the experimental exchanges (which is approximately the average length of a residence call). In 1979, usage rates were increased and the tariff structures in the three exchanges were made the same. Table 1 illustrates the changes in the residence one-party rates throughout the experiment. Business one-party, key, and PBX usage rates have been and continue to be the same as the residence usage rates although the fixed monthly charges for these classes of service are higher.

Since measured service was implemented, there has been a 20% discount on all usage between 5 and 11 p.m. daily and all day Sunday, and a 50% discount on all usage between 11 p.m. and 8 a.m. No distance sensitivity was included in the tariffs.

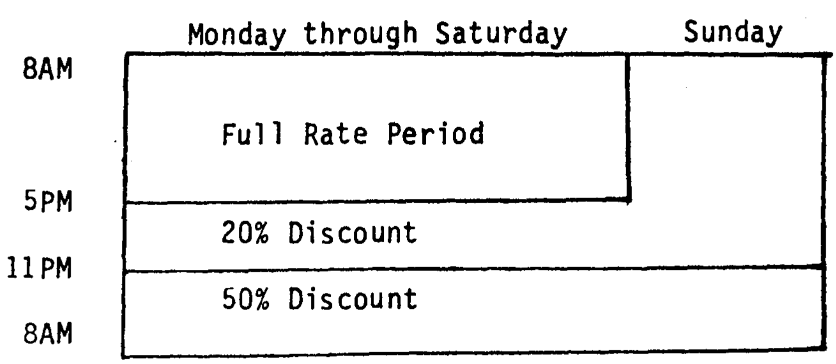
Table 1

Tariffs for the Measured Service Experimental Exchanges

Residence One-Party

	Flat Rate (5/75 - 9/77)	Measured Rate (9/77 - 6/79)		Measured Rate (6/79 - Present)	
		Monthly Charge	Usage Charge	Monthly Charge	Usage Charge
J'Ville	$7.95	$3.15	2¢/call 1¢/min.	$3.15	2.5¢/call 1¢/min.
Clinton	6.20	2.50	1.5¢/min.	2.50	2.5¢/call 1¢/min.
Tuscola	5.90	2.50	1.5¢/min.	2.50	2.5¢/call 1¢/min.

TIME-OF-DAY PRICING PERIODS



STUDY METHODOLOGY

The intent of this study was to estimate changes in usage characteristics as a result of measured service tariffs for the three experimental exchanges. The primary focus was on residence one-party service although models for business one-party service were also developed and used for comparisons. For research of this type, the ideal circumstance would have been to compare the telephone usage of control groups having very similar demographic and economic characteristics which did not experience the changes in local service tariffs with the usage of the exchanges which had been converted to measured service. However, because no equivalent control groups were established for the experimental exchanges (primarily due to the high cost of data gathering and processing), it was necessary to infer the effects of the tariff changes by estimating deviations from the historical trend which had been established prior to the change in rates. The particular data series which were analyzed in this manner were the monthly calls and minutes per mainstation.

The time-series analysis techniques developed by Box and Jenkins [4/] were chosen to identify the changes in usage characteristics. These Box-Jenkins models are particularly well suited for this type of analysis because of the seasonal characteristics of the data. For these time series, transfer function models using intervention terms were used to estimate the changes. Ideally, transfer functions using economic variables to explain the changes in the behavior of the series would provide better indicators for understanding the series. However, economic time series data which adequately describe the dynamics of consumer telephone usage in these exchanges have not been found and thus it was necessary to rely on intervention analysis.

MODEL STRUCTURE

The models which have been developed for usage per mainstation employ very simple intervention structures. In building intervention models, it is necessary to hypothesize the nature of the change attributable to a tariff (such as a permanent change in number of calls per main per month due to a price change, i.e. a step function) and then test whether the actual behaviour of the time series supports this hypothesis. (In contrast to the identification of the transfer function structure based upon cross correlation functions.) The models used in this analysis can be viewed as having an intervention component which describes step changes in the series due to tariff changes or pulses which account for unusual events, a noise component which describes all other systematic behaviour of the series, and a random error.

In order to briefly explain the model structures, let us first examine the noise component of the models. This noise component can be viewed as being the univariate model of the time series having already accounted for the effects of the interventions. Like all univariate models, the noise component describes the behaviour of the series based upon its past history using autoregressive or moving average operators (or both). Prior to identifying the types of operators which are needed and estimating their values, it may be necessary to difference and transform the series to make it stationary.

The noise structures for most of the models used in this study are of the form:

$$\nabla^{1}\nabla^{12} \ln Y_t = (1-B)(1-B^{12})a_t$$

In this model, the usage per mainstation series (Y_t) was transformed by natural logarithms (ln) and has been differenced using first order and seasonal operators $(\nabla^1 \nabla^{12})$. This transformed and differenced series which is stationary is described by a first order and seasonal moving average process $(1-B)(1-B^{12})$. The forecast errors are the a_t's.*

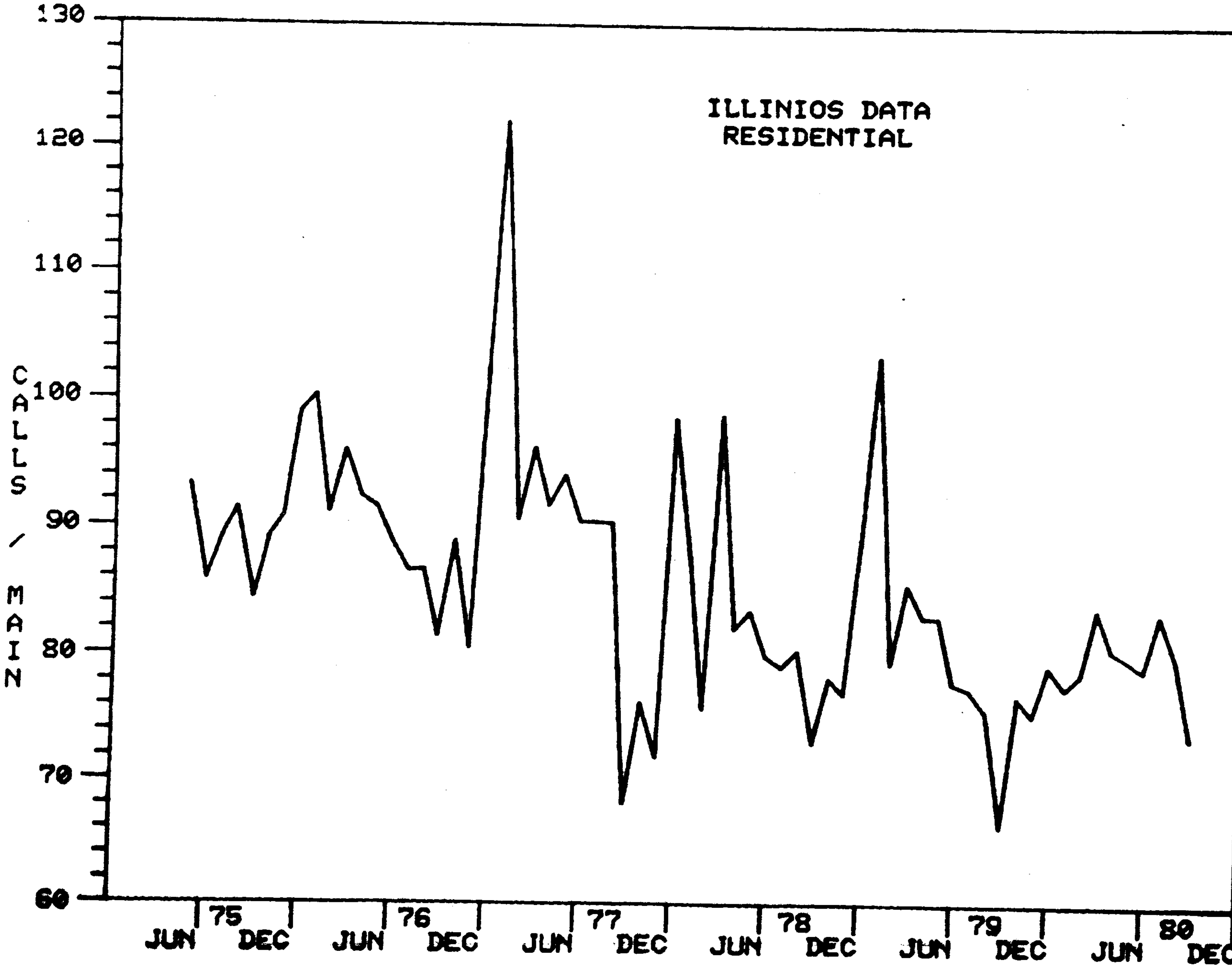
The intervention component of the models are terms which describe changes in the level of the series (steps) or unusual events (pulses). Like dummy variables in standard regression equations, these interventions are 1's when the intervention is active and 0 everywhere else. The form of the intervention structure for the models used in this study are simply $Y_t = \omega_0 \xi_t$, where ω_0 is an estimated parameter which describes the transfer function between the intervention term (ξ_t) and the usage series (Y_t).

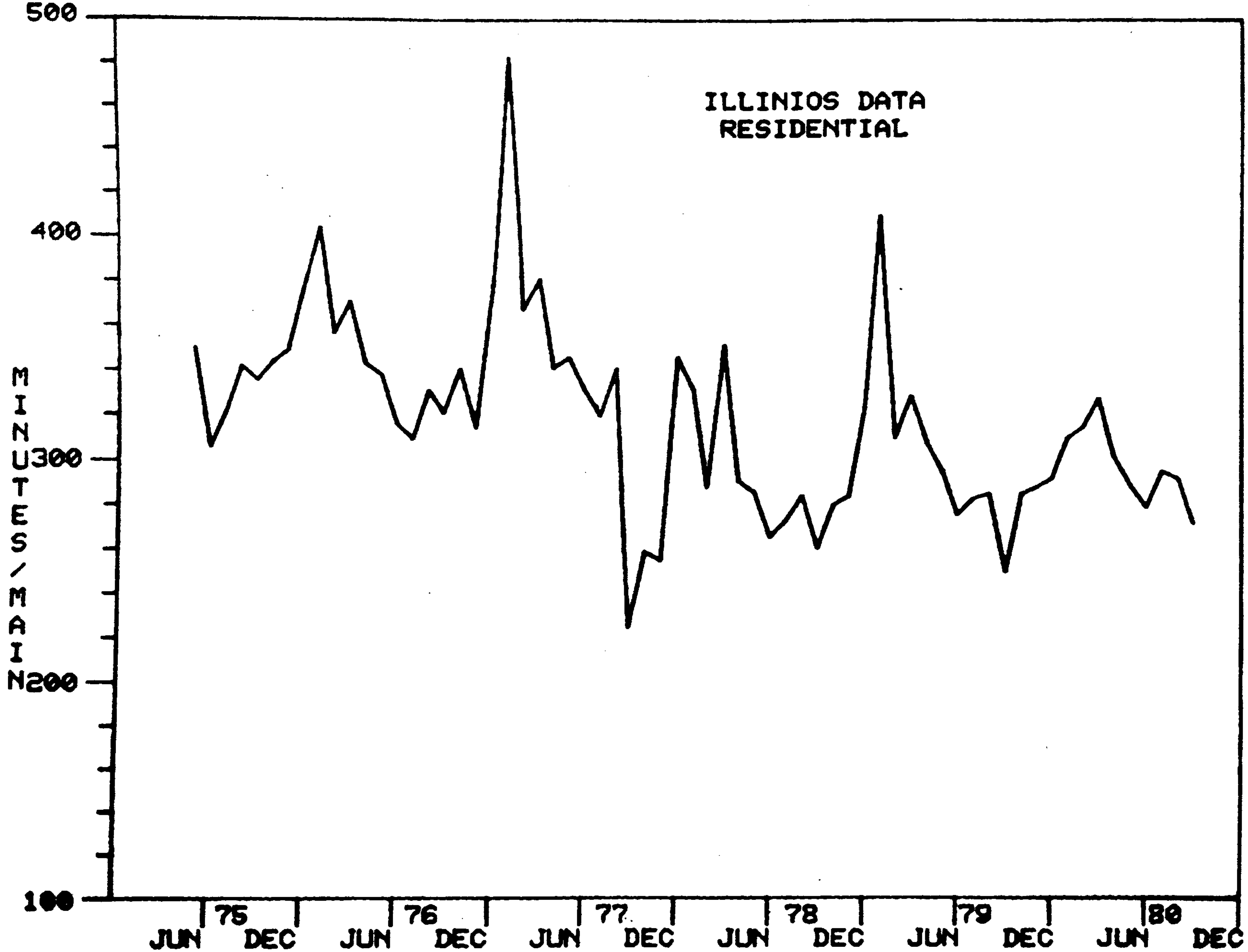
EMPIRICAL RESULTS

Initially univariate models were developed for the residence calls per mainstation and minutes per mainstation series for the aggregate of the three exchanges. (See Figure 1 and 2.) The models were fit with 65 monthly observations from May, 1975 to September, 1980. These univariate models can be very helpful in understanding the behaviour of the series and developing hypotheses for the interventions.

---

* Two operators are used in Box-Jenkins nomenclature for writing these models in a compact form: they are the backshift operator B for shifting backward in time (i.e. $(B)Z_t = Z_{t-1}$) and a differencing operator which indicates a new series has been created by differencing the original series according to the degree of the operator (i.e. $\nabla = Z_t - Z_{t-1}$) See Box and Jenkins (1970) for a full explanation of these operators and the structure of the models.


ILLINIOS DATA
RESIDENTIAL
CALLS / MAIN
130
120
110
100
90
80
70
60
JUN
75
DEC
JUN
76
DEC
JUN
77
DEC
JUN
78
DEC
JUN
79
DEC
JUN
80
DEC


ILLINIOS DATA
RESIDENTIAL
MINUTES/MAIN
500
400
300
200
100
75
76
77
78
79
80
JUN DEC JUN DEC JUN DEC JUN DEC JUN DEC JUN DEC

The largest residuals from these univariate models were associated with the change to measured service or unusual weather conditions which were known to affect calling patterns. Although no large residuals occurred following rate change in June, 1979, an intervention term for the rate case was tested for significance. Using the residuals from these univariate models as guides, transfer function models were estimated using the same time series data. The interventions for rate changes were modelled as step functions and the interventions for the months with excessively bad weather were modelled as pulses in a single intervention variable. In the formulas, the intervention terms are designated as ξ_1 for a step function beginning in September, 1977 when measured service was introduced, ξ_2 for pulses at January, 1977 and March, 1978 for unusually severe winter weather, and ξ_3 as a step function beginning in June, 1979 for the increase in the measured rates.

The transfer function model for calls per residence mainstation was the following:

$$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.03}{-.16\xi_{1t}} \underset{\pm.03}{+.17\xi_{2t}} \underset{\pm.03}{-.08\xi_{3t}} + (1-\underset{\pm.08}{.83}B)(1-\underset{\pm.07}{.75}B^{12})a_t$$

This model indicates that there has been a decline of about 15% in calls per mainstation following the introduction of measured service. * The severe weather during January, 1977 and March, 1978 accounted for an average increase

---

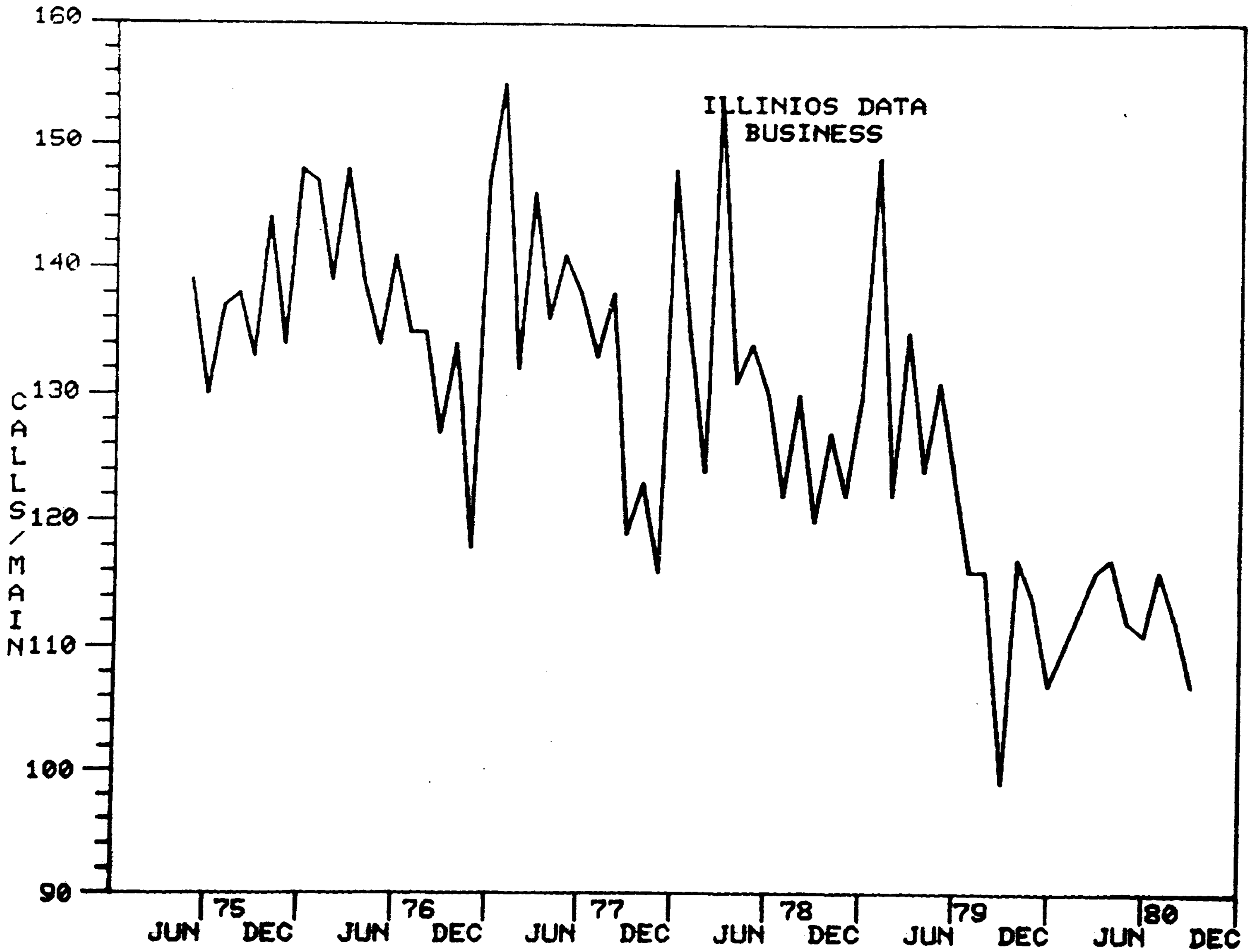
* Calculated as the complement of e to the exponent of the coefficient.

of 19% during these months. The decline of about 8% on the third intervention is difficult to interpret. It is tempting to attribute this decline in calling entirely to the June, 1979 rate increase, however, I think this would be incorrect. While the previous interventions do not appear to have other variables significantly affecting the estimate, an alternate hypothesis that the decline in calling is due to the recent recession could be formulated for this third intervention term. In order to test this intervention, the same model was estimated with only 55 data points, and the ξ_3 term was insignificant (-.05 ±.03). Further evidence that the recession was a contributing factor in this intervention coefficient were large negative residuals in the Spring and Summer months of 1980, over 9 months after the new rates were in effect (in the original 65 observation model). Thus it is probably more reasonable to attribute this 8% decline in calling to the combination of rate increase and recession effects. Since this coefficient is within two and three standard errors, it would be premature to draw strong conclusions about its effect upon the time-series.

The model for business one-party calls per mainstation indicated that businesses reacted differently than residences in their response to measured services. (See Figure 3 for a graph of this series). The model for this series was the following:

$$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.03}{-.04}\xi_{1t} \underset{\pm.05}{+.08}\xi_{2t} \underset{\pm.03}{-.11}\xi_{3t} + (1-\underset{\pm.06}{.88}B)(1-\underset{\pm.08}{.70}B^{12})a_t$$

For business calls per mainstation, there was not a significant reduction in calls per main due to measured service or an increase in calling due to the unusual weather for January, 1978 and March, 1978. However, the coefficient on the ξ_3 intervention is significant. This is probably due to a combination of rate effects and the recession.



ILLINIOS DATA
BUSINESS
CALLS/MAIN
160
150
140
130
120
110
100
90
75
76
77
78
79
80
JUN
DEC
JUN
DEC
JUN
DEC
JUN
DEC
JUN
DEC
JUN
DEC

Looking at the minutes per main series, the residence model exhibits the same structure as the calls per main model. The model was the following:

$$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.04}{-.26}\xi_{1t} \underset{\pm.04}{+.14}\xi_{2t} \underset{\pm.04}{-.05}\xi_{3t} + (1-\underset{\pm.08}{.78}\xi)(1-\underset{\pm.06}{.79}B^{12})a_t$$

The model indicates that residence minutes per main declined by about 23% due to the implementation of measured service. The weather increased minutes per mainstation by about 15%, and the third intervention (again rate case and recession related) indicated a decline in minutes per main of about 5% (this coefficient is insignificantly different from zero at two standard errors).

The business minute per main per month model (see Figure 4 for a graph of this series) was the following:

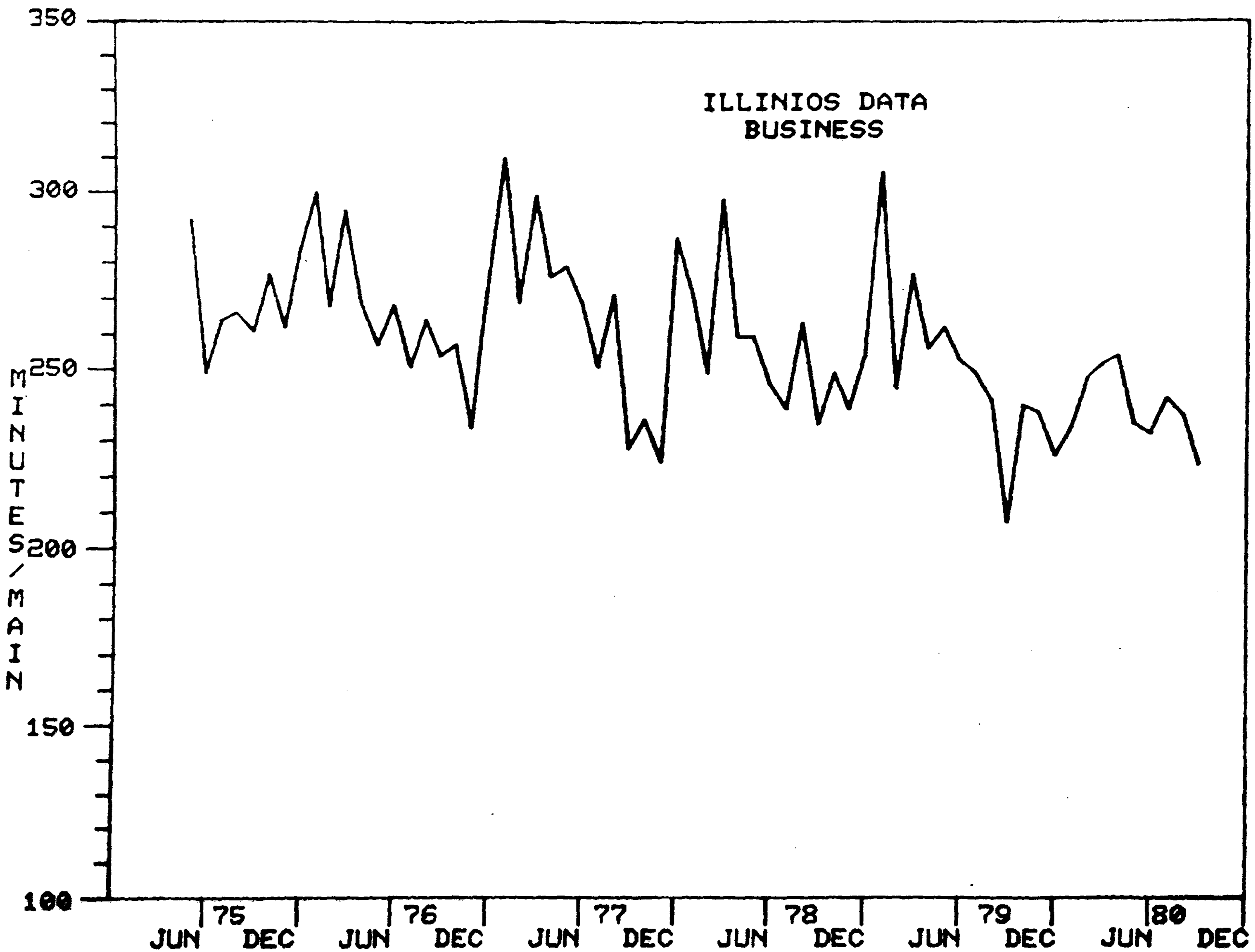
$$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.03}{-.05}\xi_{1t} \underset{\pm.05}{+.06}\xi_{2t} \underset{\pm.03}{-.05}\xi_{3t} + (1-\underset{\pm.06}{.88}B)(1-\underset{\pm.07}{.72}B^{12})a_t$$

This model is very similar to the business calls per main model with the exception of the coefficient on the ξ_3 intervention term which, unlike the calls model, is insignificant. This indicates that business calls per mainstation have been reduced to a somewhat greater degree than minutes per main as a consequence of the rate case/recession influence.

The previous models were developed using aggregate data for the three experimental exchanges. In order to examine whether residence usage was similar among the experimental exchanges separate residence minutes per mainstation models were developed.

Jacksonville: $$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.03}{-.24}\xi_{1t} \underset{\pm.05}{+.15}\xi_{2t} \underset{\pm.03}{-.04}\xi_{3t} + (1-\underset{\pm.08}{.80}B)(1-\underset{\pm.07}{.78}B^{12})a_t$$

Clinton: $$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.04}{-.25}\xi_{1t} \underset{\pm.07}{+.15}\xi_{2t} \underset{\pm.04}{-.06}\xi_{3t} + (1-\underset{\pm.06}{.86}B)(1-\underset{\pm.06}{.81}B^{12})a_t$$


ILLINIOS DATA
BUSINESS
MINUTES/MAIN
350
300
250
200
150
100
75
76
77
78
79
80
JUN
DEC
JUN
DEC
JUN
DEC
JUN
DEC
JUN
DEC
JUN
DEC

Tuscola:

$$\nabla^1\nabla^{12}\ln Y_t = -.26\xi_{1t} + .21\xi_{2t} - .07\xi_{3t} + (1-.80B)(1-81B^{12})a_t$$

$$\pm.05 \quad \pm.07 \quad \pm.04 \quad \pm.09 \quad \pm.06$$

It is evident from these models that the reaction to measured service, winter storms, and the rate case/recession effects was similar among the exchanges. All coefficients for any exchange are within one standard error of the respective coefficients for the others.

The next models identified customer reaction to changes in the tariff structure by discount period. Recall from the tariff table that a 20% discount was applied to usage charge between 5 and 11 p.m. each day and all day Sunday, and a 50% discount was applied to usage between 11 p.m. and 8 a.m. These discount periods are designated as P1 for no discount, P2 for a 20% discount, and P3 for a 50% discount. The following models were developed for residence minutes per mainstation within each discount period:

P1

$$\nabla^1\nabla^{12}\ln Y_t = -.28\xi_{1t} + .15\xi_{2t} - .06\xi_{3t} + (1-.62B)(1-68B^{12})a_t$$

$$\pm.06 \quad \pm.05 \quad \pm.07 \quad \pm.12 \quad \pm.13$$

P2

$$\nabla^1\nabla^{12}\ln Y_t = -.27\xi_{1t} + .14\xi_{2t} - .03\xi_{3t} + (1-.73B)(1-.66B^{12})a_t$$

$$\pm.04 \quad \pm.04 \quad \pm.06 \quad \pm.12 \quad \pm.13$$

P3

$$\nabla^{12}\ln Y_t = -.14\xi_{1t} + .23\xi_{2t} - .07\xi_{3t} + \left(\frac{1-.66B^{12}}{1-.59B}\right)a_t$$

$$\pm.05 \quad \pm.06 \quad \pm.08 \quad \text{numerator } \pm.13, \text{ denominator } \pm.11$$

The models for the P1 and P2 pricing periods show that the decline in minutes per mainstation (about 24%) due to the introduction of measured service was virtually the same, in spite of the 20% discount on usage in P2. The noise model for P3 usage was structured a little differently than the previous noise models using an autoregressive operator and only seasonal differencing. It indicates that a 13% decline in minutes per mainstation occurred due to the introduction of measured service. Models for business minutes per mainstation were also estimated by pricing period. Although the coefficients were very similar between the pricing period models, such a large percentage of the traffic was concentrated in the P1 period, that the P1 model was the only model which was meaningful. The business P1 model had the same coefficients as the overall model.

The final phase of this research identified a model which was somewhat better than the step function for explaining the effects of the introduction of measured service. Although it might be expected that the effects would gradually increase over a few months, the contrary occurred. A model which had a lower residual variance compared with the step function had a structure for the intervention which hypothesized an over-reaction to measured service. In this formulation, the model indicates that there was approximately a 29% reduction in minutes per mainstation during the first month of conversion, but from the second month onward, there has been only a 21% reduction in usage.

$$\nabla^1\nabla^{12}\ln Y_t = \underset{\pm.05\ \ \pm.05}{(-.36\ +.13)}\xi_{1t} \underset{\pm.04}{+.14}\xi_{2t} \underset{\pm.03}{-.05}\xi_{3t} + (1-\underset{\pm.07}{.82}B)(1-\underset{\pm.06}{.76}B^{12})a_t$$

Unfortunately, this particular model had a very high correlation between the ω_0 and ω_1 terms on the intervention for measured service. This high correlation indicates that there are probably a number of combinations of parameter estimates very close to these maximum likelihood estimates. Therefore, these results are suspect.

CONCLUSION

For these particular exchanges and tariffs, the introduction of measured service resulted in a decline in calls per residence mainstation of about 15% and for minutes per residence main, about 23%. Usage reductions (both calls and minutes per main) due to measured service for business one-party were not significant. (Although the models have not been presented here, Key and PBX usage also did not decline significantly due to the introduction of measured service). For residence, the evening discount did not significantly influence the overall impact on usage of the conversion to measured service. In fact, the coefficient indicated that approximately the same percentage repression occurred during P2 as during the daytime period. It appears that usage has declined further due to a combination of the June, 1979 rate and the current economic recession. It was not possible to isolate each effect independently in this analysis, although further work is being done in this area. Another interesting result is that severe weather significantly affects residence usage, in this case, increasing calls per mainstation by about 15%.

Overall, it seems that intervention modelling can be a useful tool for identifying the effects of price changes upon usage. The technique produces reasonable parameter estimates of usage repression. It is especially appropriate for telephone usage data because of its seasonal nature.

FOOTNOTES:

1/ See for example Garfinkel, L., and Linhart, P. B., "The Transition to Local Measured Service," Public Utilities Fortnightly, 104, Aug 1979 and Schmidt, L. W., "Local Measured Service: A Telephone Industry Perspective," in J. A. Baude, et. al. eds. Perspective on Local Measured Service, Kansas City: Telecommunications Industry Organizing Committee, 1979.

2/ Jensik, John M., "Dynamics of Consumer Usage", in J.A. Baude, et. al. eds. Perspectives on Local Measured Service, Kansas City: Telecommunications Industry Workshop Organizing Committee, 1979; and Park, R.E. and Wetsel, Bruce M., "Charging for Local Telephone Calls: Price Elasticity Estimates from the G.T.E. Illinois Experiment", The Rand Corporation, R-2635-NSS, 1981.

3/ For a more extensive description of the GTE Measured Service Experiment see G. Cohen, "Usage Sensitive Pricing", Fifth Annual Rate Symposium on Problems of Regulated Industries, Columbus, Mo. 1979.

4/ Box, G.E.P. and Jenkins, G.M., Time Series Analysis: Forecasting and Control, San Francisco: Holden Day, 1970.

LOCAL TELEPHONE PRICING: TWO-PART TARIFFS AND PRICE DISCRIMINATION

JAMES A. BRANDER
Queen's University

BARBARA J. SPENCER
Boston College

Abstract: This paper examines the optimal structure of two-part tariffs for a profit constrained public utility that can discriminate among different consuming groups. First, all groups are assumed to be final consumers and a two-stage rule characterizing optimal markups is derived. However, the important form of discrimination in telephone pricing is between final consumers and firms who use telephone service as a productive input. The paper shows that if downstream industries are perfectly competitive, the structure of optimal pricing rules is unchanged. If, however, downstream industries are imperfectly competitive, efficiency criteria imply that markups should, other things equal, be lower for these industries.

Local Telephone Pricing: Two-part Tariffs and Price Discrimination

by

James A. Brander

and Barbara J. Spencer

Introduction:

Local telephone service in Canada is priced in a rather unusual way. Specifically, consumers pay a fixed monthly fee for service and then make as many local phone calls as they wish at no extra cost: expenditure is insensitive to usage. Whether or not such a policy ever was optimal, it seems unlikely to persist given recent technological advances: the cost of monitering usage is very much lower than it was, and for large exchanges is now "small" compared with the production cost of a phone call, and the increasing use of telephone lines for computer transmissions implies that certain users will use the system so heavily that serious inefficiency may arise if usage sensitive prices are not set.[1]

From a regulator's point of view telephone companies are classic public utilities. The regulator's objective is to encourage efficient and equitable provision of telephone service while allowing a certain profit to the telephone company. This budget or profit constraint implies that, if production is being carried out under increasing returns to scale, departures from marginal cost pricing are required.

In considering departures from marginal cost pricing one principle that has emerged is the Ramsey principle: efficiency criteria imply that markups over marginal cost charged different groups or for different products should be related to elasticities of demand. Specifically, low elasticities should be associated with high markups. (Classic references on the Ramsey principle are Ramsey (1927) and Boiteux (1956). More recent work includes Baumol and Bradford (1970) on the multiproduct case, and Hartwick (1978) on price discrimination among groups consuming a single product).

An important natural question that arises concerns whether Ramsey pricing is appropriate in the empirically relevant cases. Before this question can be addressed, however, it is necessary to understand the structure of Ramsey prices in the relevant cases. One might object to the Ramsey principle either if the informational requirements are very high or if the structure of Ramsey prices implies unacceptable transfers among consumer groups. This paper, however, is concerned with the first step: characterizing the structure of Ramsey optimal prices.

In this paper we examine optimal two-part tariffs for a public utility that is capable of discriminating among different types of consumers. In particular we focus on the problem faced by a public enterprise like a telephone company, whose output is purchased both by consumers for final consumption and by firms as an input to production. Two part tariffs are pricing structures that consist of an entry or license fee that must be paid before any consumption takes place, and a constant per unit price. Standard references on two part tariffs are Gabor (1955) and Oi (1971). See also Feldstein (1972) Ng and Weisser (1974) and Faulhaber and Panzar (1977). Two part tariffs can be thought of as special case of "nonuniform" or "nonlinear" or "quantity dependent" prices. (See Spence (1977, 1980), Willig (1978), and Mirman and Sibley (1980) for general treatments of nonuniform pricing.)

There are several important features of the telephone industry that we abstract from. The cyclical pattern of demand for telephone service over the day and week is particularly important. Also the multi-product aspect of telephone service and the existence of consumption externalities among consumers have generated considerable interest. These issues are set aside in this paper. As for general background to local telephone pricing, good references are Mitchell (1978) and Baude et. al. (1979).

The optimal pricing rule for the case in which there are many groups of consumers is shown to be a two-stage Ramsey rule. The first term is the usual Ramsey term and in addition there is a correction term depending on elasticities with respect to the license fees. The same rule applies even when some users are firms who use the output of the utility as an input to further production, provided these firms are perfect competitors. If downstream firms are imperfectly competitive efficiency considerations imply that they should generally be charged a lower price for inputs, other things equal.

The problem is to maximize the sum of producers' and consumers' surplus subject to the constraint that the public utility earn a certain minimum profit. Different groups of consumers can be identified but each group is internally hetereogeneous. The utility may use discriminating two-part tariffs; that is, each group can be charged a different two-part tariff. The two-part tariff for group i consists of a license fee L^i and a price P^i. Prices greater than marginal cost will mean that consumers who use the system will underconsume from a welfare point of view. On the other hand, positive license fees may exclude consumers whose consumption is socially desirable. The optimum will generally involve a mix of these two distortions. In addition, the different groups will be treated differently.

There are n groups of consumers, indexed by the letter i, $i = 1, \ldots, n$. Within each group consumers vary, but each consumer is identified as belonging to a particular group. For example, consumers may be identified by the community in which they live. The two-part tariff facing group i is denoted (P^i, L^i). For any given tariff, group i can be divided into two subgroups: those who choose to consume, referred to as "members", and those who do not.

Let M^i = members in group i

S^i = surplus of members in group i

$$S = \sum_{i=1}^{n} S^i$$

We use surplus (areas under demand curves) as a measure of consumer welfare. The surplus measure is exact if consumer demands arise from utility functions of the form $U = u(x) + v$ where v is income spent on other goods. Alternatively, one can appeal to the approximation results of Willig (1976).

The good produced by the public utility is denoted X, and the consumption by group i is denoted X^i. Using π to denote profit,

$$\pi(P,L) = \Sigma P^i X^i + \Sigma L^i M^i - C(X(P,L))$$

where $P = (P^1, \ldots, P^n)$ and $L = (L^1, \ldots, L^n)$

Efficient pricing rules are found by maximizing the sum of producer and consumer surplus subject to the constraint that profits be non-negative. The Lagrangian function is

$$\mathcal{L}(P,L,\lambda) = S(P,L) + \pi(P,L) + \lambda\pi(P,L)$$

The first order conditions are

$$S_p^i + (1 + \lambda)\pi_p^i = 0$$

$$S_L^i + (1 + \lambda)\pi_L^i = 0$$

S_p^i is the partial derivative of S with respect to P^i. By the usual "duality" result $S_p^i = -X^i$. Similarly $S_L^i = - M^i$. Intuitively, this last equality makes sense because an increase in the license fee reduces everyone's surplus by the license fee and forces marginal consumers out. However, marginal consumers have no surplus, so the only effect is that surplus to the group as a whole falls by M^i. Using C' to denote marginal cost, the first order conditions can be rewritten:

$$[(P^i - C')X_p^i + L^iM_p^i](1 + \lambda) + \lambda X^i = 0 \qquad (1)$$

$$[(P^i - C')X_L^i + L^iM_L^i](1 + \lambda) + \lambda M^i = 0 \qquad (2)$$

$$\lambda \geq 0, \ \pi \geq 0, \ \lambda\pi = 0.$$

The logic of the problem requires that $P^i \geq 0$, $L^i \geq 0$. We assume interior solutions. Corner solutions are possible but require a lot of explanation and contribute nothing substantial.

The first point to observe is that the groups of consumers are all connected through the Lagrange multiplier λ. If a non-distorting license fee can be imposed on one group, all groups should consume at the "first best" position with $P^i = C'$.

Proposition 1.

If, at the optimum, $M_L^i = 0$ for some i (with $M^i > 0$) then $\lambda = 0$ and $P^j = C'$ for all j. If, in addition, $M_L^j \neq 0$ for some $j \neq i$, then $L^j = 0$.

Proof: See Brander and Spencer (1980).

Proposition (1) is concerned with the case in which, given the price and license fee, there are no marginal consumers for some group i. If $M_L^i = 0$ the license fee is (locally) a non-distorting lump sum transfer. The profit constraint is no longer binding ($\lambda = 0$) and all prices should be set equal to marginal cost. Also, license fees should be zero for any group for whom they would be distorting. With $P^j = C'$ and $L^j = 0$ all socially desirable consumption by group j takes place (provided second order conditions are satisfied). The distribution of license fees among groups with $M_L^j = 0$ is indeterminate. For the rest of the paper it is assumed that the profit constraint is binding: $\lambda > 0$.

Ramsey Rules for Final Consumption

If prices are to differ from marginal cost, the natural question to ask concerns how they should differ. The general insight, due to Ramsey (1927) is that large divergences should be associated with low elasticities. The reason is clear: when elasticities are low the deadweight welfare loss "triangle" is small compared to the transfer of surplus. Since the problem is to transfer surplus from consumers to the public utility at minimum deadweight loss it is not surprising that Ramsey rules arise. With two-part tariffs the Ramsey rules are more complicated, but the basic insights remain. To begin we consider the special case which corresponds to the pricing practise currently employed by Bell Canada: usage prices are set to zero. The problem is then to choose the optimal discriminating license fee.

Taking equation (2), setting $P = 0$, letting $\beta = \lambda/(1 + \lambda)$, and rearranging yields

$$-C'X_L^i + L^i M_L^i = \beta M^i$$

Letting the elasticity of membership with respect to the license fee, $M_L^i L^i / M^i$, be denoted ε_L^i, and denoting the average consumption of marginal consumers, X_L^i / M_L^i, by x_i^* we get

$$L^i = C' x_i^* / (1 + \beta/\varepsilon_L^i) \qquad (3)$$

This rule implies that low membership elasticities are associated with high license fees, as we might expect.

The more general case, in which discrimination over both a license fee and a price is allowed, is more complicated but a reasonable interpretation is possible. It is useful to define the following variables:

$e^i = P^i - C'$: the "excess price"

$\alpha^i = P^i X^i / L^i M^i$

$\eta_p^i = X_p^i P^i / X^i$: the price elasticity of demand by group i

$\varepsilon_p^i = M_p^i P^i / M^i$: the price elasticity of membership by group i

From first-order condition (1)

$$[(P^i - C')X_p^i + L^i M_p^i]/X^i = [(P^j - C')X_p^j + L^j M_p^j]/X^j$$

Rewriting this in terms of elasticities,

$$[e^i \eta_p^i / P^i] + \varepsilon_p^i / \alpha^i = e^j \eta_p^j / P^j + \varepsilon_p^j / \alpha^j$$

and rearranging yields

$$\frac{e^i/P^i}{e^j/P^j} = \eta_p^j/\eta_p^i + \frac{\alpha^i \varepsilon_p^j - \alpha^j \varepsilon_p^i}{\alpha^i \alpha^j \eta_p^i e^j / P^j} \qquad (4)$$

The proportional markup follows a two-stage Ramsey rule. The first term is the standard Ramsey rule, as in Hartwick (1978) for the price discrimination case. The second term is a correction factor which depends on price elasticities of membership. As we might expect the Ramsey rule for the two-part tariff is more complicated than for the linear pricing case, but fortunately it can be expressed as the "original" term plus a correction which may be positive or negative, and which will be close to zero if the groups are similar.

Local Telephone Service as a Joint Output

This section is based on the comments of George Hariton. As pointed out by Mr. Hariton, some observers feel that local telephone service is best thought of as two joint outputs: access to the telephone network and usage within the local calling area. In this case the cost function should be thought of as depending on arguments X and M separately:

$$C = C[X(P,L), M(P,L)]$$

Using C_X and C_M to denote partial derivatives the first order conditions corresponding to (1) and (2) are

$$[(P^i - C_X)X_p^i + (L^i - C_M)M_p^i](1 + \lambda) + \lambda X^i = 0 \qquad (1')$$

$$[(P^i - C_X)X_L^i + (L^i - C_M)M_L^i](1 + \lambda) + \lambda M^i = 0 \qquad (2')$$

This interpretation implies that having non-zero license fees is consistent with a first best pricing configuration. Specifically, $P^i = C_X$ and $L^i = C_M$ is the marginal cost pricing solution with the license fee being just the marginal cost of access.

Even if there is a real marginal cost associated with access, however, there may be large fixed costs that do not depend on total access or total usage at the margin, so that the profit constraint may still be binding and Ramsey rules would still be of interest. The analog to expression (4) is

$$\frac{e^i/p^i}{e^j/p^j} = \frac{\eta_p^i}{\eta_p^i} + \frac{A^i\varepsilon_p^j - A^j\varepsilon_p^i}{A^iA^j\eta_p^i e^i/p^j} \qquad (4')$$

where $A^i = X^iP^i/M^i(L^i - C_M)$

This reduces to (4) if $C_M = 0$.

This approach is probably more accurate for local telephone service. It does not alter the nature of the results in any substantial way. We prefer however, to continue considering the usual two part tariff problem, in which $C_M = 0$.

Optimal Two-Part Pricing of Inputs

The previous sections establish some results concerning the structure of optimal price-discriminating two-part tariffs when the public utility can discriminate among different heterogeneous groups of consumers. If the utility is capable of breaking the consuming public into homogeneous groups, prices can be set equal to marginal cost, license fees can be used to make up any deficit, and the first-best outcome can be achieved. The problem with heterogeneous groups is that license fees force out socially desirable consumers.

In the real world relatively little discrimination is possible. Indeed, for telephone companies the main form of discrimination is between business and residential customers. This type of discrimination is a little different from the model considered so far, however, because local telephone service used by businesses is an input to production rather than a consumption good. There is a fairly standard presumption in taxation theory that inputs should be priced at marginal cost and that only final outputs should be taxed[2] so we might wonder what kind of changes to the pricing formulas are required.

The difference between public utility pricing and optimal taxation is that the utility is not in a position to place a tax on the final output of firms who use telephone services as inputs. Consequently, it is efficient for the utility to charge firms a price that differs from marginal cost.

The main result of this section is that if the downstream industry is competitive, the input should be priced just as if it were being demanded for final consumption. If the downstream industry is not competitive, lower prices are called for.

Some extensions to the model are required. For ease of notation we shall assume that there is only one downstream industry and one group of final consumers. The generalization to the many-industry, many-group case follows directly upon reinterpretation of the relevant scaler variables as vectors. We shall also refer to the good as "telephone service" although the analysis is not specific to telephones and applies to any public utility.

As before, X stands for total output of telephone service. X^1 ans X^2 refer to consumption by final consumers and the downstream industry respectively. P and R are the associated prices and L and T are the associated license fees. M and N are interpreted as the total number of telephone connections for final consumers and for the downstream industry respectively, and there is a monthly license fee for each connection. When discussing consumers it was implicitly assumed that each consumer had at most one telephone connection, but in the case of firms we want to allow explicitly for the possibility that a single firm might use several telephone lines and pay a license fee for each.

Output of the downstream industry is denoted Y and the associated price is q, so the inverse demand function is q(Y). S^1 is consumers' surplus from telephone service and S^2 is consumers' surplus from good Y. Demands for telephone service and good Y are assumed independent in the sense that they enter utility functions in an additively separable way. This allows simple addition of S^1 and S^2 and lets us ignore cross elasticities between the two goods. Finally, π is the profit of the public utility and B is the profit of the downstream industry.

The Lagrangian function is

$$\mathcal{L} = S^1(P,L) + S^2(q(Y)) + (1 + \lambda)\pi(P,R,L,T) + B(R,T)$$

The first order conditions concerning P and L are as in equation (1) and (2). Using either subscripts or primes to denote derivatives, the first order conditions concerning R and T are

$$\mathcal{L}_R = S^2_q q' Y_R + (1 + \lambda)\pi_R + B_R = 0 \qquad (5)$$

$$T = S^2_q q' Y_T + (1 + \lambda)\pi_T + B_T = 0 \qquad (6)$$

These expressions can be usefully rearranged making use of the following relationships. Let k = marginal cost in the downstream industry.

$$S_q^2 = -Y \tag{7}$$

$$B_R = (q - k)Y_R + Yq'Y_R - X^2 \tag{8}$$

$$\pi_R = (R - C')X_R^2 + X^2 + TN_R \tag{9}$$

If we let $\sigma = (q - k)Y_R/X^2$ (10)

then inserting (7), (8), (9) and (10) in (5) yields

$$[(R - C')X_R^2 + TN_R](1 + \lambda) + (\lambda + \sigma)X^2 = 0 \tag{11}$$

Similarly, expression (6) becomes

$$[(P - C')X_T^2 + TN_T](1 + \lambda) + (\lambda + \delta)N = 0 \tag{12}$$

where $\delta = (q-k)Y_T/N$

Expressions (11) and (12) have exactly the same form as (1) and (2) except for the factors σ and δ. However, for competitive industries, price q is set equal to marginal cost k. Therefore, for competitive downstream industries $\sigma = \delta = 0$. Thus the following proposition can be stated:

Proposition 2

If the downstream industry is perfectly competitive, the optimal discriminating two-part tariffs for final demanders and downstream firms has exactly the same structure as in the case in which all demand is for final consumption. (This follows because the first order conditions for R and T have the same form as the first order conditions for P and L).

By Proposition 2 the two-stage Ramsey formula given by (4) applies to price discrimination among competitive downstream industries (or firms) as well as to groups of final consumers. There is, however, one important difference between the input case and the case in which telephone services are for final demand only. Specifically, Proposition 1 does not hold for downstream conpetitive industries. Even if total consumption of telephone lines does not change as the license fee, or charge per line, rises, the entire difference between marginal and average cost should not be loaded onto the competitive industry. Because higher license fees are reflected in higher prices for the output of the downstream industry, such license fees would be distorting. ($N_T = 0$ does not imply $X_T^2 = 0$ for firms, so the proof of Proposition 1 does not go through).

In any case if the downstream industry is not competitive a different setof prices and license fees should be used. Normally, imperfectly competitive firms should face lower mark-ups.

Proposition 3.

Provided that telephone service is not an inferior factor of production, an imperfectly competitive firm should be charged a lower mark-up of price over marginal cost than a corresponding competitive firm or group of final consumers.

Proof:

Consider a fairly general setting in which there are many imperfectly competitive downstream industries. We imagine setting R^i and T^i for each downstream industry i. Similarly we have the corresponding variables q_i, X^i, Y^i, N^i, and B^i. From (11) the markup over marginal cost is

$$R^i - C^i = \frac{-(\lambda + \sigma)X^i}{(1 + \lambda)X_R^i} - TN_R/XR^i$$

The difference between the competitive and imperfectly case is that σ is zero for competition but not otherwise. The definition of σ: $\sigma = (q^i - k^i)Y_R^i/X^i$ indicates that σ must be negative if telephone service is a normal factor ($Y_R^i < 0$ if X^i is normal.) Since $\lambda X^i/(1 + \lambda)X_R^i < 0$, an imperfectly competitive industry should be charged a lower mark-up over marginal cost than should a corresponding competitive industry. This completes the proof.

Thus, other things equal, imperfectly competitive industries should be charged lower markups (and lower actual prices). This comes about because it is desirable to cut back the imperfectly competitive sector less than the competitive, since the imperfectly competitive sector is already producing "too little". If, however, the input is inferior, increasing its price will increase output so a higher markup should be charged.

It is often claimed that business firms have lower demand elasticities for telephone service than do final consumers. If so this would tend to offset the relative subsidy to imperfectly competitive firms vis a vis final consumers.

Concluding Remarks

Setting price equal to marginal cost is generally the "first-best" solution to optimal pricing problems. However, if a public utility is producing with decreasing average cost and is constrained to achieve some (nonnegative) profit target, marginal cost pricing is not possible: departures from marginal cost pricing are required. Pure efficiency criteria imply the Ramsey principle: markups over marginal cost should be larger as the relevant elasticities are smaller.

The rigorous welfare-theoretic foundation of using efficiency criteria is based on one of two principles. Either there is an explicit social welfare function behind the scenes which some agent is maximizing through optimal income redistribution, or the Pareto criterion is being used in that compensating lump sum transfers are made so that everyone winds up better off (or at least no worse off).

Neither of these approaches is very practical. Indeed, if non-distorting transfers could be made easily, the problem of maximizing welfare subject to a profit constraint would not arise: prices could be set equal to marginal cost and lump sum transfers could be used to subsidize decreasing cost industries. In practise, acceptance of Ramsey pricing rules (as with cost-benefit analysis) is tantamount to adopting the policy of maximizing economic benefits "to whomsoever they may accrue".

The implicit assumption being made when efficiency measures (such as consumer and producer surplus) are used is that the social value of a dollar is the same for every consumer. Thus solutions to such problems may involve large pure transfers among consumers to achieve small efficiency gains. Furthermore, these transfers are likely to be regressive. The results of this paper suggest that such problems are probably more acute when two part tariffs are available and when downstream imperfect competition is taken into account.

Specifically, Proposition 1 shows, in extreme form, that a "captive" group of consumers will subsidize other consumers under Ramsey optimal two part pricing. Somewhat more interestingly, Proposition 3 show that imperfectly competitive downstream firms would be relatively subsidized under Ramsey optimal pricing.

It is not surprising, therefore, that subsidy free pricing has become a major concern in the study of regulation. From the regulator's point of view pricing structures that involve large cross subsidies are inequitable, and from the point of view of the regulated firms themselves, subsidized structures are dangerous because those cases in which cross-subsidization is large are precisely the cases in which private firms are likely to find competitive entry attractive, leading to so-called "sustainability" problems. (Willig (1978) considers different equity concepts for telephone pricing, and Faulhaber (1979) and Rheaume (1981) consider subsidy-free pricing for the multi-product case.

All this suggests that pure Ramsey optimal pricing may not be desirable. Also, as shown by equations (4) and (4'), the formidable array of elasticities that must be calculated makes one wonders if such formulae would ever be useful in actually setting prices in any case. Finally, although the issue has not been taken up in this paper, there is reason to believe that efficiency gains from adopting Ramsey optimal pricing structures are likely to be small. One suspects that a fairly simple usage sensitive pricing structure incorporating peak load prices close to marginal costs and subsidy free license fees to makeup deficits would be the best approach.

Footnotes

1. In the United States, of course, there is an additional rather pressing concern. Local service has traditionally been subsidized by high priced and profitable long distance service. Long distance service, however, is in competition with a growing private telecommunications industry so local service may be forced to cover its own costs. If so, usage sensitive prices promise to be an attractive method of raising revenue and cutting costs.

2. For example, Diamond and Mirrlees (1971) argue that inputs should be priced at marginal cost and that only final outputs should be taxed.

3. See Ferguson (1969), especially Ch. 6.

4. A discussion of two different concepts of equity can be found in Willig (1979).

COMMENTS ON THE PAPERS IN THE LOCAL MEASURED SERVICES SESSION

GEORGE HARITON

CRTC

I would like to thank the speakers for their contributions to our knowledge of local measured service (LMS). As we move into an era of greater competition in the telecommunications industry and as inflation increases all costs dramatically, cross-subsidy and price-averaging become more difficult to maintain. Many observers see a move toward LMS as both a way to make more efficient use of the existing plant and a way to mitigate the constant stream of rate increases that seem necessary at present. Some also see LMS as a step toward cost-based rates which, in their minds, are fairer than rates based on value of service.

Given all this, it is perhaps ironic that certain regulatory agencies are still exhorting carriers under their jurisdiction to introduce or increase extended area service, at least in certain areas. On the other hand, those regulatory agencies may know something we don't. All the more reason, therefore, to pursue research in this area.

The papers by Brander and Spencer and by Mitchell contribute to our theoretical understanding. In the first paper, the authors extend Ramsay prices to two-part tariffs (or to multiple-output situations, if "access" is distinguished as an output separate from usage), and provide a qualitative guide to low charges for LMS should be structured. Implicitly, they are operating in a second-best world. Mitchell, on the other hand, explicitly focuses on the constraints affecting telecommunications pricing. Using very simple stylized models, he shows that, under plausible assumptions as to parameter values, the constraints of price-averaging across exchanges may lead to local measured rates being less desirable than flat rates. What is even more surprising, this "inferiority" of LMS appears even when metering costs are completely ignored. (Maybe the regulatory agencies do know something after all.) Mitchell's work-in-progress may clarify the situation further. In the meantime, it is clear that caution is necessary in introducing LMS.

One of the frightening aspects of theoretical studies of LMS, at least for practitioners like me who have to actually evaluate "real-world" rate designs, is the sensitivity of the models to a host of parameters, particularly demand elasticities. This is particularly alarming because, as the participants in the session on measuring demand showed, there is very little consensus as to what those elasticities might be. Even for long distance calls, where massive data is available, price elasticity estimates vary from less than one half to more than one-and-a-half. How much worse the situation for local service, where data are scarce and hard to come by.

Nevertheless, empirical information is imperative if LMS is to be introduced in a national way. I am particularly pleased that two of the papers today are empirical studies of the impact of the introduction of LMS on local calling patterns. Williamson's paper is a careful analysis of the changes in numbers of calls resulting from changing from flat rate calling to LMS. The results are interesting in showing the dynamic response, and adjustments over time, of customers. Wong's paper concentrates on a different aspect, namely the change in the frequency distribution of calls. This is particularly important in estimating revenues from introduction of LMS with a free calling allowance.

..../3

Neither study tried to isolate the influence of other variables, reflecting the social and economic environments in which the change-overs occurred. In this sense, it is difficult to be sure that the changes in local calling patterns they report are due solely to the change to LMS. In general, it is desirable for demand estimates to be "robust", i.e. applicable to other geographical locations and to other time periods. This is especially important for estimates of the introduction of LMS, if these are to be useful for planning purposes elsewhere. Clearly, the results reported here today must be replicated elsewhere before they are usable.

Finally, it is interesting that demand studies for telecommunications services typically use aggregate time series data. I believe it would be worthwhile to carry out demand studies based on modelling decisions by individual households or firms, and using disaggregate data. Such disaggregate demand models have been constructed elsewhere in economics, most recently in transportation. There, they have had mixed success, in part because data is difficult to obtain and because models estimated from cross-sectional data have been used to forecast many years into the future. For our problem, however, I believe that the situation is more promising. In estimating the impact of changing to LMS, such data must be gathered anyway. Further, the purpose would be, not to forecast the future, but to estimate the (more or less) immediate impact of customer's reactions.

Author's Response to Discussant

T. F. Wong

Mr. Hariton asked a reasonable question regarding the transferability of my result to other situations. I would like to answer this question in two steps.

The prototype model, constructed from METROPAC data, is potentially transferable: the particular combination of unit charge, CPI and MHI [eqs. 33 and 34] was motivated by Favarini's result, which in turn is consistent with the maximization of the utility function of a household. In this form, my result could be applicable to other environments and times as characterized by MHI and CPI.

I have done a crude estimation based on the GTE results reported in this conference.

Assuming

$$p = \$0.015$$
$$CPI = 195 \ (1978)$$
$$MHI = \$20,000 \ (\text{Central Illinois, 1978})$$

Eqs. (33) and (34) give

$$\delta = 0.976$$
$$\epsilon = 0.934$$

Using flat rate usages of 350 minutes/month and 90 calls/month [Wilkinson, this conference], eqs. (15) and (24) predict measured rate usages of 304 minutes/month and 84 calls/month, or a repression of 13.1% in minutes and 6.6% in calls. These results compared favorably with the "true" repression [after the "substitution" effect was corrected] of 13.5% in minutes and 7.7% in calls calculated by Park and Wetzel [quoted by B. Mitchell, this conference]. This agreement [between widely different environments in Colorado 1971 and Illinois 1977] is really encouraging and indicates that my results could be useful in forecasting usage change in future MS conversions.

T.F. Wong, Day 2, Session 2

OPTIMAL PRICING

AN ECONOMIC ANALYSIS OF MEASURED SERVICE OPTIONS

R.E. DANSBY

J.C. PANZAR

Bell Telephone Laboratories

ABSTRACT

This paper examines the impact of an ex post billing option on a public utility's profit and its customers' net surplus, given that an optional two-part tariff is initially used to price the utility's service. The ex post billing option allows a customer to pay a premium so that their bill may be computed each billing period according to the optional two-part tariff which yields the smallest expenditure. It is shown that this ex post billing option can reduce the divergence between ex ante and ex post optima. Moreover, for some demand distributions the ex post tariff option is pareto superior to the simple flat-measured option, i.e., the ex post tariff option will increase consumers' surplus without reducing profits of the firm. However, for many demand distributions, use of the ex post option will reduce profits and aggregate welfare.

I. Introduction

Traditionally telephone utilities have primarily used flat rate tariffs to price local exchange service.[1] However, in recent years, the utilities have seen the need to change the method used to price these services. Several considerations have led to this perceived need for new pricing techniques.[2] Among the primary considerations are: (1) flat rate tariffs do not permit the utilities to base customers' bills on their usage; (2) measurement of customers' usage is now economically feasible though this has not always been the case; and (3) competition has dictated that all utility services be priced in relation to their cost. These considerations have led utilities to seek alternatives to the traditional flat rate tariffs for local exchange service.

For example, GTE has chosen to implement mandatory measured tariffs for local exchange service;[3] which requires that all customers in a given exchange be charged on a measured basis rather than by flat rate. The measured service tariff requires payment of a fixed monthly charge plus a usage charge per unit of local calling. In some cases, a measured service tariff includes a specified number of call allowances; no usage charge is imposed on local usage which does not exceed the call allowance.

An alternative scheme is typically used by Bell System companies; customers are offered the option of having their usage billed under either a measured service tariff or flat rate. Each customer must choose the tariff under which their usage is subsequently to be billed.

From the customer's point of view, the merits of the alternative tariffs may be judged in terms of their consumer surplus and income distributional consequences. Under mandatory flat-to-measured tariff conversions, high volume customers will tend to be losers while low volume customers will gain from mandatory measured service implementation. A desirable aspect of optional flat rate-measured is that customers are given the option to choose the tariff which best suits their needs. Customers whose usage is low and would achieve a larger net surplus by being billed under the measured tariff can choose that option. Because of customers' ability, under optional flat rate-measured, to select the tariff that best suits them, all customers can in theory be made better off than is possible with either flat rate or mandatory measured.[4]

However, when given such tariff options, it has been observed that customers frequently choose a tariff option which

does not yield the minimum possible bill based on actual usage.[5] That is, customers often choose a tariff which is not cost minimizing, ex post. Several explanations for this empirical observation have been advanced. For example, some researchers customers may be reluctant to
accept measured service because of their long experience and familiarity with flat rate, e.g. habit formation. On the other hand, customers may choose flat rate, when measured would be cheaper ex post, because of the prospect of a high bill under measured, i.e. customers are risk averse. There are many possible explanations for the divergence between a customer's ex ante optimal tariff and the ex post optimal tariff, even if customers are making economically rational tariff choices.

In this paper, we focus on one factor that can cause the ex ante optima to deviate from the ex post optima: inter-temporal variations in customers' usage. It is shown that the ex post billing option discussed in Mitchell can reduce the divergence between ex ante and ex post optima. Moreover, for some demand distributions the ex post tariff option is pareto superior to the simple flat-measured option. That is, for demand distributions specified herein, an ex post tariff option will increase consumers' surplus without reducing profits of the firm. However, it is important to note that for many demand distributions, use of the ex post option will reduce profits and aggregate welfare. Section II characterizes the customers who would benefit from ex post billing. Section III explores the cost and profit consequences of ex post billing. These results are derived under the assumption that customers use expense as the criterion for choice among optional tariffs; Section IV derives similar results when customers use a consumer surplus criterion for tariff selection. Concluding remarks are stated in Section V.

II. Consumers' Expenditures and Firm's Revenue

In this section, we examine the consumer surplus and profit consequences of optional flat rate-measured tariffs and ex post billing. The focus is on the potential consumer benefits of ex post billing as compared to the benefits of optional flat rate-measured service (FR-MS) tariffs.

Suppose that F is the monthly charge prescribed for flat rate, local exchange customers. The measured rate tariff has a monthly charge ρ and a price per unit of usage, P. In this intertemporal scenario, customers are given an option between flat rate and measured service; however, customers may choose between tariffs only once every n periods. Therefore, a customer who chooses flat rate will incur a total bill of

$$E^f = nF \tag{1}$$

during the n billing periods. A customer whose usage is $q_i(P)$, $i = 1,\ldots,n$, during the n periods will incur a total bill of

$$E^m = n\rho + P \sum_{i=1}^{n} q_i(P) \tag{2}$$

if the measured service option is chosen. Therefore, the expenditure of a customer who is offered optional measured service, will be

$$E^0 = \text{Min}\left[nF, n\rho + P \sum_{i=1}^{n} q_i(P)\right] = \text{Min}[E^f, E^m] \tag{3}$$

The ex post billing option allows a customer to pay a premium $\mathcal{P}$ so that their bill may be computed each billing period according to the ex ante tariff which yields the smallest expenditure. That is, under the ex post option, a customer whose usage is $q_i(P)$, $i = 1,\ldots,n$, would incur the expense

$$E^P = \mathcal{P} + \sum_{i=1}^{n} \text{Min}[F, \rho + Pq_i(P)] \tag{4}$$

The customer's choice among the optional FR-MS and ex post billing will clearly depend on the distribution of usage across billing periods. If a customer's usage is the same in each period, the customer will have no incentive to choose ex post billing. If the ex post premium is positive then a customer will prefer ex post billing only if there is sufficient variation in usage. These ideas are made precise in what follows. Suppose that

$$M^{+}(P) = \{i:\ q_i(P) \geq (F-\rho)/P\},$$
$$M^{-}(P) = \{i:\ q_i(P) < (F-\rho)/P\}. \qquad (5)$$

Hence, at usage price P, M^{+} is the set of billing periods in which usage exceeds or equals $(F-\rho)/P$. Thus, for the m^{+} billing periods in M^{+}, the customer's expenditure will be smaller under flat rate than under the measured service tariff. Similarly, for the m^{-} billing periods in M^{-}, a customer's expenditure will be smaller if usage is billed under the measured tariff. The average usage in the M^{+} periods is $\bar{q}^{+} \equiv \sum_{M^{+}} q_i/m^{+}$, while the average usage in the M^{-} periods is $\bar{q}^{-} \equiv \sum_{M^{-}} q_i/m^{-}$. It follows that[6]

<u>Proposition I:</u> If $P > 0$ then $E^{P} < E^{0}$ only if M^{+} and M^{-} are non-empty.

That is, if the ex post premium is positive, then the expenditure under the ex post tariff is smaller than the expenditure under the ex ante FR-MS option only if there are some billing periods in which flat rate is the minimum cost option and others in which measured is the minimum cost option.

It also follows that if there is no ex post premium, then a customer's expenditure under the ex post option will always be at least as small as their expenditure under the minimum cost ex ante option.[7]

Proposition II: If $P = 0$, then $E^P < E^0$ if M^+ and M^- are non-empty, while $E^P = E^0$ if M^+ or M^- is empty.

This result says that ex post billing is **valuable** to any customer whose usage varies across periods, if no premium is charged for this option. However, it should be clear that the value, to any customer, of the ex post billing option decreases as the ex post premium rises. But even if no ex post premium is charged, a customer whose demand has no intertemporal variation will be indifferent between ex post billing and optional FR-MS. The latter result presumes of course that each customer is able to compute the minimum cost ex ante option.

The relative expense to the customer, under the alternative tariff options, depends crucially on the distribution of demand. Thus, it will be useful to characterize the demand distributions for which a given tariff option is cost minimizing. For this purpose, suppose that

$$X = \left\{(m^+,\bar{q}^+,\bar{q}^-):\quad m^-\left[\frac{\mathcal{F}-\rho}{P} - \bar{q}^-\right] < \mathcal{P}/P\right\}, \tag{6a}$$

$$Y = \left\{(m^+,\bar{q}^+,\bar{q}^-):\quad m^+\left[\bar{q}^+ - \frac{\mathcal{F}-\rho}{P}\right] < \mathcal{P}/P\right\}, \tag{6b}$$

and

$$Z = \{(m^+,\bar{q}^+,\bar{q}^-):\quad \bar{q} \geq (\mathcal{F}-\rho)/P\}. \tag{6c}$$

where $\bar{q}$ is the average usage in all periods, i.e. $\bar{q} \equiv \{m^+\bar{q}^+ + m^-\bar{q}^-\}\frac{1}{n}$. The demand distributions in Z are ones for which flat rate is the minimum cost tariff under the optional FR-MS. The demand distributions in Z', the complement of Z, are ones for which measured service is the minimum cost tariff under optional FR-MS.[8]

The distributions for which flat rate is the minimum cost tariff under the ex post option are those for which $E^f < E^m$ and $E^f < E^P$. The distributions that satisfy $E^f < E^m$ are those in Z, while the distributions satisfying $E^f < E^P$ are those in X. Therefore, the demand distributions for which flat rate is the minimum cost tariff under the ex post option are those in $X \cap Z$. Similarly, the demand distributions for which measured service is the minimum cost tariff under the ex post option are those in $Y \cap Z'$; since $E^m < E^f$ in Z' and $E^m < E^P$ for distributions in Y. Finally, the ex post billing option is the cost minimum for distributions in $X' \cap Y'$, where X' and Y' are respectively the complements of X and Y.[9] The foregoing relationships among demand distributions and cost minimizing tariffs are summarized in Figure 1.

With these definitions, Proposition II may be restated simply as[10]

> Proposition IIA: If $P = 0$ then X and Y are empty sets; hence all demand distributions are in $X' \cap Y'$, i.e. $Z \cup Z' = X' \cap Y'$.

Consequently, all consumers will benefit from the use of ex post billing if no premium is charged for this billing option. Customers who are billed according to the flat rate tariff under optional FR-MS will incur a smaller expense under the ex post option in periods when usage does not exceed $(F-\rho)/P$. Customers who are billed according to the measured tariff under optional FR-MS will incur a smaller expense under the ex post option in periods when usage exceeds $(F-\rho)/P$. Thus, implementation of the ex post option, with a zero premium, may be beneficial to both flat rate and measured customers.

If the ex post premium is positive,

then not all customers will benefit if the underlying tariff parameters remain at their status quo levels. The next results provide insight concerning the customer groups that would benefit from implementation of an ex post option with a positive premium. First, consider the effect on customers whose expense is smaller when billed on a flat rate basis under optional FR-MS; these are the customers whose demand distributions are in Z. Note that the set Z is independent of P. On the other hand, the number of elements in $X' \cap Y'$ is a non-increasing function of the ex post premium. To be precise, if $h(m^+, \bar{q}^+, \bar{q}^-)$ is the number of customers with demand distributions having parameters $(m^+, \bar{q}^+, \bar{q}^-)$, then the number of customers whose expense would be lower under the ex post option, i.e. the number of customers in $X' \cap Y'$, is

$$N(X' \cap Y') \equiv \int_{X' \cap Y'} h(m^+, \bar{q}^+, \bar{q}^-) \, dm^+ d\bar{q}^+ d\bar{q}^- \qquad (7)$$

and must not increase with P.[11] This property of $X' \cap Y'$ leads us to conclude that there must be a finite P such that $X' \cap Y' \subset Z$.[12] In particular,[13]

<u>Proposition III</u>: If $P \geq \max_{Z \cup Z'} [m^-](F-\rho)$, then $X' \cap Y' \subset Z$.

This result means that if the ex post premium is greater than or equal to $(F-\rho)$ times the maximum, over all demand distributions, of the number of periods in which an individual's demand is less tha $(F-\rho)/P$, then the only beneficiaries of ex post billing will be customers whose expenditure is smallest under the flat rate tariff when they face the FR-MS option. Similarly, if $P \geq n(F-\rho)$ then the only beneficiaries of ex post billing are customers whose demand distributions are in Z when the FR-MS option is used. Consequently, when $P \geq (F-\rho)\mathrm{Max}[m^-]$, no customers in Z' will have a smaller bill under the ex post option.

The fact that $X' \cap Y'$ is a non-increasing function of P, and that $X' \cap Y' = \emptyset$ for some P also implies that there must exist an ex post premium for which customers in Z would not benefit from the ex post option.

<u>Proposition IV</u>: Suppose $P > \overline{P}$, where

$$\overline{P} = \mathrm{Min}\left[\max_{Z \cup Z'} \left\{\frac{F-\rho}{P} - \bar{q}^-\right\} Pm^-, \max_{Z \cup Z'} \left\{\bar{q}^+ - \frac{F-\rho}{P}\right\} Pm^+\right], \tag{8}$$

then $X' \cap Y' = \emptyset$.

If $P > \overline{P}$, then each customers' usage cost saving from ex post billing will be smaller than the ex post premium. Therefore, if the ex post premium is greater than $\overline{P}$, there will be no customer whose bill is smaller under the ex post option than it is under the FR-MS option.[14]

Propositions II-IV may be summarized as follows: compared to the expense incurred under the FR-MS option (1) if $P = 0$ then all customers will have a lower bill under the ex post option; (2) if $0 < P < (F-\rho)\mathrm{Max}[m^-]$, then the customers who have a lower bill under the ex post option will include some (but not necessarily all) measured service customers, those in Z', and all flat rate customers, those in Z; (3) if $P \geq (F-\rho) \max_{Z \cup Z'} [m^-]$ then the only customers who will have a lower bill under the ex post option are the flat rate customers, those in Z; and (4) if $P > \bar{P}$ then no customer will have a lower bill under the ex post option.

These results also have important implications concerning properties of the revenue function under the ex post option. To facilitate comparison of the revenues earned under the alternative tariff options, let $I(g;A) = \int_A ghdm^+d\bar{q}^+d\bar{q}^-$ denote the integral of gh(·) over the region A. The revenue, R^0, generated under the FR-MS option is then

$$R^0 = I(nF;Z) + I(n[\rho+P\bar{q}];Z'); \qquad (9)$$

the total expense incurred by all flat rate customers, $I(nF,Z)$, plus the total expense incurred by all measured service customers, $I(n\{\rho+P\bar{q}\},Z')$. The revenue, R^P, earned under the ex post option is

$$R^P = I(nF;X \cap Z) + I(n[\rho+P\bar{q}];Y \cap Z') + I(P+m^+F+m^-[\rho+P\bar{q}^-];X' \cap Y') \qquad (10)$$

Customers in $X \cap Z$ are those for whom flat rate is the minimum cost tariff and $I(nF;X \cap Z)$ is the total expense incurred by those customers. Similarly, measured service is the minimum cost tariff for customers whose demand distributions are in

$Y \cap Z'$, the total expense incurred by this consumer group is $I(n[\rho+P\bar{q}];Y \cap Z')$; note that $n[\rho+P\bar{q}]$ is the total expense incurred by a given customer over n billing periods. Finally, under the ex post option, a given customer incurs a total expense of $m^{+}F$ for periods that are billed on a flat rate basis, a total expense of $m^{-}(\rho+P\bar{q}^{-})$ for periods that are billed under the measured service tariff, and a premium $\mathcal{P}$ for the ex post billing privilege. Thus, the total expense incurred by customers under the ex post option is $I(\mathcal{P}+m^{+}F+m^{-}(\rho+P\bar{q}^{-});X' \cap Y')$.

It follows from the results in Propositions II-IV that if $\mathcal{P} = 0$ then $R^{0} > R^{P}$, and if $\mathcal{P} > \bar{\mathcal{P}}$ then $R^{0} = R^{P}$.[15] These properties of the revenue functions are evident upon noting that[16]

$$R^{0}-R^{P} = I(nF;Z - [X \cap Z]) + I(n[\rho+P\bar{q}];Z' - [Y \cap Z']) - I(\mathcal{P}+m^{+}F+m^{-}[\rho+P\bar{q}^{-}];X' \cap Y') \tag{11}$$

If $\mathcal{P} = 0$ then $X = \emptyset$ and $Y = \emptyset$, thus $X' \cap Y' = Z \cup Z'$, but $n \geq m^{+}$ and $n(\rho+P\bar{q}) \geq m^{-}(\rho+P\bar{q}^{-})$ with the strict inequality holding in some periods. Hence $\mathcal{P} = 0$ implies that $R^{0}-R^{P} > 0$. If $\mathcal{P} > \bar{\mathcal{P}}$ then $X' = \emptyset$ and $Y' = \emptyset$, hence $X \cap Z = Z$ and $Y \cap Z' = Z'$, thus $R^{P} = R^{0}$.

We also conclude from previous discussion that R^{P} is a non-decreasing function of $\mathcal{P}$ if $I(\mathcal{P}+m^{+}F+m^{-}[\rho+P\bar{q}^{-}];X' \cap Y')$ is a non-decreasing function of $\mathcal{P}$. This follows from the fact that $X \cap Z$ and $Y \cap Z'$ are non-decreasing functions of $\mathcal{P}$, while nF and $n[\rho+P\bar{q}]$ are independent of $\mathcal{P}$. Thus, the total revenue earned from flat rate and measured service customers, under the ex post option, is a non-decreasing function of the ex post premium. The last term in R^P is the total revenue earned from customers who are billed under the ex post tariff, see Eq. (4).

No customer will incur a higher bill under the ex post option than is incurred under the FR-MS option, i.e., $R^P \leq R^0$ for all demand distributions and tariff parameters. These considerations indicate that if there exists an ex post premium for which profits are larger under ex post billing than under optional FR-MS, then it must be in the interval $(0, \bar{P})$.

III. Cost of Service and Profits

The total cost incurred by the firm offering local service under the FR-MS option is

$$C^0 = c+I(c^u n\bar{q}(0);Z) + I(n[c^u+c^m]\bar{q}\ (P);Z') \tag{12}$$

where c is a fixed cost, c^u is the cost per unit of usage and c^m is the cost of measurement per unit of usage. The second term in C^0 is the total cost of usage generated by flat rate customers and the third term in C^0 is the total cost of usage and measurement generated by measured service customers.

The total cost incurred by the firm if it offers local service under the ex post option is

$$C^P = \bar{c} + I(c^u n\bar{q}(0); X \cap Z) + I(n[c^u + c^m]\bar{q}; Y \cap Z') + I(c^u m^+ \bar{q}^+(0); X' \cap Y') + I(m^-[c^u + c^m]\bar{q}^-; X' \cap Y') \tag{13}$$

Here, $\bar{c}$ is the fixed cost incurred under the ex post option; in general, we would expect that $\bar{c} \geq c$. The second term in C^P is the total cost generated by customers who are served on a flat rate basis under the ex post option. The third term in C^P is the total cost of usage and measurement generated by customers who are served under the measured service tariff. The fourth term is the total cost of usage generated by customers on the ex post tariff in periods when they are billed ex post on a flat rate basis. Finally, the fifth term in C^P is the total cost of usage and measurement generated by customers on the ex post tariff in periods when they are billed ex post on a measured service basis.

The total profit of the firm when the optional FR-MS tariff schedule is used is then

$$\pi^0 = R^0 - C^0,$$

(see Eqs. (9) and (12)), which reduces to

$$\pi^0 = I(n[F - c^u\bar{q}(0)]; Z) + I(n[\rho + \{P - c^u - c^m\}\bar{q}]; Z') - c. \tag{14}$$

The first term is the net profit earned from customers billed on flat rate under the optional FR-MS. Alternatively, the second term is the net profit earned from customers billed on a measured basis under the optional FR-MS. On the other hand, the firm earns the profit $\pi^P = R^P - C^P$ if service is provided under the ex post tariff option. From Eqs. (10) and (13), it follows that

$$\begin{aligned}\pi^P = {} & I(n[\mathcal{F}-c^u\bar{q}(0)];X \cap Z) + I(n[\rho + \{P-c^u-c^m\}\bar{q}];Y \cap Z') \\ & +I(\mathcal{P}+m^+[\mathcal{F}-c^u\bar{q}^+(0)] + m^-[\rho + \{P-c^u-c^m\}\bar{q}^-];X' \cap Y') - \bar{c}\end{aligned} \tag{15}$$

When the ex post option is used, there will be three classes of customers: (1) those customers who are served in all periods under the flat rate tariff, (the first term in Eq. (15) describes the net profit earned from them); (2) customers who are served in all periods under the measured service tariff, (the second term is the net profit generated by them); and (3) customers who are served under the ex post tariff, i.e. they pay the ex post premium, and are billed in some periods on a flat rate basis while in other periods they are billed on a measured service basis, (the third term in Eq. (15) characterizes the net profit earned from this class of customers).

The primary question now is whether profits of the firm are higher when the ex post option is used than when service is provided under the FR-MS option. The answer depends crucially on the characteristics of the distribution of demand across billing periods and across customers. The answer also depends on the size of the ex post premium P and the parameter (F,ρ,P) of the basic tariffs.

If $P > \bar{P}$, then no customer will choose to be billed under the ex post tariff, i.e. $X' \cap Y' = \emptyset$ and the differential profit of the firm will be equal to the difference between fixed cost under the ex post option and the fixed cost under the FR-MS option, i.e.

$$\Delta\pi = \pi^0 - \pi^P = \bar{c} - c \qquad (16)$$

Since we would generally expect the fixed costs of administration and billing to be higher under the ex post option it follows that if $P > \bar{P}$ then $\pi^0 > \pi^P$.

However, depending on the demand distributions, there may exist an ex post premium $P < \bar{P}$ for which $\pi^P > \pi^0$. From Eqs. (14) and (15) we deduce the general formula for the profits differential

$$\begin{aligned} \Delta\pi = (\bar{c}-c) &+ I(n[F-c^u\bar{q}(0)];Z - \{X \cap Z\}) \\ &+ I(n[\rho + \{P-c^u-c^m\}\bar{q}];Z' - \{Y \cap Z'\}) \qquad (17) \\ &+ I(m^+[c^u\bar{q}^+(0) - F] + m^-[(c^u+c^m-P)\bar{q}^- - \rho] - P, X' \cap Y') \end{aligned}$$

Our previous discussions indicate that this profits differential $\Delta\pi = \pi^0 - \pi^P$ will not be negative for all demand distributions. However, we will demonstrate that it is negative, i.e. $\pi^P > \pi^0$, for some demand distributions. Such a demonstration will of course require that restrictions be placed on the parameters of the tariff structure.[17]

<u>Proposition V</u>: If the tariff parameters satisfy

$$[(\bar{c}-c)/I(1,X' \cap Y')] < P < \bar{P} \; , \tag{18a}$$

$$p = c^u + c^m , \tag{18b}$$

$$F = c^u I(\bar{q}(0);Z)/I(1;Z) , \tag{18c}$$

and the demand distribution satisfies

$$\frac{I(\bar{q}(0);Z)}{I(1;Z)} > \frac{I(\bar{q}(0);Z - \{X \cap Z\})}{I(1;Z - \{X \cap Z\})} \; , \tag{18d}$$

$$I(n\; Z' - \{Y \cap Z'\}) < I(m^-;X' \cap Y') \; , \tag{18e}$$

then $\pi^P > \pi^0$ and the ex post option is a Pareto improvement.

The condition in (18a) requires the ex post premium to exceed the additional cost of administration and billing per customer served under the ex post option; but the ex post premium must be less than the maximum price that any customer is willing to pay for the ex post option. The condition (18b) says that measured service usage is priced at its marginal cost; while (18c) presumes that the flat rate monthly charge has been chosen so that under optional FR-MS the total revenue from flat rate customers equals the total variable cost of usage generated by them. The (18d) restriction on the demand distribution requires that the average usage of flat rate customers, when facing the FR-MS option, be smaller than the average usage of customers who would choose flat rate when facing the FR-MS option but would not choose flat rate when facing the ex post option. Finally, condition (18e) requires that the number of customer-periods, which are billed under the measured service tariff, must be larger under the ex post option than when the FR-MS option is used.

If these conditions held then implementation of the ex post option would be a pareto improvement over the FR-MS option, since customers' surplus and profit would increase. But there are many demand distributions for which the ex post option would not be a pareto improvement, since the firm's profits would decline. Moreover, one may easily construct cases in which the ex post option will decrease aggregate welfare; for example, if $\bar{\rho} = 0$ and $\bar{c}$ is much larger than c. (See summary in Figure 2.) Therefore, we conclude that for status quo tariff parameters the ex post option will increase customers' surplus but whether it simultaneously increases aggregate welfare or the firm's profit becomes an empirical question.

IV. Consumer Surplus Choice Criterion

The results in Section II and III (which are summarized in Figure 2) were derived under the assumption that each consumer chooses the tariff option which gives them the lowest total bill. An alternative choice criterion that consumers might use is to select the tariff option which yields the largest consumer surplus. However, even if consumers used the consumer surplus criterion for tariff selection, the principle results would be qualitatively identical to those summarized in Figure 2. That is to say, if consumers made their tariff selection decisions on the basis of consumer surplus, then the benefits of ex post billing (as compared to the FR-MS option) would be summarized as follows:

1) If the ex post premium, P, were equal to zero then ex post billing will increase the consumer surplus of all customer classes but will decrease the firm's profit.
2) There exists a finite $\tilde{P}$ such that if $P > \tilde{P}$ then ex post billing will not increase the consumer surplus of any customer and will decrease the firm's profits.
3) Ex post billing will increase the firm's profit only if $0 < P < \tilde{P}$; however, there are demand distributions for which it is not possible to choose a value of P which increases profits.
4) If $0 < P < \tilde{P}$, then customers who choose flat rate under the FR-MS option will receive a larger relative share of

the consumer surplus benefits of ex post billing

the larger is the ex post premium.

Therefore, the ex post billing option will tend to benefit some consumers but will tend to adversely affect profits of the firm. Thus, in general, the ex post option will not be pareto superior to the FR-MS option; even if customers use the consumer surplus criterion to choose among tariff options.

These insights are derived by the same methods used in Sections II and III; hence for brevity the details are merely sketched here. Let $S_i(P) = \int_P^{\infty} q_i(y)dy$, i.e. $S_i(P)$ is an individual's consumer surplus in period i when usage is priced at P dollars per unit. The total consumer surplus derived by an individual during the n billing periods is $S = \sum_{i=1}^{n} S_i(P)$. The net consumer surplus of an individual billed on a flat rate basis is therefore

$$S^f = S(0) - nF$$

while the individual's net consumer surplus if billed under the measured service tariff is

$$S^m = S(P) - n\rho.$$

Therefore, an individual who uses the consumer surplus criterion for tariff selection would choose flat rate if $S^f \geq S^m$, i.e. if

$(S(0) - S(P))/nP \geq (F-\rho)/P$. Alternatively the individual would choose measured service if $S^m > S^f$. Let[18]

$$\hat{Z} = \{(q_1,\ldots,q_n): \; S^f \geq S^m\},$$

then individuals having demand distributions in the set $\hat{Z}$ will choose flat rate under the FR-MS option while individuals having demand distributions in the complement of $\hat{Z}$, i.e. in $\hat{Z}'$, will choose measured service.

Under ex post billing, an individual would achieve a net surplus of

$$S^P = \sum_{i=1}^{n} \text{Max}[S_i(P) + Pq_i(P) - \text{Min}\{F,\rho+Pq_i(P)\};\; S_i(0) - \text{Min}\{F,\rho+Pq_i(0)\}]$$

This expression can be simplified upon noting that no rational consumer would consume $q_i(P)$ even though $\text{Min}[F,\rho+Pq_i(P)] > \text{Min}[F,\rho+Pq_i(0)]$. Hence there are three possible combinations of consumption levels and ex post billing rate: (1) a customer consumes $q_i(P)$ and $\text{Min}[F,\rho+Pq_i(P)] < \text{Min}[F,\rho+Pq_i(0)]$; (2) a customer consumes $q_i(0)$ while $\text{Min}[F,\rho+Pq_i(P)] > \text{Min}[F,\rho+Pq_i(0)]$; and (3) a customer consumes $q_i(0)$ even though $\text{Min}[F,\rho+Pq_i(P)] < \text{Min}[F,\rho+Pq_i(0)]$.[19] With this information the billing periods can be separated into three sets, based on the classifications in (1)-(3) above, that accomplish the same purpose as the sets

M^+ and M^- used in Sections II and III. Having simplified the expression for S^P by using these sets we define

$$\hat{X} = \{(q_1,\ldots,q_n): \quad S^f > S^P - P\}$$

and

$$\hat{Y} = \{(q_1,\ldots,q_n): \quad S^m > S^P - P\}$$

Therefore, when customers use the consumer surplus criterion to choose among tariff options, the tariff option chosen for different demand distributions is summarized in what follows. Under the FR-MS option, a customer chooses: (1) flat rate if $(q_1,\ldots,q_n)\varepsilon\hat{Z}$; and (2) measured service if $(q_1,\ldots,q_n)\varepsilon\hat{Z}'$. Under the ex post option a customer chooses: (1) flat rate if $(q_1,\ldots,q_n)\varepsilon\hat{X} \cap \hat{Z}$; (2) measured service if $(q_1,\ldots,q_n)\varepsilon\hat{Y} \cap \hat{Z}'$ and ex post billing if $(q_1,\ldots,q_n)\varepsilon\hat{X}' \cap \hat{Y}'$.[20] (Note the similarity of these results to the tariff choices summarized in Figure 1.) With these definitions the analysis proceeds exactly as developed in Sections II and III.

V. Conclusions

In this paper, it has been shown that the ex post billing option will not, in general, be a pareto improvement over the simple flat rate-measured service tariff option. This result was shown to be true when consumers choose among alternative tariffs on the basis of expense or consumer surplus. In general, the ex post option will increase consumers' benefits but may decrease profits of the firm. It was demonstrated, however, that there are some demand distributions for which the ex post option would increase the firm's profit. The conditions, which are sufficient for this result, place severe restrictions on the demand distributions and price parameters. Hence, on these theoretical grounds, ex post billing does not appear to be a pricing option whose implementation should be advocated without detailed empirical analysis of the tradeoffs noted in this paper.

Footnotes

* The author's interest in the subject of this paper was stimulated by discussion with John C. Panzar who also influenced the selection of issues addressed in the paper. I adhere to the usual error liability protocol.

1. For example, in 1950, 81.3% of residence-local exchange customers in the Bell System received local service under flat rate tariffs.

2. See Linhart-Garfinkel.

3. See Alleman.

4. Faulhaber-Panzar shows that offering more tariff options usually is pareto optimal.

5. See Pavarini.

6. Suppose that M^+ is empty, then $M^- = \{i = 1,\ldots,n\}$ which implies that $\bar{q} = \sum_{i=1}^{n} q_i/n < (F-\rho)/P$, thus $E^0 = n(\rho+P\bar{q})$ and $E^P = P + n(\rho+P\bar{q})$ consequently, $E^0 < E^P$ if $P > 0$. Alternatively, if M^- is empty, then $M^+ = \{i = 1,\ldots,n\}$ which implies that $\bar{q} \geq (F-\rho)/P$, thus $E^0 = nF$ and $E^P = P+nF$, hence $E^0 < E^P$ if $P > 0$. Therefore, if either M^+ or M^- is empty then E^P cannot be less than E^0 if $P > 0$. It follows that $E^P < E^0$ for $P > 0$ only if M^+ and M^- are non-empty.

7. Suppose $P = 0$, then from Eq. (4) it follows that M^+ and M^- being non-empty implies that

$$E^P = m^+F+m^-(\rho+P\bar{q}^-).$$

If $E^0 = nF \equiv (m^++m^-)F$, then $E^P < E^0$ since $(\rho+P\bar{q}^-) < F$ by definition of M^-. Similarly, if $E^0 = n\rho+P\sum_i q_i \equiv m^+(\rho+P\bar{q}^+) + m^-(\rho+P\bar{q}^-)$, then $E^P < E^0$ since $F < (\rho+P\bar{q}^+)$ by definition of M^+. If M^+ is empty then $E^0 = E^m$ and $E^P = E^m$ similarly if M^- is empty then $E^0 = E^f$ and $E^P = E^f$.

8. Under optional FR-MS, flat rate is the minimum cost option if and only if $E^f \leq E^m$, see Eqs. (1) and (2); which implies that $nF \leq n\rho+Pn\bar{q}(P)$ and is equivalent to $\bar{q} \geq (F-\rho)/P$, the boundary constraint of Z. The complement of Z is $Z' \equiv \{(m^+,\bar{q}^+,\bar{q}^-)\colon \ \bar{q} < (F-\rho)/P\}$ and is the set of distributions for which $E^m < E^f$.

9. The boundaries of the sets X and Y are derived as follows.

$$E^f \leq E^P \text{ iff } nF < P + \sum_i \text{Min}[F, \rho+Pq_i(P)] \equiv P+m^+F+m^-(\rho+P\bar{q}^-)$$

which is equivalent to $m^-F < P+m^-(\rho+P\bar{q}^-)$; the definition of X follows from this inequality. Y is the set for which $E^m < E^P$, thus it is defined by $n(\rho+P\bar{q}) < P+m^+F+m^-(\rho+P\bar{q}^-)$ which is equivalent to $m^+(\rho+P\bar{q}^+) < P+m^+F$. Thus, $E^P < E^f$ and $E^P < E^m$ iff $(m^+,\bar{q}^+,\bar{q}^-)\varepsilon X' \cap Y'$.

10. If $P = 0$ then $X = \{(m^+,\bar{q}^+): m^-[(F-\rho)/P - \bar{q}^-] \leq 0\}$ but $\bar{q}^- < (F-\rho)/P$ by definition, thus $X = \emptyset$ the empty set. Similarly, $P = 0$ implies that $Y = \{(m^+,\bar{q}^+,\bar{q}^-): m^+[\bar{q}^+ - (F-\rho)/P] < 0\}$, while by definition $\bar{q}^+ > (F-\rho)/P$ thus $Y = \emptyset$. Since X and Y are both empty, their complements must both be equal to the universal set; therefore, $X' \cap Y'$ must be equal to the universal set of demand distributions.

11. From Eqs. (6a) and (6b) it follows that

$$X' = \{(m^+m\bar{q}^+,\bar{q}^-): m^-[(F-\rho)/P - \bar{q}^-] \geq P/P\}$$

and

$$Y' = \{(m^+,\bar{q}^+,\bar{q}^-): M^+[\bar{q}^+ = (F-\rho)/P] \geq P/P\}.$$

Therefore, the number of elements in X' and Y', respectively, must be a non-increasing function of P; consequently, the same must be true of $X' \cap Y'$.

12. Since $N(X' \cap Y')$ is a non-increasing function of P, and is strictly decreasing for all regions in which the distribution $h(\cdot)$ has compact support, it follows that there exists a finite P such that $X' \cap Y' = \emptyset$ and $N(X' \cap Y') = 0$, if usage is finite. Since $\emptyset \subset Z$, there is a P for which $X' \cap Y' \subset Z$.

13. Any demand distribution $(m^+,\bar{q}^+,\bar{q}^-)$ which is in $X' \cap Y'$ must satisfy

$$m^+\bar{q}^+ - m^-\bar{q}^- \geq 2\, P/P + (m^+ - m^-)(F-\rho)/P.$$

this follows from adding the bounds which specify the sets X' and Y', see Footnote 11. If $P \geq m^-(F-\rho)$, then $2\,P/P + (m^+ - m^-)(F-\rho)/P \geq [2\bar{m}^- + \bar{m}^+ - m^-](F-\rho)/P = n(F-\rho)/P$. Thus $P \geq m^-(F-\rho)$ implies that $m^+\bar{q}^+ - m^-\bar{q}^- \geq n(F-\rho)/P$, and consequently that $m^+\bar{q}^+ + m^-\bar{q}^- > n(F-\rho)/P$. Distributions which satisfy the last inequality are in Z. Hence, if $P \geq (F-\rho)\text{Max}\{m^-\}$ and if $(m-,\bar{q}^+\bar{q}^-)\varepsilon X' \cap Y'$ it follows that $(m^+,\bar{q}^+,\bar{q}^-)$ Z, i.e. $X' \cap Y' \subset Z$. Since $X' \cap Y' \subset Z$ if $P > (F-\rho)\text{Max}[m^-]$ the same must be true if $P \geq n(F-\rho)$, since $n \geq \text{Max}[m^-]$.

14. From Footnote 11, it follows that $X' = \emptyset$ and $Y' = \emptyset$, respectively, if $P > \max_{Z \cup Z'} \left\{ \frac{F-\rho}{P} - \bar{\bar{q}}^{-} \right\} m^{-}P$ and $P > \max_{Z \cup Z'} \left\{ \bar{\bar{q}}^{+} - \frac{F-\rho}{P} \right\} m^{+}P$. Since P is defined to be the minimum of these two quantities, Proposition IV follows immediately, since $X' \cap Y' = \emptyset$ if $X' = \emptyset$ or $Y' = \emptyset$.

15. Recall that customers are presumed to always choose the minimum cost tariff, given their demand distribution and the tariff option they face.

16. These equivalent expressions for the total cost functions were derived by using the following properties of the integral function I: (1) $I(f;A) + I(g;A) = I(f+g;A)$; and (2) $I(f;A) + I(f,B) = I(f;A \cup B)$.

17. From Eq. (17) it follows that if $P = c^u + c^m$ and $F = c^u I(\bar{q}(0);Z)/I(1;Z)$ then $I(n[F-c^u\bar{q}(0)];Z - \{X \cap Z\})$ is negative if condition (18d) holds; (18a) and (18e) then insure that the remaining terms in $\Delta\pi$ have a net value which is negative. Therefore, conditions (18a)-(18e) imply that $\pi^P > \pi^0$.

18. Here it is assumed that $S^f > 0$ for all relevant demand distributions.

19. These considerations imply that more customer-periods are billed under flat rate when customers use the consumer surplus criterion than when the minimum expense criterion is used for tariff selection.

20. With this notation, $\tilde{P}$ is defined as

$$\tilde{P} = \text{Min}\left[\max_{\hat{Z} \cup \hat{Z}'} \{S^P - S^f\},\ \max_{\hat{Z} \cup \hat{Z}'} \{S^P - S^m\}\right].$$

Hence, from considerations similar to those in Footnote 14, it follows that $\hat{X}' \cap \hat{Y}' = \emptyset$ if $P > \tilde{P}$.

FIGURE 1

Tariff Option	If Demand Distribution Is In	then	Cost Minimizing Tariff Is
FR-MS	Z		E^f, Flat Rate
	Z'		E^m, Measured Service
Ex Post Billing	X ∩ Z		E^f, Flat Rate
	Y ∩ Z'		E^m, Measured Service
	X' ∩ Y'		E^P, Ex Post Billing

FIGURE 2

Who Would Benefit From
Ex Post Billing?

	IF	THEN
Consumers	$P = 0$	All Customers Benefit
	$0 < P < n(F-\rho)$	Some Measured Service Customers and All Flat Rate Customers Benefit
	$n(F-\rho) < P < \bar{P}$	Only Flat Rate Customers Will Gain
	$\bar{P} < P$	No Customer Will Benefit
The Firm	$P = 0$	Firm's Profits Will Decline
	$0 < P < n(F-\rho)$	$\pi^0 - \pi^P$
	$n(F-\rho) < P < \bar{P}$	Depends Crucially on Distribution of Demand Across Customers and Periods
	$\bar{P} < P$	Firm's Profits Will Decline

THE REGULATORY PROCESS UNDER PARTIAL INFORMATION

G. WARSKETT
Carleton University

ALAIN DE FONTENAY
Government of Canada
Department of Communications

INTRODUCTION

The common principle of public utility regulation in Canada is the allowed rate of return (ROR) on capital assets, and yet in most situations and in particular telecommunications the regulatory bodies have gone further and ruled on relative prices. Were the ROR the only policy goal, then the regulator need only establish an 'absolute price level' to permit the firm to achieve the allowed ROR. However, other goals are evidently on the regulator's mind, as it sets each price individually in order to first, generate sufficient revenue which is needed to cover production operating expenses plus the allowed return, and second, possibly to produce a cross-subsidy between services thought just and appropriate. In terms of relative pricing, however, these procedures are arbitrary and may be improved by adjusting prices in the appropriate way to produce savings in consumer surplus.

At the theoretical level, this problem of achieving efficiency in consumption while simultaneously satisfying cost of supply requirements has been solved (Ramsey, 1927; Baumol & Bradford, 1970). From the operational point of view, however, the task appears forbidding since it requires knowledge of every own-price elasticities and every cross-price elasticities. Even though there have been numerous attempts at calculating cross-price elasticities, at least at a very aggregate level (Corbo et al., 1979), those do not appear to be successful. In fact, just the problem of calculating own-price elasticities is formidable enough to prevent any conclusive estimates from being made (Taylor, 1980), as the range of actual results obtained testifies (CRTC, 1980; Breslaw & Smith, 1980). It seems therefore that Ramsey prices are hardly more than a theoretical ideal to wish upon public utility regulation, having little operational substance. This issue forms the subject of our paper.

An operational solution for achieving the goal of optimal pricing under a regulatory environment has been found, in a design suggested by Vogelsang and Finsinger (1979). Yet their solution does not constitute a fully satisfactory process which is ready for actual use. One problem with the process is its static formulation which makes it inapplicable to commonly occuring situations such as cost inflation or demand cutbacks. A second and most unfortunate property of the V-F procedure is its capacity to cause the regulated utility with any increasing costs[1] to experience spells of potentially severe losses, threatening the financial viability of the company. Their method will be reviewed in the first part of this paper.

The question of optimal prices under ROR regulation has still to be fully investigated. The V-F algorithm is formally based on the assumption of a zero profit regulatory goal, although the authors indicated their intention to eventually generalize the algorithm to the ROR situation. The ROR-type regulation, however, raises serious problems for achieving optimal pricing, and we shall show briefly in the next section that the ROR regulated firm, maximizing profit by adjusting relative prices, will behave in accordance with its perceived cost. Thus, in general, it will not produce socially optimal prices, because of the Averch-Johnson overcapitalization effect.

The positive profit which the regulated firm is allowed to make under the ROR return regulation can be conceptualized as a cost of regulating a monopoly.[2] An alternative approach which minimizes on these costs and at the same time avoids the problems associated with the A-J effect, can be found in an analogue to the zero profit type of regulation that gives the profit maximizing firm the incentive to move efficiently toward optimal prices (this view also constitutes an alternative justification of the V-F rule).

In the next section, we introduce an algorithm which attempts to cope with the problems found with the V-F rule, together with a demonstration of feasibility. In contrast to the V-F rule (which alternates, depending upon the sign of the profits realized by the firm), this algorithm is a two-step procedure which is independent of the characteristics of the technology. Following this, the properties of the suggested alternative rule are described, and comparisons are made with the V-F rule. Even though our approach is free of V-F's sustainability problem, nevertheless neither rule can be shown to always dominate the others in terms of consumer welfare. In the appendix, we address the important issue of the incentive to waste created by the attempt to avoid regulation through deception, and how to overcome this misinformation problem.

There is no telling whether our rule dominates the rate of return regulation, since the latter involves a differential between the market cost of capital and the allowed rate of return, as set exogeneously by the regulator. To the extent that the difference tends toward zero, the rate of return rule will clearly dominate, but at the same time the firm will lose more and more its incentive to minimize cost. Furthermore, it is important to note that the operational use of the ROR criteria has its own problems, which apparently can be quite serious (British Columbia Telephone Company, 1980).

Inherent to regulation are inescapable information problems, as was illustrated earlier. The only hope is to minimize costs associated with the regulatory body gathering information on the economic state of the firm. This can be achieved by confining information requirements to bookkeeping data and have the firm through its actions reveal the needed cost and demand characteristics. However, these incentives, in terms of profits for the keeping, at the same time constitutes a social cost, and the regulatory situation we are describing becomes one of selling or foregoing temporary welfare gains for information. The importance of the V-F analysis for practical regulation is to be found in the light they shed on this central issue.

2. THE VOGELSANG AND FINSINGER REGULATION MODEL

The V-F approach consists of an algorithm composed of instructions issued by the regulator to the firm, designed to bring the firm from a position of positive profits to one of zero profits, with prices in their optimal ratio.[3]

The algorithmic procedure consists of two loops to cover the two situations where average costs do not increase along any ray,[4]

Figure 1 about here

---

and the case where they do so increase. In both instances the regulator sets the price level while the firm is free to set relative prices subject to the constraints imposed in that period. In setting up the evolving series of constraints the regulator is not required to know anything about current costs or demand elasticities; only observations on last period operating costs and output levels are used.

The object of allowing the firm profits when it adjusts relative prices is to avoid the information problem mentioned earlier. In effect the firm volunteers to do the job for the regulator of finding the optimal price level, by bringing prices into line with the direction of greatest welfare increase at each iteration of the regulatory process. This it does by using its own knowledge of cost and demand conditions. As a reward the firm is allowed to keep profits taken in each step. The trick lies in the fact that profit maximizing behaviour does indeed get the prices alined in the (socially) correct ratio.

Each successive step of the algorithm applied to non-increasing ray average cost (loop 1) gives rise to a stronger constraint on profits by adjusting downward a Laspeyres (chain) price index. Eventually, the routine converges to the desired regulatory target.

To deal with the more awkward case of costs which do not conform with a nonincreasing ray average cost, a second loop to the algorithm is activated, one which mitigates the regulatory impact whenever negative profits are caused. A sequence of negative profits can in fact be elicited by the V-F rule, but at some point this must end and positive profits re-emerge. However, the process can on occasions return to the mode of negative profits and remain for an unknown number of steps.

In reviewing the assumptions and results of the V-F model, we can see that its shortcomings are of three kinds. First, the model is static in its formulation and results, and does not address the central regulatory problem of the day - inflation in operating costs, and technological change that may bring about changes in productivity. Inflation places the firm, formerly at a regulatory equilibrium, into a situation of negative profits, a situation which is explicitly excluded by the model (by restricting the starting point to positive profits). This exclusion thus constitutes a serious shortcoming of the model. Thirdly, the occasions of negative profits induced by regulation bring with them the danger of bankruptcy, making the process politically as well as economically unacceptable as a practical regulatory procedure.[5]

Moreover, under dynamic conditions created by shifts in either or both demand and supply conditions, the regulatory process could allow positive profits for an indefinite period of time, contrary to the intent of the model. This would be so if convergence occurred at too slow a rate to keep pace with exogenous changes.

From the preceeding remarks, we see that it is possible for the V-F regulation to be too harsh or too generous to the utility, sometimes allowing too little revenue or other times too much in relation to costs. Shortly we examine a different regulatory process which is designed around these concerns and problems. First, however, we consider the ROR regulation as a procedure for achieving optimal pricing.

3. THE ROR RULE AND OPTIMAL PRICING

From one point of view the allowed ROR is a recognition of less than full information on the regulated monopoly that is available to the regulators. The method of regulation would fail if the allowed rate, s, did not exceed the competitive rate r, since the bookkeeping data made available to the regulator is insufficient to guard against wasteful practices or to ensure that costs are in fact being minimized. In this light the margin the allowed rate has over the competitive rate, i.e. (s-r), is a premium to be paid for the lack of full information. The issue now becomes the following. Does this 'premium' interfere with the formation of a socially optimal price ratio? This issue is set up analytically below.

First, the prices, set by the multi-output regulated natural monopoly are the solution to the constrained maximization problem

$$\max_p \{\pi(p) \mid \pi(p) \leq (s-r)k\},$$

where k is the aggregate capital stock of the firm and $\pi(p)$ is the profit function. The necessary conditions for an (internal) solution are

$$\gamma_1 \operatorname{grad} \pi(p) = (s-r) \frac{\partial k}{\partial p}$$

where $\gamma_1 = 1+\mu_1$, and μ_1 is the Lagrange multiplier. The equilibrium point will be denoted $p = p^a$, and $\pi(p^a) = \pi^a$.

In appendix 1 we develop a cost function $C = g(w,r,s,x)$ corresponding to a technical specification of production. Also, we have

$$k = g_r .$$

Thus, the above condition can be expressed as

$$\gamma_1 \text{ grad } \pi(p^a) = (s-r)\left(\frac{\partial g_r}{\partial x}\right)\Big|_{x=x^a} \frac{\partial x^a}{\partial p} .$$

Now, at p^a, the firm is enjoying a certain level of profits π^a. This profit represents as it were the social cost of less than full information to the regulator. Consider now the maximization of 'social welfare', $W(p)$, subject to this information 'cost':

$$\max_p \{W(p) \mid \pi(p) = \pi^a\}$$

yielding the necessary conditions to the equilibrium solution p^b:

$$-\text{grad}\, W(p^b) = \gamma_2 \text{ grad}\, \pi(p^b)$$

or

$$x^b = \gamma_2 \text{ grad}\, \pi^b .$$

We state the following theorem.

<u>Theorem</u>

The profit maximizing monopoly subject to ROR where $s > r$ generate socially optimal prices only if the additional condition

why not "if" once you properly specify α i.e. γ_1 & γ_2

is satisfied

$$\frac{\partial g_r}{\partial x}\frac{\partial x}{\partial p} = \alpha x$$

at the equilibrium point $p = p^e$, where $\alpha = (s-r)\gamma_2/\gamma_1$. Here the cost gradient (the gradient of the constraint) is alined in the same direction as the welfare gradient.

In the above formulation this corresponds to $p^a = p^b = p^e$.

The fact that an additional condition is needed for optimal prices under ROR regulation can be readily appreciated from the following diagram:

figure 2 about here

3. AN ALTERNATIVE REGULATORY PROCESS

This process consists of two parts, one where the regulator requires the firm to set the price level consistent with zero profits, and the other where the firm adjusts relative price under a profit ceiling. The process can be described through an example starting at positive profits. The regulator calls for a reduction in price along a ray to zero profits. In the absence of any knowledge regarding elasticities, this step on the part of the regulator represents a neutral stance. It is no more arbitrary a procedure on relative prices than the current regulatory action, and it has the virtue of a simple rule that is likely to reduce the welfare loss associated with a monopoly restriction of output in one step. Furthermore, were the regulator to have a priori knowledge regarding elasticities, this knowledge could be used to modify the rule in order to quicken convergence. The second stage allows profits under a constraint on the price level. This profit incentive will produce the appropriate adjustment in relative prices. If the firm has positive profits, the above procedure is repeated until no further adjustments are made by the firm. Thus the process converges to an equilibrium at which prices are optimal and the firm realizes zero profits.

The proof of the viability of this process consists of two parts. The first demonstrates that any ray along which prices are reduced intersects the surface in R^n corresponding to zero profits, where n is the number of services offered by the firm. The second part shows that the process converges to the desired point.

We make the following four assumptions concerning the profit function defined on the space of output prices:

A1 $\{p \mid \pi(p) = 0\} \in R^n$ is nonempty

A2 $\Pi = \{p \mid \pi(p) \geq 0\}$ is compact and convex, $\Pi \in R^n$

A3 $\pi(p)$ is continuous real valued function

A4 $\theta \notin \{p \mid \pi(p) \geq 0\}$.

Define a norm $\|p\|$ on R^n (R^n becomes a normed vector space) and define the set $S(p) \subset \{p \mid \pi(p) = 0\}$ as follows:

$$S(p) = \{p \mid \max_{p \in \pi} \|p\|\} \cap \{p \in \Pi \mid \eta p \notin \Pi, \text{ all } \eta \in (0,1)\} .$$

Note that since $\{p \mid \pi(p) = 0\} \neq \theta$, $S(p)$ is nonempty. Also, if $S(p)$ is a singleton set, write $\{p^o\} = S(p)$. Then

$$\{p \mid \pi(p) > 0\} \subset \{p \mid \theta p^o, \ \theta > 1\} .$$

If $S(p)$ contains more than a single element, write $p^1, p^2 \in S(p)$. Then for any $p \in \Pi$, $\exists \mu$, $\lambda \in R$ such that $\lambda p = \mu p^1 + (1-\mu)p^2$, $\mu \in (0,1)$. Note by convexity, $\pi(\lambda p) > 0$.

<u>Theorem</u>

There exists $\mu \in \{0,1\}$ such that for any element of $\{p| \ \pi(p) > 0\}$ we have $\pi(\mu p) = 0$.

<u>Proof</u>

Note that for any $p \in \pi$, $\exists p^1, p^2 \in S(p)$, $\begin{matrix} \eta \in R \\ \mu \in (0,1) \end{matrix}$ $\eta p = \mu p^1 + (1-\mu)p^2$, with $\pi(\eta p) > 0$ (In the case where $S(p) = p^o$ the construction is obvious).

By convexity, $\exists \mu \in (0,1)$ such that $\pi(\mu \eta p) \leq \pi(\eta p)$. Since $\theta \notin \{p| \ \pi(p) > 0\}$, by continuity, there is a neighbourhood of θ, N_θ, such that $\mu \in N_\theta$ $\pi(\mu \eta p) < 0$. Using continuity again and the fact that $\Pi(p) \gg \theta$, there is one value, μ^0, such that $\pi(\mu^0 p) = 0$. QED

To show that welfare increases under the rule, whenever $p \rightarrow \lambda p$, $\lambda < 1$, write

$$W(\lambda p) \geq W(p) + \text{grad } W(p) \cdot (\lambda p - p).$$

Then $W(\lambda p) - W(p) \geq (1-\lambda)p \cdot x(p)$, from grad $W(p) = -x$. This says that the gain in welfare is bounded from below by the revenue saving obtained from the ray price reduction, measured in the original quantities.

Under $(p-p_j)x_j = 0$, we have

$$W(p) \geq W(p_j) + \text{grad } W(p_j)(p-p_j) = W(p_j) .$$

earlier you used λ_i?

The regulatory process T is characterized by two subsequences $R = \{r_0 p_0, r_1 p_1, r_2 p_2, \ldots\}$ and $P = \{p_0, p_1, p_2, \ldots\}$ such that $T = RUP$. The subsequence $\{r_j p_j\}$ give the sequence of ray reductions and $\{p_j\}$ give the profit maximization steps under constraint. The process starts at p_0 with $\pi(p_0) \geq 0$, continues to $r_0 p_0 (r_0 \leq 1)$, p_1, $r_1 p_1 (r_1 \leq 1)$, and so on.

We assume that the consumption patterns for the firms output and technology is such that set

$$\{p \mid \pi(p) \geq 0\} \cap \{p \mid W(p) \geq c\}, \quad \text{all } c < \infty ,$$

is closed and bounded. This set is not convex in general although it is compact; therefore consider its <u>convex hull</u> H. H_j will denote the convex hull corresponding to the j step in the r subsequence.

<u>Theorem</u>

The sequence T is a convergent series with $\lim_{k \to \infty} p_k = p^*$, where p^* has the optimality property (is the solution of):

$$(p - \frac{\partial c}{\partial x})' \frac{\partial x}{\partial p} = -\lambda x, \quad 0 \leq \lambda \leq 1 .$$

Note that the sequence $\{r_j\}$ is not necessarily monotonic. However $\lim_{j \to \infty} r_j = 1$. Further, it is significant that our approach does not encounter the V-F problem of unsustainability associated with increasing costs.

Proof

First we show the strict inclusion $H_{j+1} \subset H_j$, all $j = 0, 1, \ldots$ H_j is characterized as the convex hull of

$$\{p|\ \pi(p_j) \geq 0\} \cap \{p|\ W(p_j) \geq c_j\}$$

such that $r_j p_j \in \{p|\ \pi(p) = 0\}$, i.e. $r_j p_j$ is on the boundary of H_j. Profit maximization subject to $px_j \leq r_j p_j x_j$ will produce a point in the interior of H_j, provided $r_j p_j$ is not the equilibrium point. This is because $x_j = $ -grad $W(r_j p_j)$ and $p_{j+1} \in \{p|\ p \neq r_j p_j,\ px_j \leq r_j p_j x_j\} \subset \{p|\ W(p_j) > c_j\}$. Now by construction the point $r_{j+1} p_{j+1}$ is on the boundary of H_j but interior to $\{p|\ W(p) > c_j\}$. Write c_{j+1} to correspond to $\{p|\ W(r_{j+1} p_{j+1}) \geq c_{j+1}\}$. Then $c_{j+1} > c_j$ and $H_{j+1} \subset H_j$. The process must stop where $r_j = 1$ and $r_j p_j = p_j$. At this point the firm is earning zero profits $(p_j \in \{p|\ \pi(p) = 0\}$ though it is maximizing profits (under constraint). Also welfare is maximum subject to non-negative profits and the theorem is proved. QED

4. CONCLUSION

Any realistic dimension of optimal pricing for a regulated natural monopoly must contain some assurance of its practical application. We expressed the problem as one of lack of full information to the regulator. A beginning in this direction is offered by the V-F regulator, although, as we have seen, it suffers several severe shortcomings. For example, the V-F regulation process can jeopardize the firm's viability as a profitable business; an alternative procedure has been proposed in this paper to overcome this particular problem. Secondly, the problem of wilful waste presents itself and a procedure to meet this concern has also been proposed (see the appendix). Also we investigated the difficulty of achieving socially optimal prices under ROR regulation. Finally, there remains the problem of dynamic change or shocks (inflation, technological change, change in consumer preferences). Our proposed model, appears very open to being modified so as to deal with these complications; this area will be the subject of further work.

APPENDIX - 1

THE ROR REGULATION AND OPTIMAL PRICING

The application of ROR regulation to a multi-output natural monopoly constitutes a constraint over the level of prices, leaving to the firm (n-1) degrees of freedom in which to set relative prices. In the following, we use results developed by Fuss and Waverman (1978) to construct the cost function of a ~~to show that the regulated firm will~~ profit maximizing firm under a ROR regulation, the firm ~~maximize profits by setting relative prices in ratios that are~~ maximizes profit in two steps; first for any output bundle, it minimizes cost, ~~in general not optimal (because of the A-J overcapitalization effect).~~ then, ~~given any~~ it selects that output bundle which maximizes its profits.

Consider a monopoly producing n outputs (x) and $m+1$ inputs (y,k) under the technology $F(x,y,k) \leq 0$, $(x,y,k) \in R^{n+m+1}$; k denotes aggregate capital inputs. Input prices $(w,r) \in R^{m+1}$ are given to the firm, and selling prices are denoted by $p \in R^n$. r here is the competitive rate of return on capital; let $s > r$ denote the allowed rate of return.

Lemma (Fuss and Waverman)

The cost function g^* dual to the technology $F(x,y,k) \leq 0$, when subject to the ROR constraint, is expressed as

$$C^* = g^*(x,w,r^*),$$

where

$$C^* = wy + r^*k,$$

and

$$r^* = \frac{r-\lambda s}{1-\lambda} \quad \text{where } \lambda < 0. \rightarrow 1 < \lambda < 0.$$

Proof

Profit maximization subject to ROR regulation is given by the solution to the Lagrangian maximization:

$$\max_{p,y,k} L = (1-\lambda)(px-wy) - (r-\lambda s)k - \gamma F(x,y,k)$$

where λ and γ are Lagrangian multipliers. By rescaling F we can set $\gamma = 1$ (for w,r,s given):

$$\frac{\partial L}{\partial p} = (1-\lambda)(x+p\frac{\partial x}{\partial p}) - \frac{\partial F}{\partial x}\frac{\partial x}{\partial p} = 0 , \qquad 0 \leq \lambda \leq 1$$

$$\frac{\partial L}{\partial y} = -(1-\lambda)w - \frac{\partial F}{\partial y} = 0 ,$$

$$L_k = -(r-\lambda s) - F_k = 0 ,$$

supposing no corner solution and constraints are binding. Here scalar differentials are denoted by subscripts, $F_k = \partial F/\partial k$, $\partial F/\partial y_i = F_i$, $i = 1, \ldots, m$. From above,

$$F_i/F_j = w_i/w_j , \qquad F_i/F_k = w_i/r^*$$

and the conditions are thus fulfilled for cost minimization with respect to the input prices (w,r^*). Assuming F satisfies the necessary regularity conditions, there exists a cost function dual to F expressed as $C^* = g^*(x,w,r^*)$, where in addition

$$C^* = wy + r^*k . \qquad \text{QED}$$

* g^* is in terms of the shadow price ... function g^* corresponds to the cost function of an unregulated monopoly facing a market rate of r^*. The observed cost level, given the ROR

Notice that the cost ~~without~~ regulation is, of course,

$$C = wy + rk \; .$$

Thus

$$\sout{C^* = C - \lambda sk} \qquad (1-\lambda)C^* = C - \lambda px$$

or

$$\sout{C = C^* + \lambda sk} \qquad C = (1-\lambda)C^* + \lambda px$$

Define

$$g(x,w,r,s) = (1-\lambda)\, g^*(x,w,r^*) + \lambda px$$

and observe that (using Shepard's lemma)

$$g^*_{r^*} = g_r \; \sout{= k}$$

$$g_r = (1-\lambda)\, g^*_{r^*} = k$$

6. ~~RELATED~~ REVEALED AVERAGE COST AND WASTE

The application of the algorithm presented in the preceding section raises a major problem, that of waste. While in the s subsequence of t, the firm adjusting relative prices in order to maximize profits has no incentive to waste, since any waste would curtail its profits, this is not the case in the r subsequence of the algorithm. There the firm is required to contract prices by a factor of $(1-r_j)$. As the factor r_j is dependent upon cost and demand, it is up to the firm to determine its values, yet the firm's reward is independent of r_j. Unless the regulator has some knowledge regarding the cost structure in terms of cost minimization, it has to take the firm's word that the r_j applied to eliminate the firm's profits is indeed the smallest r_j value feasible. To the extent that the firm is able to pad its cost through waste, to the extent the adjustment cost of introducing and removing waste every other period without the knowledge of the regulator is not too high,[6] the firm can in fact take advantage of waste. By introducing waste so as to moderate the proportionate reduction in prices, it decreases to the same extent the constraint in the following period. This waste strategy provides the possibility of increasing profits in each period to the extent that the waste introduced in the previous period can be eliminated. The direct cost of waste whenever profits are to be reduced to zero is then directly born by the consumers and the producer does not receive any incentive to eliminate it, except in the next period.

The problem we face with the algorithm we propose comes from its very strength in relation to the V-F algorithm, namely its independence of the firm's technology. In the first loop of the V-F algorithm, the regulator had to know that the technology exhibited increasing returns to scale. Not knowing the extent of the increase in return to scale, V-F based their algorithm on the lower bound of any increase in return to scale, namely constant return to scale. They used as information the cost level of the preceding period, as a measure of average cost. As in practice the regulator cannot be expected to know whether the average cost is decreasing or increasing, V-F introduces their second loop, and with it the possibility that the firm be unsustainable.

In practice, it would seem that the regulator must have information on cost (under conditions of cost minimization by the firm) if the firm is to be prevented from incurring waste. Moreover, bookkeeping data is not sufficient to establish whether the firm indeed is miminizing cost.[7] Our goal (the same as that of V-F) is in effect to design an algorithm with a built-in incentive for the firm to minimize cost. In our algorithm the mechanism is operative every other period, whenever the firm is left free to modify relative prices in order to maximize profits. In the other periods, a modification to the regulation rule is made whereby the information of the firm's cost structure is revealed whenever it maximizes profits (minimizes cost). Whenever average cost is decreasing, this information would cause the process to converge faster than under the V-F algorithm. This is because our approach yields a better approximation of the average cost curve than does theirs, based on the preceding period average cost.

To be able to render the average cost concept meaningful, we shall make certain assumptions concerning the technology. First of all we shall assume that there exists a cost function which can be approximated locally (i.e. over short time spans) by a separable flexible functional form and that this cost function is the relevant cost minimizing function whenever, as in every other period, the firm maximizes profits. Separability implies that there exists an output aggregator function used to define, in terms of the total observed cost, an appropriate average cost. Then we assume that the average cost function, as a function of the aggregate output, is concave to the abcissa. In other words, if X_{2j} and $X_{2(j+1)}$ denote the aggregate output level in periods 2j and 2(j+1), given any α such that $0 < \alpha < 1$, and defining X_α as $\{\alpha X_{2j} + (1-\alpha)X_{2(j+1)}\}$, then the average cost a_α, defined as C_α/X_α where C_α is the total cost corresponding to X_α, is assumed to be bounded from above by $\{\alpha a_{2j} + (1-\alpha)a_{2(j+1)}\}$. Here the average costs a_{2j} and $a_{2(j+1)}$ are defined in the same way as a_α. This corresponds to assuming that the technology is such that, over the three periods 2j, 2j+1 and 2(j+1), it is bounded by

$$C_\alpha < X_\alpha(b_0 + b_1 X_\alpha) ,$$

where b_0 and b_1 are dependent upon the observed cost in the periods 2j and 2(j+1). If b_1 is negative then the technology will be of the decreasing cost family, while a positive b_1 indicates an increasing cost technology. As the firm maximizes profit in periods 2j and 2(j+1), it will be revealing to the regulator that its technology has increasing, constant or decreasing cost, and the regulator can now use the information to set an upper bound on the firm's profit in the interim by imposing a zero profit with respect to the <u>revealed</u> upper bound to the average cost. That is, the zero profit constraint with the proportional contraction of all prices is not with respect to the effective cost C_α but in terms of the upper bound $X_\alpha(b_0+b_1X_\alpha)$, such that the constraint becomes

$$P^T_{2j+1}X_{2j+1} - X_{2j+1}(b_0+b_1X_{2j+1}) = 0,$$

where b_0 and b_1 are functions of a_{2j} and $a_{2(j+1)}$.

The firm will be able to make a profit by taking advantage of any curvature of the average cost curve and its profit, π_{2j+1}, will be

$$\pi_{2j+1} = X_{2j+1}(b_0+b_1X_{2j+1}) - C_{2j+1} \quad .$$

It should be noted at this stage that, even if the assumption regarding the average cost curve is invalid, the firm's sustainability is not at stake since the firm always has the option to set its cost and revenues in period $(2j+1)$ equal to those of period $2(j+1)$, in which case its profits would be reduced to zero in both periods.

In practice, the regulator can obtain the estimate of X, the aggregate output, for any appropriate flexible functional form through the corresponding superlative index number (Diewert, 1976; Fontenay, 1980). It follows that the rule is easy to apply. It should also be noted that it can be enforced only every other period. The enforcement will thus be ex-post, similar to the

ROR regulation in this respect. Such a modification to our regulatory procedure is possible and most importantly it preserves the Ramsey character of S prices.

WELFARE OPTIMAL SUBSIDY-FREE PRICES UNDER A REGULATED MONOPOLY

GILLES C. RHEAUME

Bell Canada

1) Introduction

One particularly important issue in economic theory is optimal pricing. In the literature, two of the discussions on regulated monopolies have been Ramsey pricing and subsidy-free pricing. They are pricing rules that are related to costs and the problems of sustainability, predatory measures and inefficient competition. Ramsey pricing has focused on the efficient allocation of resource when a profit constraint is binding. Its sustainability has been demonstrated under strict cost conditions of a natural monopoly, by W.J. Baumol, E.E. Bailey, and R.D. Willig[1]. On the other hand, subsidy-free pricing has implied revenue conditions related to costs that have been associated with sustainability and the Pareto criterion. Such pricing was analysed extensively by G.R. Faulhaber[2].

Ramsey prices are not necessarily compatible with subsidy-free prices. Such a comment has been mentioned in the literature[3] and examples of this possible incompatibility have been provided[4]. But the economic consequences from a Pareto criterion and other points of view have not been sufficiently explored.

The purpose of this theoretical paper is first, to discuss the analytical issue of incompatibility between Ramsey pricing and subsidy-free prices. Then a theoretical model is constructed to develop welfare optimal prices that are subsidy-free.

Ramsey pricing, cross-subsidization and subsidy-free prices are described in the first sections. Then, cross-subsidized Ramsey pricing leads to a discussion on the efficient allocation of resources and equity. Afterwards, the welfare model that satisfies allocative efficiency, anonymous equity[5] and the profit constraint is explored.

2) Major Assumptions

Because the following discussion is theoretical, some simplifying assumptions can be made. First, the industry is assumed to be a regulated monopoly under a profit constraint. Open entry is possible, at least for some outputs.

Second, the prices are linear. Linearity in the price structure is a constraint that could be impractical. But because of the theoretical simplicity of linear prices, they have been adopted for the purpose of this study. Services and outputs are used interchangeably although actual services should not be implied from them.

Third, there are no cross-elastic effects on demand and no externalities. Eliminating the cross-elasticities of demand and the externalities from the discussion simplifies the analysis and permits us to focus on the specific issues addressed in this paper.

The models presented in this paper use the concepts of consumer surplus, and of producer surplus or economic rent. Such concepts as measures of social welfare have been extensively debated in the literature of economic theory. They may not represent the true benefits of consumers and producers. Their assumptions have been defended and criticized. It is not the purpose of this paper to elaborate on the appropriateness of such concepts.

Services are defined according to consumer groups such that each service or subset of services is identified with a consumer group. Such a structure simplifies the discussion about subsidy-free prices and equity.

3) Ramsey Pricing

In this section, the Ramsey model is developed and discussed. The model was originally elaborated in 1927 by F.P. Ramsey[6] as a solution to a taxation problem of obtaining given revenues. It then became popular in the literature on public financing. It is only later in the development of optimal price theory that the Ramsey rule was applied to a regulated monopoly under a profit constraint. Such an adaptation became popular in 1970 with the publication of an article entitled "Optimal Departures from Marginal Cost Pricing", written by W.J. Baumol and D.F. Bradford. But in the Ramsey pricing literature, cross-subsidization, competitive entry and predatory pricing were not directly discussed.

In the Ramsey model associated with public utilities, the regulated monopoly is assumed to have a binding profit constraint and the optimization goal is to set prices such that social welfare is maximized. The maximization process is to efficiently allocate resources without having the regulated monopoly incurring excess economic profits or economic losses.

Furthermore, as mentioned earlier, let us assume that there are no externalities that need to be incorporated in the maximization problem. There are also no cross-elastic effects of demand between services. Then, the consumer surplus is defined as follows:

$$(3.1) \quad \sum_{i=1}^{n} \int_{0}^{q_i^*} p_i(q_i)\, dq_i - \sum_{i=1}^{n} p_i(q_i) q_i$$

Where	p_i	is the price of output i
	q_i	is the quantity of output i
	$p_i(q_i)$	is the inverse demand function for output i, and
	q_i^*	is the optimal quantity of output i.

The producer surplus is defined as follows:

$$(3.2) \quad \sum_{i=1}^{n} p_i(q_i)q_i - C(q)$$

Where	q	is a vector of outputs $(q_1, q_2, \ldots, q_n)$
	C(q)	is the total opportunity cost to the regulated monopoly to produce the vector of output q

At prices $p_1, p_2, \ldots, p_n$, the quantity purchased of each output is equal to the quantity sold such that the welfare function is:

$$(3.3) \quad \sum_{i=1}^{n} \int_0^{q_i^*} p_i(q_i)\, dq_i - C(q)$$

Therefore, the optimization problem is to maximize the welfare function (3.3) subject to a profit constraint that could be defined as:

$$(3.4) \quad \sum_{i=1}^{n} p_i(q_i)q_i - C(q) \leq \Pi^*$$

where Π^* is the economic profit allowed.

The profit constraint is usually assumed to be between zero and the entry costs.[7]

The solutions to the problem are:

$$(3.5) \quad p_i - \partial C/\partial q_i - \lambda\, ((\partial p_i/\partial q_i)\, q_i + p_i - \partial C/\partial q_i) \leq 0 \qquad i = 1, 2, \ldots, n$$

(3.6) $\sum_{i=1}^{n} p_i(q_i)q_i - C(\dot{q}) - \Pi^* \leq 0$

The inequalities become equalities if q_i's are strictly positive and the constraint is binding.

Transforming equation (3.5)

(3.7) $p_i^* = \varepsilon_i \; MC_i/(\varepsilon_i + a)$

Where p_i^* is the Ramsey price

ε_i is the price elasticity of demand for output i, $\varepsilon_i < 0$

MC_i is the marginal cost of output i

a is the Ramsey number, $a = -\lambda/(1 - \lambda)$

(For more information, see Appendix A).

Therefore, according to the Ramsey rule, the Pareto optimal solution under a profit constraint requires that the price of each output deviates from marginal cost according to the inverse elasticity of demand. Services with inelastic demands will have prices that deviate more from marginal cost than those with more elastic demands. Such a pricing policy assures that the quantities of each service are the least different from those under marginal cost pricing.

In the literature on Ramsey pricing, the weak inequalities of the Kuhn-Tucker conditions are assumed to be strict equalities. In other words, it is assumed that all services are provided. But, if instead of an equality, an inequality was assumed for a particular service (see Appendix B), it could still be an optimal solution where that particular service would not be offered. Provision versus non-provision of service at a particular price is usually not questioned in the Ramsey pricing literature.

4) Burden Test

In discussing Pareto optimal solutions, J. Rohlfs mentions the possibility of optimal pricing where some services are not provided[8]. To verify whether provision is the optimal solution, the "burden test" is suggested. It is defined at a particular set of prices, in this case, Ramsey prices, as follows: Suppose the service or subset of services considered was not offered by the multi-output monopoly while prices of the other services remain constant. If profits of the firm decrease from this action, then the service or subset of services passes the burden test.

If a service or subset of services passes the burden test at given prices, then its provision at optimal prices is Pareto superior to its non-provision. Such a test is similar to the subsidy-free concept.

The Pareto criterion applied in this paper and the related literature, refers to an industry not the overall economy. Therefore, it is necessary to assume that the industry's purchase of resources does not affect their allocation in the rest of the economy. Furthermore, the prices of services supplied in the industry are assumed to not affect significantly the output levels of products outside the industry.

5) The Theory of Subsidy-Free Rate Structures

The literature on cross-subsidization has defined Pareto optimal criteria for the issue of provision versus non-provision of a particular service or subset of services. It presents conditions of a subsidy-free rate structure based on the Pareto superiority principle of welfare economics.

A price structure is subsidy-free only if each customer prefers the state in which each service is provided by the multi-output firm to the state in which only some of the services are offered. Such a situation occurs because no consumer pays more for a service (or subset of services) with the provision of the other services than without, and the consumers of the other services are better off with the provision than non-provision by the multi-output enterprise. Therefore, at least some consumers are made better off and no consumer is made worse off with the provision than with the non-provision of each service by the multi-output firm. In other words, provision of service under a subsidy-free rate structure is Pareto superior to non-provision.

On the other hand, if there is cross-subsidization, then the consumers of the subsidizing service or subset of services are made worse off by the provision of the subsidized services than they would be without their provision. They are paying more for the consumption of the subsidizing services than they would if the subsidized services were not provided. Therefore, consumers of the subsidized services are made better off with the provision of these services than without by making the consumers of the subsidizing services worse off. In other words, provision of the subsidized services is not Pareto superior to their non-provision.

But non-provision of subsidized services is not Pareto superior to their provision. Under cross-subsidization, consumers of the subsidizing services are made better off by the non-provision of the subsidized services than with their provision, by making the consumers of the subsidized services worse off.

Since the consumers of the subsidizing services prefer non-provision of the subsidized services to provision while the consumers of the subsidized services prefer the opposite alternative, under cross-subsidization, provision is Pareto noncomparable to non-provision[9]. Such a solution can also be found using the Koopmans' efficiency ranking[10].

In order to identify a subsidy-free rate structure from a welfare economics' point of view, the following two alternative criteria are developed in the theoretical literature:

a) Any service or proper subset of services offered by a multi-output firm is not providing a subsidy if its revenues do not exceed its "stand alone" costs.

Let N be the set of n services considered. For any proper subset of N, $S \subset N$, q^S is a n-vector such that:

$$q_i^S > 0 \quad , \quad i \in S$$

and $$q_i^S = 0 \quad , \quad i \notin S \quad \text{but} \quad i \in N.$$

Then, any q^S is not providing a subsidy if:

$$(5.1) \qquad \sum_{i=1}^{n} p_i q_i^S \leq C(q^S)$$

where $C(q^S)$ is the stand alone cost of q^S

p_i is the given price of service i

b) Equivalently, if the revenues of any service or proper subset of services offered by the enterprise are greater or equal to its incremental costs, then it is not being subsidized.

Any q^s is not receiving a subsidy if:

$$(5.2) \quad \sum_{i=1}^{n} p_i q_i^s \geq C(q) - C(q^t)$$

where q^t is a n-vector such that

$q_i^t = 0 \quad i \in S$

$q_i^t > 0 \quad i \notin S$ but $i \in T \subset N$.

T is a proper subset of N where $S \cup T = N$ and $S \cap T = \phi$

Subsidy-free prices are not necessarily unique. On the contrary, there can be more than one subsidy-free price structure. Subsidy-free prices are such that, for any subset of outputs, its revenues must be greater than or equal to its incremental costs and smaller than or equal to its stand alone costs. If the incremental costs are significantly smaller than the stand alone costs for any subset of services, then a number of different subsidy-free price structures could be found.

Economies of scope are an essential requirement for the existence of subsidy-free prices when the regulated monopoly has a binding profit constraint. The following proof demonstrates the necessity of such cost characteristics.

For any $S, T \subset N$, $S \cap T = \phi$, $S \cup T = N$ there is a n price vector p such that

a) $pq^s + pq^t - C(q^s + q^t) = 0$

where

q^s is a n output vector such that $q_i^s > 0$ if $i \in S$

and $q_i^s = 0$ if $i \in T$

q^t is a n output vector such that $q_i^t > 0$ if $i \in T$

and $q_i^t = 0$ if $i \in S$

Furthermore, let us assume that p is a subsidy-free price structure. That is,

b) $pq^s - C(q^s + q^t) + C(q^t) \geq 0$

and

c) $pq^t - C(q^s + q^t) + C(q^s) \geq 0$

Subtracting (b) and (c) from (a),

$$C(q^s + q^t) - C(q^t) - C(q^s) \leq 0$$

Therefore, if there is a binding zero profit constraint, economies of scope are a necessary condition for the existence of subsidy-free prices from a welfare economics point of view.

It is possible for a regulated monopoly with a binding profit constraint to have the paradoxical situation where each service or subset of services is, at the same time, subsidizing and being subsidized, given the criteria of subsidy-free prices mentioned above.

For any $S, T \subset N$, $S \cap T = \phi$, $S \cup T = N$, there is a n price vector p such that

d) $pq^s + pq^t - C(q^s + q^t) = 0.$

Furthermore, let each subset have revenues greater than their stand alone costs. That is,

e) $pq^s - C(q^s) > 0$ and f) $pq^t - C(q^t) > 0$.

Subtracting e) from d), for example, g) $pq^t - C(q^s + q^t) + C(q^s) < 0$

On the other hand, if a) is subtracted from d), for example, then c) results.

Therefore, each subset of services is subsidizing and is being subsidized. In this paradoxical case, provision of service by the multi-output enterprise is not Pareto superior to non-provision. On the contrary, all consumers are worse off with the provision of service by the multi-output enterprise than they would be on a stand alone basis. They are paying more not because they are subsidizing other consumers but because they are subsidizing the multi-output enterprise such that it meets its revenue requirement. Hence, provision of service by the multi-output firm is Pareto inferior to provision of a stand alone basis.

The paradoxical case occurs only when there are diseconomies of scope. Subtracting e) and f) from d), $C(q^s) + C(q^t) - C(q^s + q^t) < 0$.

If the price structure is not subsidy-free, that is, its criteria do not hold then it implies cross-subsidization only if there are economies of scope. With the existence of diseconomies of scope, no subsidy-free price structure can be found and cross-subsidization may or may not occur.

6) <u>Cross-Subsidization and Ramsey Prices</u>

In order to have a subsidy-free price structure, each service and proper subset of services should not be subsidizing or be subsidized by the other services. Therefore, subsidy-free Ramsey prices must have revenues of each service and proper subset of services that are greater than or equal to their incremental costs and smaller than or equal to their stand alone costs. That is, for each

$q^s \in S \quad S \cap T = \phi, \quad S \cup T = N,$

(6.1) $C(q) - C(q^t) \leq \sum_{s} p_i^* q_i \leq C(q^s)$

where p_i^* is the Ramsey price for service i

$q \in N$

$q^t \in T$

For each output, it is possible to derive the boundaries for the Ramsey number in order to have subsidy-free Ramsey prices. For any output, q_i, $p_i^* = \varepsilon_i MC_i/(\varepsilon_i + a)$. Furthermore, let q^s be the vector of n outputs such that only one output i is greater than 0.

Then, from (6.1),

(6.2) $$C(q) - C(q^t) \leq \frac{\varepsilon_i MC_i}{\varepsilon_i + a} \; q_i \leq C(q^s)$$

where $q^t \in T$, $S \cap T = \emptyset$, $S \cup T = N$.

Dividing inequalities (6.2) by q_i,

(6.3) $$(C(q) - C(q^t))/q_i \leq \varepsilon_i MC_i/(\varepsilon_i + a) \leq C(q^s)/q_i .$$

The left hand side of the inequalities can be defined as the average incremental cost of q_i (AIC_i). It is the additional cost per unit of output i incurred only for the supply of that output. The right hand side of the inequalities can be defined as the average stand alone cost of q_i (ASC_i). It is the cost per unit of output i, to produce it on a stand alone basis.

By a number of mathematical operations on inequalities (6.3), it is possible to derive the following:

(6.4) $$\varepsilon_i (1 - MC_i/ASC_i) \leq -a \leq \varepsilon_i (1 - MC_i/AIC_i)$$

The Ramsey number, a, must satisfy the boundaries for each output as a necessary condition for a subsidy-free solution. But, if the outputs have significantly different elasticities of demand, marginal to average cost ratios, or both, then Ramsey prices are likely to lead to cross-subsidization.

Let us assume that Ramsey pricing does induce cross-subsidization. As was demonstrated in the previous section, a cross-subsidized pricing structure implies that the provision of subsidized services is Pareto noncomparable to their non-provision if the strict Pareto criterion is applied.

Furthermore, cross-subsidized Ramsey prices have the potential of increasing inefficiency when a regulated monopoly is susceptible to competitive entry. When a public utility has natural monopoly characteristics but the prices do not reflect such cost advantages, entrants may compete, increasing the overall costs to supply the markets. Such a situation is likely to occur under cross-subsidization. The profits from the subsidizing services offer an incentive for firms to enter the respective markets by profitably offering those services at prices that are lower than or equal to those of the monopoly. The regulated natural monopoly which cannot realize its cost advantages because of cross-subsidization would likely lose at least a portion of the profitable market making cross-subsidized Ramsey pricing unsustainable and increasing the industry's overall production costs.

Cross-subsidized Ramsey prices could also produce another type of inefficiency. Cost characteristics could possibly indicate that competition for the subsidized services would be more efficient than monopoly. But the cross-subsidized Ramsey prices for such outputs would likely discourage competition. In such a case, the monopoly supply of the subsidized outputs would have production costs above the optimal market structure solution.

7) Allocative Efficiency and Equity

In order to compare economic states from an allocative efficiency point of view, the Pareto criterion and Koopmans' efficiency rule are widely used. The Pareto criterion enunciates that one economic state has a more efficient allocation of resources than another, if it can make at least one individual better off without making anyone else worse off. Two economic states are said to be Pareto noncomparable if neither of them has resource allocations more or less efficient than the other. Noncomparable resource allocations imply that either of them is neither Pareto superior or Pareto inferior to the other.

Koopmans' efficiency ranking states that a feasible vector of outputs is more efficient than another if it has at least a greater quantity of one output and no smaller quantity of the other outputs[11]. If, between two feasible vectors of outputs, none satisfies Koopmans' efficiency ranking, then they are noncomparable from an allocative efficiency point of view.

As was discussed earlier, provision of service is Pareto noncomparable to non-provision under cross-subsidization. For a regulated monopoly under similar conditions, Koopmans' efficiency rule also cannot establish the ranking between provision and non-provision under cross-subsidization.

Under a profit constraint, $\Pi^* = 0$, provision of service would give $p^s q^s + p^t q^t - C(q^s, q^t) = 0$ where $S \cap T = \phi$, $S \cup T = N$. Furthermore, let subset S be the set of subsidized services and subset T be the set of subsidizing service

Then, $p^s q^s - C(q^s, q^t) + C(0, q^t) < 0$

and $p^t q^t - C(0, q^t) > 0$

Let the profit function, $\Pi (q^t)$, defined as $p^t q^t - C(0, q^t)$, be concave to the prices p_i, for all $i \in T$, which is a usual property of the profit functions with no cross-elasticities. Furthermore, given the profit constraint, $\Pi^* = 0$, non-provision of the subsidized services would imply $\Pi (q^t) = 0$.

Since, under provision $\Pi (q^t) > 0$, there exists at least one price, p_i, $i \in T$, which is lower under non-provision that would satisfy $\Pi (q^t) = 0$ [12]. Given the inverse relationship between quantity demanded and prices, and assuming that price elasticity of demand is not zero, then, for all $i \in T$, $q_i \geq q_i'$ and there exists at least one $q_i > q_i'$ where q_i is the level of output under non-provision of the subsidized services and q_i' is the level of output of the provision scenario. Therefore, under provision, the subsidized services have positive quantities but smaller quantities of the subsidizing services. On the other hand, under non-provision, the quantities of the subsidizing services are greater but the subsidized services have zero output levels. Therefore, provision is Koopmans' efficiently noncomparable to non-provision.

For subsidy-free prices, it has been demonstrated earlier that provision is Pareto superior to non-provision. It is more efficient from Koopmans' ranking criterion since no service or subset of services is provided less because of provision of the other services. Provision of the latter services may even permit or require (given the profit constraint) a reduction in the prices of the former service or subset of services, increasing their quantities (if they do not have zero price elasticities of demand).

The discussion above was on comparisons between provision and non-provision of service at particular prices. But it is also important to compare different price structures from an allocative efficiency point of view. Each price structure implies an economic state in which given quantities of outputs are being supplied. Furthermore, to each subset of outputs corresponds a consumer group. For the purpose of this paper, it is necessary to compare cross-subsidized Ramsey prices to subsidy-free prices.

Cross-subsidized Ramsey prices are Pareto noncomparable to a subsidy-free price structure. Consumers of the Ramsey subsidized services would lose if a subsidy-free rate structure was adopted while the subscribers to the Ramsey subsidizing services would gain by such a change. The gains and losses would be the opposite for a change from the subsidy-free rate structure to the cross-subsidized Ramsey prices.

In going from cross-subsidized Ramsey prices to subsidy-free prices, smaller quantities of the subsidized services are supplied at higher prices while greater quantities of the subsidizing services are provided at lower prices, under a profit constrained regulated monopoly[13]. The inverse situation would occur from subsidy-free prices to cross-subsidization. Therefore, applying Koopmans' efficiency ranking, cross-subsidized Ramsey prices are noncomparable to subsidy-free prices.

Using either the Pareto criterion or the Koopmans' efficiency ranking, it is impossible to state that cross-subsidized Ramsey prices are more or less allocatively efficient than subsidy-free prices. They are noncomparable from the point of view of allocative efficiency[14].

In the literature on welfare theory, compensation rules have often been discussed to provide an answer to such Pareto noncomparable states. There are a number of such rules that exist. The most stringent of them is the Scitovsky criterion. It can be defined as follows: *If*, from a comparison between economic states A and B, those that benefit from A can both potentially compensate the losers and still be in a better position than at B while those that benefit from B cannot do the same, then A is socially preferable to B.

Given the structure of the paper, the Scitovsky criterion can be applied in the following way. If, from subsidy-free prices to cross-subsidized Ramsey prices, the consumer surplus gain to the subsidized customers is greater than the consumer surplus loss to the subsidizing ones, then cross-subsidized Ramsey prices is socially preferable to the subsidy-free prices. The subsidized customers can compensate the total loss to the subsidizing consumers and still be better off, while the subsidizing cannot do the same. On the other hand, if the gain is smaller than the loss, subsidy-free prices is socially preferred to cross-subsidized Ramsey prices.

A number of unsolved theoretical issues arise from the compensation rule. First, it compares consumer benefits by using monetary value as a cardinal index of consumer welfare gains and losses. Such interpersonal comparisons of consumer welfare involve the same unsolved debate as the one for consumer surplus[15].

Second, the potential compensation rule has greater problems than consumer surplus. The latter is a measurement that leads to interpersonal comparisons. The former has the additional criticisms of embodying personal value judgements. The compensation is potential and hence, one economic state compared to another has a redistribution of income where some consumers benefit, others lose. The rule makes the value judgement that one economic state and thus, one distribution of income, is better than another because the winnings from the gainers are greater in monetary value than the losses of the losers. Since some consumers are better off while others are worse off, the compensation rule transcends the problem of allocation of resources and addresses the policy issues of distributtion. Therefore, the compensation rule makes value judgements that transcend the scope of economics[16].

One may argue that actual compensations would achieve the socially preferred state if they are adequate to make the losers at least as well off and the gainers better off. Then, such a case would be based on the Pareto criterion. But actual compensations can be made only through some form of taxation. Such an instrument may not be available or socially desirable.

Although cross-subsidized Ramsey prices and subsidy-free prices are noncomparable from an allocative efficiency point of view, they each imply a distributional principle. Since it is assumed that the producer has the same level of profit from each of the two alternatives[17], distributional considerations are between consumer groups, each group identified by the subset of outputs they are willing to purchase.

In the context of this paper, equity is the set of rules that provides a distribution of outputs between consumers. There are at least two such distributional principles that may be applied to a regulated monopoly: willingness-to-pay and anonymous equity.

The principle of willingness-to-pay describes a distribution of outputs as follows. Consumer groups that are willing to pay more for their subset of services should pay more. This implies that services associated with consumer groups that have a high willingness-to-pay should have relatively higher prices than those with consumers that have a low willingness-to-pay. Ramsey prices implicitly imply such a principle of equity. Since they are based on the inverse price elasticity of demand, prices depend on the willingness-to-pay.

Another principle of fairness is anonymous equity. It states that customers of each service or subset of services should pay at least the incremantal costs of its provision. No consumer groups should be subsidized or be subsidizing. Such a principle is equivalent to subsidy-free prices.

There may be conflicts between anonymous equity and the willingness-to-pay principle, for example, in the case of cross-subsidized Ramsey prices. As mentioned above, such prices are likely to lead to inefficient competition or could discourage efficient competition. Such circumstances would then result in an economic state that is less efficient than a subsidy-free rate structure.

Nevertheless, if the principle of willingness-to-pay is adopted , then Ramsey prices are welfare optimal prices that will satisfy allocative efficiency and meet such a form of equity. It would be socially preferable to any other price structure even if it implies cross-subsidization.

On the other hand, if anonymous equity is the principle of fairness , a welfare optimal subsidy-free price structure would then be socially preferable to cross-subsidized Ramsey prices. In order to establish such optimal prices, a welfare model needs to be developed. Such a model is found in the next section.

8) Welfare Optimal Subsidy-Free Prices

Let us attempt to define optimal prices that satisfy efficiency and anonymous equity according to technological and market considerations. To build the associated model, it is not possible to impose anonymous equity implicitly. Furthermore, it cannot be incorporated in the objective welfare function. Therefore, it can only be added by means of additional constraints.

The multi-output regulated monopoly is assumed to have a profit constraint. Furthermore, subsidy-free prices are necessary and sufficient to satisfy the principle of anonymous equity. Therefore, anonymous equity will be expressed in the model as subsidy-free prices' constraints.

For the purpose of the analysis, let us assume there are only two services that have independent demands.
The welfare function can be defined as:

$$(8.1) \quad \int_0^{q_1^*} p_1 dq_1 + \int_0^{q_2^*} p_2 dq_2 - C(q_1, q_2).$$

The profit constraint is described as follows:

$$(8.2) \quad p_1(q_1)q_1 + p_2(q_2)q_2 - C(q_1, q_2) \leq \Pi^*$$

where $\Pi^* = 0$[18] which implies that the constraint is necessarily binding[19]. Furthermore, in order that provision of service be Pareto superior to non-provision, subsidy-free prices are required. That is, for the two-output case,

$$(8.3) \quad \begin{aligned} &- p_1(q_1)q_1 + C(q_1, q_2) - C(0, q_2) \leq 0 \\ &- p_2(q_2)q_2 + C(q_1, q_2) - C(q_1, 0) \leq 0 \end{aligned}$$

Inequalities (8.3) state that each service covers at least its incremental costs such that no service will be subsidized.

The welfare optimization model is the maximization of the welfare function (8.1) subject to the profit constraint (8.2) and the subsidy-free constraints (8.3). The Kuhn-Tucker conditions are:

$$(8.4)\quad \partial L/\partial q_i = p_i\,(1+(\lambda_i-\lambda)(1+1/\varepsilon_i)) - MC_i(1+\lambda_i-\lambda) - \lambda_j\,(MC_i - MC_i') \leq 0$$

$$i,j = 1,2, \quad i \neq j, \quad \varepsilon_i < 0$$

where λ_i is the Lagrange multiplier of the subsidy-free constraint of output i,

λ is the Lagrange multiplier of the profit constraint,

MC_i is the marginal cost of output i when produced jointly with output j,

and MC_i' is the marginal cost of output i on a stand alone basis.

$$q_i\,\partial L/\partial q_i = 0, \qquad q_i \geq 0$$

$$(8.5)\qquad \partial L/\partial\lambda_1 = p_1q_1 - C(q_1,\, q_2) + C(0,\, q_2) \geq 0$$

$$\lambda_1\,\partial L/\partial\lambda_1 = 0 \qquad \lambda_1 \geq 0$$

$$(8.6)\qquad \partial L/\partial\lambda_2 = p_2q_2 - C(q_1,\, q_2) + C(q_1,\, 0) \geq 0$$

$$\lambda_2\,\partial L/\partial\lambda_2 = 0 \qquad \lambda_2 \geq 0$$

$$(8.7)\qquad \partial L/\partial\lambda = \Pi^* - p_1q_1 - p_2q_2 + C(q_1,\, q_2) \geq 0$$

$$\lambda\,\partial L/\partial\lambda = 0 \qquad \lambda \geq 0$$

Initially, these conditions do not indicate a solution since we do not know which constraints are binding. But from this two-output model, it is possible to derive some conclusions.

Since the anonymous equity principle is imposed in the model, provision of service is necessarily Pareto superior to non-provision[20]. Therefore, the outputs are strictly positive and $\partial L/\partial q_i = 0$

Reworking equation (8.4), the optimal prices are

(8.8) $p_i^* = \varepsilon_i MC_i/(\varepsilon_i + a_i) - a_{ij} \varepsilon_i(MC_i - MC_i^{'})/(\varepsilon_i + a_i)$

where $\varepsilon_i < 0$

$a_i = (\lambda_i - \lambda)/(1 + \lambda_i - \lambda)$

$a_{ij} = \lambda_j/(1 + \lambda_i - \lambda)$

Furthermore, it is not possible that $\partial L/\partial\lambda_1 = \partial L/\partial\lambda_2 = \partial L/\partial\lambda = 0$ unless there are no strict economies of scope and the difference between the costs of providing the services together and the costs of providing them separately (i.e., each on a stand alone basis) is equal to the allowed profits, Π^*.[21] Therefore, if strict economies of scope are assumed, at least one constraint must not be binding.[22]

According to our assumptions, the profit constraint is satisfied as an equality. Therefore, either both subsidy-free constraints are strict inequalities or only one is an equality.

<u>Case 1: $\lambda_i = 0, \quad i = 1,2$</u>

If both subsidy-free constraints are strict inequalities, the problem becomes the standard Ramsey pricing rule where,

$$p_i^* = \frac{\varepsilon_i MC_i}{\varepsilon_i + a} \qquad \text{where } a = -\lambda/(1 - \lambda) \text{ and } \varepsilon_i < 0.$$

<u>Case 2: $\lambda_i = 0, \quad \lambda_j > 0, \quad i,j = 1, 2$ and $i \neq j$</u>

If constraint for service j is binding, then service j covers only its incremental costs from the fact that $\partial L/\partial\lambda_j = 0$. Therefore, in order to satisfy the profit constraint equality, service i must have revenues to cover its stand alone costs. Therefore, in such a situation, the optimal solution is to price the otherwise Ramsey subsidized service according to its average incremental costs and the otherwise Ramsey subsidizing service according to its average stand alone costs.

That is, for service j,

(8.9) $p_j^* = \varepsilon_j MC_j/(\varepsilon_j + a_j)$

and

(8.10) $p_j^* q_j - C(q_i, q_j) + C(q_i, 0) = 0$

or

(8.11) $p_j^* = (C(q_i, q_j) - C(q_i, 0))/q_j = AIC_j$.

For service i,

(8.12) $p_i^* = \varepsilon_i MC_i/(\varepsilon_i + a) - a' \varepsilon_i (MC_i - MC_i')/(\varepsilon_i + a)$

where $a = -\lambda/(1 - \lambda)$

$a' = \lambda_j/(1 - \lambda)$

and

(8.13) $p_i^* q_i - C(q_i, 0) = 0$

or

(8.14) $p_i^* = C(q_i, 0)/q_i$

$= ASC_i$.

Therefore, a few interesting insights are found from this exercise. First, if the welfare optimum is an interior solution of the opportunity set, then it is the usual Ramsey pricing solution. Second, if the welfare optimum is on the boundary of the opportunity set, then the solution is at the intersection or tangency of the binding profit constraint and one of the subsidy-free constraints. Third, if Ramsey pricing leads to cross-subsidization, then the welfare model specified above suggests as a welfare optimal solution that the otherwise Ramsey subsidized service should cover only its incremental costs and the otherwise Ramsey subsidizing service should cover only its stand alone costs.

To generalize the welfare optimal subsidy-free model to n services, (n being greater than 2), a number of cumbersome problems occur. First, the number of constraints increase drastically. The formula to calculate such a number is the following:

$$\sum_{r=1}^{n-1} \frac{n!}{r!\,(n-r)!} + 1$$

where n is the number of services,

r is the number of services selected in a combination.

The first part of the formula is the number of subsidy-free constraints to take into account all possible proper subsets of services that could be produced separately. It is a summation of combinations of r services. The addition of one in the formula corresponds to the profit constraint.

Second, the number of different marginal costs required to calculate the optimal price follows the number of constraints. The total number of marginal costs for each service is:

$$0.5 \sum_{r=1}^{n-1} n!/(r!\,(n-r)!)+1$$

where n and r are defined above.

For example, if there are three services, then the welfare model has seven constraints and four different marginal costs for each service in its pricing solution.

The general model is defined as follows:

(8.15) Maximize $\sum_{i=1}^{n} \int_0^{q_i^*} p_i dq_i - C(q)$

Subject to: $-\sum_{S} p_i(q_i)q_i + C(q) - C(q^t) \leq 0$

$i \in S,\ S \cap T = \phi,\ S \cup T = N$

$S,T = 1, 2, \ldots, \sum_{r=1}^{n-1} n!/r!\,(n-r)!$

and

$\sum_{i=1}^{n} p_i(q_i)q_i - C(q) \leq \Pi^*$

where q is a n-output vector.

Solving for the n services and the $\sum_{r=1}^{n-1} n!/r!\,(n-r)! + 1$ Lagrange multipliers, the following Kuhn-Tucker conditions are found:

(8.16) $\partial L/\partial q_i = p_i\ (1 + \sum_s \lambda_s - \lambda) + p_i/\varepsilon_i\ (\sum_s \lambda_s - \lambda)$

$- MC_i\ (1 + \sum_s \lambda_s - \lambda) - \lambda_t\ (MC_i - MC_i^s) \geq 0$

$i \in S,\ S \cap T = \phi,\ S \cup T = N,\ \varepsilon_i < 0$

$S,T = 1, 2, \ldots, \sum_{r=1}^{n-1} n!/r!\ (n - r)!$

where λ_t is a $0.5 \sum_{r=1}^{n-1} n!/r!\ (n - r)!$ row vector of Lagrange multipliers,

$(MC_i - MC_i^s)$ is a $0.5 \sum_{r=1}^{n-1} n!/r!\ (n - r)!$ column vector

λ_s is the Lagrange multiplier of the s^{th} subsidy-free constraint

$q_i\ \partial L/\partial q_i = 0, \quad q_i \geq 0$

MC_i is the marginal cost of $C(q)$

MC_i^s is the marginal cost of $C(q^s)$

$i = 1, 2, \ldots, N$

(8.17) $\partial L/\partial \lambda_s = \sum_s p_i(q_i)q_i - C(q) + C(q^t) \geq 0$

$s = 1, 2, \ldots, \sum_{r=1}^{n-1} n!/r!\ (n - r)!$

$\lambda_s \partial L/\partial \lambda_s = 0, \quad \lambda_s \geq 0$

(8.18) $\partial L/\partial \lambda = \Pi^* - \sum_{i=1}^{n} p_i(q_i)q_i + C(q) \geq 0$

$\lambda \partial L/\partial \lambda = 0$

Since under the constraints of the model, provision of service is Pareto superior to non-provision, all outputs are strictly positive at the optimal solution.

That is,

$$\partial L/\partial q_i = 0.$$

Therefore,

$$(8.19)\quad p_i^* = \varepsilon_i\, MC_i/(\varepsilon_i + a_i) + a_{it}\,(MC_i - MC_i^S)\,\varepsilon_i/(\varepsilon_i + a_i)$$

where $\varepsilon_i < 0$

$$a_i = (\sum_s \lambda_s - \lambda)/(1 + \sum_s \lambda_s - \lambda)$$

a_{it} is a $0.5 \sum_{r=1}^{n-1} n!/r!(n - r)!$ row vector of λ_t multiplied by the scalar $1/(1 + \sum_s \lambda_s - \lambda)$.

If none of the subsidy-free constraints are binding, that is, $\sum_s \lambda_s = \sum_t \lambda_t = 0$ for every S, $T \subset N$, then the usual Ramsey pricing rule is the optimal solution. In such a situation,

$$a = a_i = a_j = -\lambda/(1 - \lambda) \quad \text{for any } i, j = 1, 2, \ldots, n$$

and $a_{it} = 0$ such that

$$p_i^* = \varepsilon_i\, MC_i/(\varepsilon_i + a)$$

Furthermore, if there are strict economies of scope for any combination of proper subsets such that their union equal the total set of n outputs and the profit constraint is binding, then, at most, half the subsidy-free constraints are binding. Therefore, under strict economies of joint production, at most half the subsets of services will cover only their respective incremental costs.

Equivalently, if there are strict economies of scope and the profit constraint is binding, then there are at least half the subsidy-free constraints that are not binding. Therefore, at least half of the Lagrange multipliers are equal to zero.

In the multiple output case ($n > 2$), there are, however, a number of unsettled problems where more research is required. First, for those subsets of services which have binding subsidy-free constraints, the price of each output must be found such that each subset covers its incremental cost. Second, for those subsets of services which do not have binding subsidy-free constraints, the price of each output must be found such that each subset covers both its incremental cost and a proportion of the common costs. Third, further exploration of the existence of a unique solution is required when Ramsey pricing leads to cross-subsidization. Fourth, it is possible that no single welfare optimal subsidy-free pricing structure can be found. Such a situation occurs only if there are diseconomies of scope.

9) Conclusions

The discussion on welfare optimal pricing was based on the criterion of allocative efficiency and distributional principles of equity under regulated monopolies with a profit constraint. The Pareto criterion and Koopmans' efficiency ranking were used to test for allocative efficiency.

Cross-subsidized Ramsey prices were demonstrated to be Pareto noncomparable to subsidy-free prices because of the partial order of the Pareto criterion. They were also proven to be Koopmans' efficient noncomparable under the given assumptions.

Ramsey pricing has been proposed in the literature as a means of eliminating the producer loss under marginal cost pricing. It is a welfare optimal pricing rule that satisfies a profit constraint.

Ramsey pricing does not explicitly state the welfare gain or welfare loss between consumer groups. This paper demonstrated such welfare differences between cross-subsidized Ramsey pricing and a subsidy-free rate structure. A compensation rule to choose which alternative was socially preferable, was critically analysed. Then, equity principles were discussed.

Ramsey pricing implicitly has a willingness-to-pay equity principle. It states that consumers that are willing to pay more for their service or subset of services should pay more.

Anonymous equity enunciates that no consumer group should be subsidized by, or be subsidizing, the prices of services. It is synonymous to subsidy-free pricing.

The choice between a willingness-to-pay distributional principle and anonymous equity remains a policy issue. If the former is chosen, Ramsey prices are suggested as the welfare optimal solution. On the other hand, if anonymous equity is preferred, then welfare optimal subsidy-free prices would be the suggested solution.

The paper elatorated a welfare model to achieve anonymous equity efficiently. The producer meets its allowed profit as in Ramsey pricing. The model then focuses on the distribution of the services between consumer groups.

The two-output welfare optimal subsidy-free model gave interesting insights. First, it is equivalent to Ramsey prices if the latter are subsidy-free. Second, if Ramsey prices lead to cross-subsidization, then the otherwise

Ramsey subsidized service would cover in this model its incremental costs while the otherwise Ramsey subsidizing service would cover its stand alone costs. Unfortunately, further research is required when there are more than two outputs. The solution becomes very complex to analyse in a general way.

If a profit constraint is binding, economies of scope are a necessary condition for the existence of subsidy-free prices. If there are diseconomies of scope, then provision of service by a multi-output firm is Pareto inferior to provision on a stand alone basis.

If legal barriers to entry are eliminated or become less stringent, cross-subsidized Ramsey prices, as defined in this paper, can encourage inefficient competition for the subsidizing services and can discourage competition where it could be socially desirable. On the other hand, subsidy-free prices may not completely eliminate the possibility of a non-optimal market structure, but it is a necessary condition to the sustainability of natural monopoly or to the existence of socially desirable competition.

Appendix A

From inequality (3.5) on page 4,

(A.1) $p_i^* - \partial C/\partial q_i - \lambda\ ((\partial p_i/\partial q_i)\ q_i + p_i^* - \partial C/\partial q_i) = 0$

if $x_i > 0$

(A.2) $p_i^* - MC_i - \lambda\ p_i^*\ (1 + \frac{1}{\varepsilon_i}) - \lambda MC_i = 0$

(A.3) $(1 - \lambda)\ (p_i^* - MC_i) = \lambda\ p_i^*/\varepsilon_i$

(A.4) $p_i^* - MC_i = -ap_i/\varepsilon_i$ where $a = -\lambda/(1 - \lambda)$

(A.5) $(\varepsilon_i + a)\ p_i^*/\varepsilon_i = MC_i$

(A.6) $p_i^* = \varepsilon_i MC_i/(\varepsilon_i + a)$

Appendix B

The Lagrange multiplier, λ, of Ramsey's welfare optimization problem can be interpreted as the change in social welfare for a given change in allowed profit. Let us specify $\lambda < 0$, that is, social welfare decreases as the allowed profit increases.

From inequality (3.5) on page 4,

$$(1 - \lambda)\ (p_i - MC_i) - \lambda p_i/\varepsilon_i \leq 0$$

where $\varepsilon_i < 0$.

Then,

$$p_i - MC_i - \lambda p_i/(1 - \lambda)\varepsilon_i \leq 0$$

since $1 - \lambda > 0$.

Let us specify, $a = -\lambda/(1 - \lambda) > 0$.

Thus,

$$p_i\ (\varepsilon_i + a)/\varepsilon_i < MC_i.$$

Therefore, if $|\varepsilon_i| > a$, then $p_i \leq \varepsilon_i\ MC_i/(\varepsilon_i + a)$

if $|\varepsilon_i| < a$, then $p_i \geq \varepsilon_i\ MC_i/(\varepsilon_i + a)$

if $|\varepsilon_i| = a$, then it is undefined.

Footnotes

1. Baumol, W.J., Bailey, E.E. and Willig, R.D., "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly", American Economic Review, June 1977

2. see Faulhaber's "Cross-Subsidization in Public Enterprise Pricing" and "Cross-Subsidization: Pricing in Public Enterprises".

3. see G.R. Faulhaber, AER, December 1975; W.J. Baumol, E.E. Bailey and R.D. Willig, AER, June 1977.

4. see G.R. Faulhaber, AER, December 1975.

5. Anonymous equity is the principle of fairness where no consumer subsidizes or is subsidized by another consumer. A definition is provided in the paper.

6. Ramsey's paper is entitled: "A Contribution to the Theory of Taxation".

7. The allowed economic profit is usually equal to zero. But there are models that have a wider scope for the allowed profit. They specify that it is between zero and the entry costs when the latter are strictly positive.

8. see J. Rohlfs' "Economically Efficient Bell System Pricing" pages 7-8.

9. Cross-subsidization thus implies that the Pareto criterion would not provide an argument for the firm to change its status on the provision of subsidized services. Based on the Pareto principle, it would neither eliminate subsidized services if they were already being provided nor supply subsidized services if they are not already being provided. However, such a situation does not imply that the status quo is Pareto superior to non-provision of existing subsidized services or to provision of new subsidized services. It is neither Pareto superior nor Pareto inferior to such alternatives: It is Pareto noncomparable to them.

Footnotes (cont'd)

10. Such a ranking is based on output levels as found in Koopmans' Three Essays on the State of Economic Science.

11. Koopmans' ranking criterion is based on a number of assumptions. In order that it applies to an industry, the assumptions need to be specially interpreted. First, it is assumed that the consumers of the services considered are not saturated by their consumption of these outputs. Second, the technology and resources available to the industry are supposed to have limitations. Third, the industry's purchase and allocation of resources are assumed not to affect significantly the rest of the economy.

12. There would also exist at least one price, p_i, $i \in T$, which is higher under non-provision that would satisfy $\Pi (q^t) = 0$ given the concavity of the profit function. Since the latter would be welfare inferior to the existing price under cross-subsidization and to the lower price where $\Pi (q^t) = 0$, it is socially preferable under non-provision of q^s, to choose the lower price. J.C. Panzar and R.D. Willig refers to such a choice as the undominated zero profit price vector in their paper, "Free Entry and the Sustainability of Natural Monopoly". In the AER, December 1975, G.R. Faulhaber mentions that the lower price would be chosen because of increased consumer welfare and the indifference of the firm between the higher and lower prices unless, as Faulhaber states, the firm is malicious or ignorant.

13. Such an analysis is based on a number of underlying assumptions. First, the profit functions are concave to prices. Second, the price elasticity of demand is strictly greater than minus infinity and strictly smaller than zero. Third, there exists cross-subsidized Ramsey prices. Fourth, there exists a subsidy-free price vector.

14. This conclusion is based on the arguments presented above and their assumptions.

Footnotes (cont'd)

15. Interpersonal comparisons of the compensation rule assume that the marginal utility of income for each individual is defined equal whatever the level of income and whichever the individual. Furthermore, the consumer welfare indicators can be expressed as a cardinal measurement for each individual and given the proper assumption of the marginal utility of income, interpersonal comparisons can be made. Alternative assumptions to measure variations in consumer welfare have been discussed in the literature of welfare economics. The debate remains unresolved as in the case of consumer surplus. For further information, one can refer to W.J. Baumol's Welfare Economics and the Theory of the State, or J. Rothenberg's, The Measurement of Social Welfare.

16. see W.J. Baumol, Welfare Economics and the Theory of the State.

17. Since the level of profit is the same, that is, $\Pi = 0$, it is assumed that the producer is indifferent between the two alternatives.

18. The allowed economic profit is assumed to be equal to zero because a strictly positive allowed profit is difficult to incorporate in the subsidy-free constraints. But it is not unusual to assume allowed profit to be equal to zero in a regulated monopoly. Such an assumption seems acceptable especially for the purpose of this paper.

19. The profit constraint is necessarily binding because if the profits are less than zero, not all factors are being paid their opportunity cost. Hence, the producer would incur economic losses.

20. This issue is discussed in section 5 above.

21. In the particular case being considered, since $\Pi^* = 0$,

$$C(q_1, q_2) = C(q_1, 0) + C(0, q_2).$$

Footnotes (cont'd)

22. If there are weak economies of scope and they are defined as $C(q_1, q_2) = C(q_1, 0) + C(0, q_2)$, then the welfare optimal subsidy-free solution would be that each output covers its incremental costs which are equal to its stand alone costs.

EFFICIENCY, EQUITY AND REGULATION
A MODEL OF BELL CANADA

JON BRESLAW

J. BARRY SMITH

Concordia University

1- Introduction

In this paper we study equity-efficiency pricing issues for the case of Bell Canada - a large telecommunications carrier operating as the sole supplier for almost all of Quebec and Ontario. We begin by formalizing the equity-efficiency pricing trade-off within a general economic model. The model involves an optimization problem which yields as a solution residential service prices which incorporate both efficiency and equity considerations. We next specify and estimate an econometric multi-input multi-output cost and demand model which is used to study characteristics of the Bell Canada production process. Information resulting from this empirical model is then introduced into the optimal pricing model. When the pricing model is simulated, sets of efficiency-equity prices result. Differences in the set of prices reflect different efficiency-equity weightings. An important result is that given the assumptions of the model, the stronger is the equity weighting, the closer are the historic prices to the optimal prices. We conclude with a demonstration of the fact that, within our model, the first movement towards adopting optimal prices for residential services will supply the greatest benefit to consumers.

2- Efficiency-Equity Pricing

We begin this section by considering the general problem of choosing service prices of a regulated industry so as to maximize the welfare of residential consumers of the services. The choice of prices is constrained by the requirement that the regulated industry earn no more than a predetermined level of profit. We conclude this section by developing a model in which econometric cost and demand information can be combined with the theory to provide a rigorious, consistent and tractable application of the pricing problem. The resulting "efficiency-equity" model is used to simulate socially optimal departures from the historic pattern of prices of telephone services of Bell Canada.

The Model

The canonical solution to the problem of choosing welfare maximizing prices subject to constraint is attributed to Ramsey (1927). Feldstein (1972a, 1972b, 1972c) extended the analysis to include distributive or equity considerations. The analysis presented here is similar to that of Feldstein (1972a). There are however some interesting differences and extensions. In the first place, the optimality conditions are derived using the (dual) indirect utility function approach. Secondly, a diagrammatic solution to the problem is presented.

The problem considered here can be written:

$$\underset{(p_1,p_2)}{\text{maximize}} \quad W = N\int_0^{\infty} V(p_1,p_2,p,y)\ f(y)\ dy \qquad (2.1)$$

$$\text{subject to} \quad \Pi\ (p_1,p_2,p;K) \geq \Pi_0 \qquad (2.2)$$

where N = number of consumers

V = indirect utility function of the representative consumer assumed quasi-convex in (p_1,p_2,p,y). All households are assumed to have the same preferences and face the same prices. They differ to the extent that they have different income levels (y).

p_1, p_2 = service prices to be chosen

p = price index for a composite of all other goods

$f(y)$ = relative density function of household income

$\Pi(\ )$ = profit of the regulated firm assumed quasi-concave in prices

Π_o = minimum required profit

K = a vector of parameters including characteristics of the income distribution and factor prices.

Two points should be made at the outset. First, this problem is posed for consumers only. It is assumed that the prices faced by firms for the variety of services are unchanged. Secondly, welfare (W) to be maximized can be interpreted as the sum of utilities of consumers of various incomes (or income classes) weighted by the number of consumers in these classes. Class differences, as determined by income, will provide the basis for equity considerations in the model. Consistent with the income distribution and inequality literature, the indirect utility function is assumed concave in the income argument.

The maximization problem is solved by first constructing the Lagrange function:

$$L = N\int_0^\infty V(p_1,p_2,p,y)\ f(y)\ dy + \lambda(\Pi(p_1,p_2,p;K)-\Pi_o) \qquad (2.3)$$

The first order necessary conditions for an interior constrained maximum are given by:

$$\frac{\partial L}{\partial p_1} = N\int_0^\infty \frac{\partial V}{\partial p_1}(p_1,p_2,p,y)f(y)\ dy + \lambda\frac{\partial \Pi}{\partial p_1}(p_1,p_2,p;K) = 0 \qquad (2.4)$$

$$\frac{\partial L}{\partial p_2} = N\int_0^\infty \frac{\partial V}{\partial p_2}(p_1,p_2,p,y)f(y)\ dy + \lambda\frac{\partial \Pi}{\partial p_2}(p_1,p_2,p;K) = 0 \qquad (2.5)$$

$$\frac{\partial L}{\partial \lambda} = \Pi(p_1,p_2,p;K) - \Pi_o = 0 \qquad (2.6)$$

The second order sufficient conditions require that the determinant of the Hessian matrix of the Lagrange function be positive.

The multiplier λ can be interpreted as $-\frac{dW}{d\Pi_o} > o$ or the increase in welfare arising from reducing the minimum required profit by a 'small' amount. Thus, equations (2.4) and (2.5) have the interpretation that price is set where the increase in welfare arising from lowering prices a 'small' amount is just offset by the social value of the associated decrease in profit of the regulated firm.

Is is possible to express the first order conditions in a fashion which aids interpretation of the equity contribution to the pricing scheme.

First, it will be recalled that Roy's Identity yields:

$$\frac{\partial V(p_1,p_2,p,y)}{\partial p_i} = -q_i(p_1,p_2,p,y)\ \frac{\partial V(p_1,p_2,p,y)}{\partial y} \qquad (2.7)$$

Secondly, the aggregate demand for good i, Q_i, can be defined:

$$Q_i = N\int_o^\infty q_i f(y)\ dy \qquad (2.8)$$

Thirdly, the profit derivatives can be re-written:

$$\frac{\partial \Pi(p_1,p_2,p;K)}{\partial p_i} = (MR_i - MC_i)\frac{\partial Q_i}{\partial p_i} \qquad (2.9)$$

where MR_i and MC_i are respectively the marginal revenue and marginal cost for service i and $\frac{\partial Q_1}{\partial p_2} = \frac{\partial Q_2}{\partial p_1} = 0$ by assumption.

Finally, it is convenient to follow Feldstein (1972a) in defining the distributional coefficient of i as:

$$R_i = \frac{N}{Q_i}\int_o^\infty q_i(p_1,p_2,p,y)\ \frac{\partial V(p_1,p_2,p,y)}{\partial y}\ f(y)\ dy \qquad (2.10)$$

Feldstein (1972a, p. 33) notes that:

"the ratio R_i is a weighted average of the marginal social utilities, each household's marginal social utility weighted by that household's consumption of good i. The conventional welfare assumption that $\frac{\partial V}{\partial y}$ declines as y increases implies that the value of R_i will be greater for a necessity than for a luxury. The higher the income elasticity of demand for a good, the lower the value of R_i".

Substituting (2.7), (2.8), (2.9) and (2.10) into (2.4) and (2.5) and eliminating λ yields:

$$\frac{\dfrac{MR_1 - MC_1}{p_1}}{\dfrac{MR_2 - MC_2}{p_2}} = \frac{\varepsilon_2}{\varepsilon_1}\frac{R_1}{R_2} \quad ; \quad \varepsilon_i = \frac{\partial Q_i}{\partial p_i} \cdot \frac{p_i}{Q_i} \qquad (2.11)$$

Equation (2.11) represents the optimal divergences of marginal revenues from marginal costs for both goods given both efficiency and equity considerations. The case treated by Ramsey ignored equity considerations and can be derived from (2.11) by imposing the restrictions that $R_1 \equiv 1 \equiv R_2$. With these restrictions, (2.11) reduces to the familiar Ramsey Rule:

$$\frac{\dfrac{p_1 - MC_1}{p_1}}{\dfrac{p_2 - MC_2}{p_2}} = \frac{\varepsilon_2}{\varepsilon_1} \qquad (2.12)$$

when one makes use of the result $MR_i = p_i(1 + \frac{1}{\varepsilon_i})$. Equation (2.12) has the interpretation that the optimal percentage divergence of regulated price from marginal cost for a good is inversely related to the price elasticity of demand for the service. Alternatively, the optimal tax on a good is higher the lower is the price elasticity of demand.

There is an element of discrimination in the Ramsey Rule and it is equivalent to th[illegible] normal conditions, a discriminating monopolist will charge a higher price in the less elastic of two markets. The unsettling feature of the Ramsey Rule is that, as Atkinson and Stiglitz (1972) and Pestieau

(1975) point out, less elastic goods are also often necessities and thus the brunt of the 'optimal' tax will be borne by those with lower income streams. Thus, Ramsey optimality may not be distinguishable from regressivity in this context. Equity considerations suggest that whenever a good is a necessity, the optimal tax on the good should be lower than the Ramsey Rule suggests. This latter requirement is present in equation (2.11). The fact that R_i is smaller for luxuries than for necessities reduces the optimal tax from the levels which would obtain under a Ramsey, or pure efficiency, regime.

This latter fact has been proved by Feldstein (1972a). Thus, rather than reprove the general case, it is useful to describe a specification of a welfare model which can be used in conjunction with econometric information to simulate optimal efficiency-equity prices for Bell Canada.

A Welfare Specification: Initial Considerations

Two sets of constraints arising from empirical and theoretical considerations are important in determining the specified form of the welfare model.

With respect to the constraints placed on the model by data considerations, one of the features of available telecommunications time series data is that double-log demand systems with constant own-price and income elasticity parameters will in general, provide as good a fit as is reasonable to expect (see for example, Taylor (1980)). Problems including multicollinearity and small sample size effectively preclude the accurate estimation of cross-elasticity terms or terms which would allow the own-elasticities to vary with price or income.

Theoretical constraints are perhaps the more difficult to satisfy. The empirical constraints noted above tend to support the acceptance of double-log demand models with constant own-price and income elasticities. As well, the existing evidence suggests that price and income elasticities differ across goods. Unfortunately, economic theory suggests that the only exact demands consistent with strict constancy of the own-price and income elasticicies are everywhere unit elastic and come from Cobb-Douglas utility functions. However unit elastic demands are inconsistent with observed consumer behavior. It would therefore be pointless to proceed by adopting double-log demands with elasticity parameters different from those consistent with economic theory. Fortunately, there are some conditions under which the demands derived from utility maximization are virtually indistinguishable from double-log demands. Frisch (1959) and Sato (1972) have studied the properties of demands derivable from additive utility functions. They show that if the demand data satisfy certain conditions then the demands will be almost double-log in own-price and income. It is useful to briefly rederive these results since the Bell Canada data can be shown to satisfy the "almost double-log" conditions.

Using the notation and arguments of Sato (1972) we define:

q_i = quantity of the i^{th} good $\quad i=1,\ldots n$

p_i = price of the i^{th} good $\quad i=1,\ldots n$

y = total expenditure = $\sum_{i=1}^{n} p_i q_i$

Θ_i = budget share of good i, = $\frac{\Sigma p_i q_i}{y}$; $\sum_{i=1}^{n} \Theta_i = 1$

η_i = income elasticity of good i, $= \frac{\partial q_i}{\partial y} \cdot \frac{y}{q_i}$ $\quad i=1,\ldots n$

λ = marginal utility of income

σ : $-\frac{1}{\sigma}$ = the income elasticity of the marginal utility of income $= \frac{\partial \lambda}{\partial y} \cdot \frac{y}{\lambda}$

Sato shows that if the utility function is additive of the form $V(q) = \Sigma f_i(q_i)$, then the price elasticities of demand can be written:

$$\frac{\partial q_i}{\partial p_j} \cdot \frac{p_j}{q_i} = \eta_i \theta_j (\sigma \eta_j - 1) \qquad i \neq j \tag{2.13.1}$$

$$\frac{\partial q_i}{\partial p_i} \cdot \frac{p_i}{q_i} = -\sigma \eta_i + \eta_i \theta_i (\sigma \eta_i - 1) \tag{2.13.2}$$

Examining equation (2.13.1) we note that the cross-price elasticity of demand for good i with respect to the price of good j can effectively be ignored if the budget share of good j (θ_j) is small. Similarily, if the own-budget share of a good (θ_i) is small then equation (2.13.2) states that the own-price elasticity of demand will be proportional to the income elasticity of demand with the factor of proportionality given by the inverse of the elasticity of marginal utility of income.

In general, both σ and η_i (and thus the own-price elasticity of demand) will not be constant. It is however possible to constrain each own-price elasticity to be almost constant at the value σ_i. Thus σ and η_i are constrained to satisfy $\sigma_i = -\sigma\eta_i$ everywhere. These constraints implicitly define the utility function:

$$V = \sum_{i=1}^{n} c_i q_i^{-\rho_i} \quad ; \quad \rho_i = (1-\sigma_i)/\sigma_i \quad ; \quad c_i\left(1-\frac{1}{\sigma_i}\right)>0 \tag{2.14}$$

Further, since Engel's aggregation yields:

$$\Sigma\eta_i\Theta_i = 1$$

it follows that:

$$\sigma = \Sigma\sigma_i\Theta_i \qquad (2.15)$$

In general, σ is still not constant. However, if the constants σ_i are all close to the same size or if Θ_i, the budget shares, are relatively constant, then σ will be almost constant.

Consider now the special case where utility is defined over three goods (q_1, q_2, q_3). Assume that the budget shares of goods q_1 and q_2 are small and stable and that q_3 is a composite commodity. If we adopt the utility function defined by (2.14) and use the results presented in (2.13.1), (2.13.2) and (2.15) we can almost exactly write the demands for q_1 and q_2 in double-log form as:

$$\ell n q_i = \ell n \alpha_i - \sigma\eta_i \, \ell n(p_i/p) + \eta_i \, \ell n(y/p) \qquad i=1,2 \qquad (2.16)$$

where p, the price index, and $-\sigma$ are defined by:

$$p \simeq p_3 \qquad (2.17.1)$$

$$\sigma \simeq \sigma_3 \qquad (2.17.2)$$

Finally, the near constancy of σ implies that the marginal utility of income function can be written:

$$\ell n \, \lambda \simeq \ell n k - \frac{1}{\sigma} (\ell n \, y - \ell n p) - \ell n \, p \qquad (2.18)$$

where k is independant of prices. It will be noted that λ is homogeneous of degree (-1) in prices and incomes as required.

An Exact Specification

In the last section, we provided a demonstration that under some conditions, double-log demands, which are desirable from the point of view of estimation and computation, are almost exact. In this section we introduce the results of the last section into the welfare maximization model to illustrate the final form of the model.

We begin by noting that when (2.16) is substituted into (2.8) and (2.10) the distributional coefficients for goods 1 and 2 can be written:

$$R_i = \frac{kp^{1-\frac{1}{\sigma}} \int_0^{\infty} y^{\eta_i - \frac{1}{\sigma}} f(y)\, dy}{\int_0^{\infty} y^{\eta_i} f(y)\, dy} \qquad (2.19)$$

It will be noted that R_i is independant of p_1 and p_2. Similarily , the ratio R_1/R_2 is independant of all prices and scale (k). Turning our attention to equation (2.11) we assume that marginal costs for goods 1 and 2 are given by c_1 and c_2.
Equation (2.11) can be then re-written as:

$$\frac{\dfrac{(p_1 + p_1/\varepsilon_1 - c_1)}{p_1}}{\dfrac{(p_2 + p_2/\varepsilon_2 - c_2}{p_2}} = \frac{\varepsilon_2}{\varepsilon_1}\frac{R_1}{R_2} \qquad (2.20)$$

where the R_i are given by (2.19) and the ε_i are constant and given by equation (2.16) as $-\sigma\eta_i = \varepsilon_i$.

At this point it is useful to present a diagrammatic representation of the problem. We will tailor this discussion to the case of Bell Canada so that the numerical results of the sections which follow can be visually interpreted. In particular, we will assume that good 2 (later identified as local residential services) is price inelastic and that good 1 (later identified as toll residential services) is price elastic. Since $\varepsilon_i = -\eta_i\sigma$, it follows that good 1 is more of a luxury than good 2 and therefore that $R_1 < R_2$. Demands are assumed to be given by (2.16) and for ease of exposition in the diagrammatic case, the marginal costs are assumed constant at the levels c_1 and c_2.

The equation of the indirect indifference curve for utility level $\bar{V}$ is given by:

$$N \int_0^{\infty} V(p_1, p_2, p, y)\ f(y)\ dy = \tilde{V}(p_1, p_2, p, N) = \bar{V} \qquad (2.21)$$

Clearly $\tilde{V}$ is a quasi-convex function of (p_1, p_2), and the indirect indifference curves in (p_1, p_2) space are convex to the origin with direction of improvement towards the origin. Similarily, the preceding arguments guarantee that the iso-profit contour given by $\Pi(p_1, p_2, p) = \Pi_0$ takes the general form of a parabola with minimum at $p_1^M = c_1/(1+1/\varepsilon_1)$ which is the unconstrained profit maximizing price for good 1. The curves Π_0 and $\bar{V}$ are drawn in Figure 1. Higher iso-profit contours lie to the north of contour Π_0. Similarily, that part of the contour Π_0 corresponding to prices of good 1 in excess of p_1^M is unimportant since a rational social manager could move to the left of p_1^M and lower both prices and thereby raise welfare.

The equilibrium point is given by E where the indirect indifference curve is just tangent to the profit constraint. The second-order conditions require that the iso-profit constraint lie everywhere inside the indirect indifference curve.

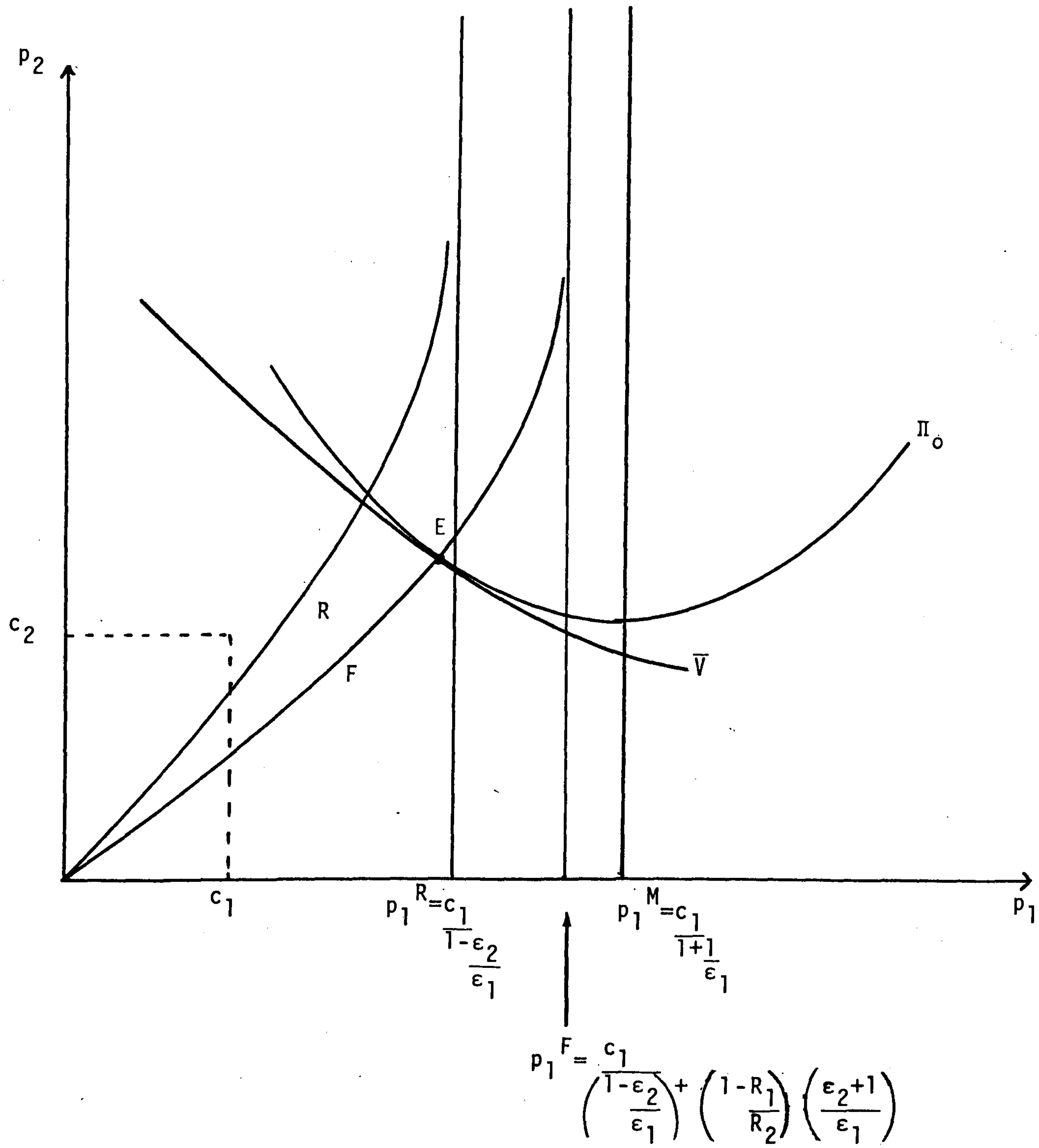
F (for Feldstein) in Figure 2.1 is the 'efficiency-equity' price locus defined by equation (2.20). At this point, it is useful to re-write (2.20) as:

$$p_2 = \frac{Rc_2 p_1}{p_1 \left(1 - \frac{\varepsilon_1}{\varepsilon_2}\right) + p_1 (R-1) \frac{(\varepsilon_2 + 1)}{\varepsilon_2} + \frac{\varepsilon_1}{\varepsilon_2} c_1} \qquad (2.22)$$

where $R = R_1/R_2 < 1$. From (2.22) it is clear that the Feldstein

FIGURE 2.1

FELDSTEIN (EFFICIENCY-EQUITY) AND RAMSEY LOCI



locus will go through all equilibrium points such as E and will be asymptotic to the line $p_1^F = c_1 / \left\{ (1-\varepsilon_2/\varepsilon_1) + (1-R)\ (\varepsilon_2+1)\varepsilon_1 \right\}$. As $R \rightarrow o$ or alternatively, as the equity importance of the relative necessity increases or the relative luxury decreases, $p_1^F \rightarrow p_1^M$. However, as $R \rightarrow 1$, equity becomes less important and the Feldstein locus converges on the Ramsey locus given by R. It may seen odd at first that strong equity weights push one towards the profit maximization point. However, this can be explained by the fact that good 2 is inelastic and the price of good 2 successively decreases as the profit maximization point is approached.

3- Cost Model

In this section we discuss the econometric cost model which will be used to estimate the characteristics of the Bell Canada production process. The estimates are later introduced into the pricing model. We begin with a discussion of our motivations in chosing a translog specification, and continue with an investigation of the properties of the model and the restrictions placed upon the estimating model by economic theory. We then turn to a brief analysis of how the estimated model can be used to test various hypotheses concerning economies of scale and scope and other properties of the underlying production process. We conclude with a discussion of the estimated model and its properties.

Specification

Background

In this paper we have chosen to model the Bell Canada production process over the period 1956-1978 with a three-input three-output translog cost function. By selecting a cost function we have made the implicit assumption that Bell Canada will choose inputs of capital, labour and materials so as to minimize the cost of producing any output vector. We further assume that of the three classes of service outputs of Bell Canada, message toll and toll private line services are supplied to firms and consumers at a rate which maximizes profits whereas local services to firms and consumers are supplied at a rate which just exhausts demand at the regulated price. The implication of our output assumptions is that regulation is effective only for local services and that even though message toll prices are in principle regulated, this regulation does not form a binding constraint.[1] It is further assumed that no important distortions arise due to A-J type rate-of-return constraints. Thus, optimality is examined in the absence of rate-of-return constraints.[2] We do not model the accumulation of capital in Bell Canada. We do however assume that capital service flows are instantaneously optimal at each data point.[3] Finally, we assume that planning and forecasting within Bell Canada are accurate and therefore that all factors adjust to their optimal levels in the year between time series observations. Our estimated cost function is therefore long-run in form.

The translog cost function is sufficiently general to allow testing of restrictions on the functional form. For example, one can directly test whether the cost function is significantly different from Cobb-Douglas. As well, the translog cost function is linear in factor and output revenue cost shares. This feature is important when the equations of large models such as this are all estimated simultaneously.

The Model

The symmetric translog cost function is written:

$$\ell n\ (COST) = C_0 + \sum_i C_i\ \ell n\ X_i + .5 \sum_i\sum_j C_{ij}\ \ell n\ X_i\ \ell n\ X_j \qquad (3.1)$$

where: $i,j\ \varepsilon(w,r,v,QL,QM,QP,T(i))$

and $C_{ij} = C_{ji}$ by assumption

In the cost function, (X_w, X_r, X_v) are respectively the factor prices (w,r,v) for labour, capital services and material. Similarily, (X_{QL}, X_{QM}, X_{QP}) are respectively the outputs (QL,QM,QP) of local, message toll and private line services. Cost is defined at each point in time by COST =wL+rK+vM where (L,K,M) are the inputs of labour, capital services and material. A complete description of the data can be found in Appendix 1.

The T's are technical change variables and their specification requires some elaboration. Very little is known ex-ante about the way in which technological change has affected costs. Clearly some of the technical improvement is embodied in the capital stock and directly enters the cost function through the user cost of capital (r). Unfortunately, there is not sufficient additional information to construct an exact hedonic constant-quality capital index which could be compared to the capital series supplied by Bell Canada. In addition, learning-by-doing type arguments can be introduced to support an argument that all service outputs and factor inputs have had certain amounts of cost savings associated with them over time. As well, it is not unreasonable to suppose that there may have been some Hicks neutral technical change. Finally, proxy indicators of technical change for all inputs and outputs are not available. Excluding a simple time trend, three indicators of technological change are readily available from the data in the public domain: NEW - an index of the percent of main stations switched by new technology including cross-bar and electronic switching, ACCESS - an index of the percent of phones with access to direct-distance dialing, and PDPH - an index of the percent of phones which dial.

Given the constraints discussed above, the following approach was adopted to introduce technical change into the cost function. First, it was noted that potentially five separate indices of technical progress could enter the cost function. These include, one for neutral improvements (To), one for all factors (T_F) (since homogeneity of degree 1 of the cost function in factor prices and and the translog specification rule-out different technical change variables for each input) and three separate technical change variables for outputs - one each for local (T_{QL}), message toll (T_{QM}), and private line (T_{QP}). Since it was reasonable to expect the separate technical change indices to be related to the general indices (NEW, ACCESS, AND PDPH) and further given that the exact relationship in each case was unknown, a Box-Cox (1964) technique was used to simultaneously fit the cost parameters and the parameters determining the functional form of the separate technology variables. Thus, the technology variables were defined by:

$$T_k = \alpha_k\left[\frac{NEW^{A_k}-1}{A_k}\right] + \beta_k\left[\frac{ACCESS^{B_k}-1}{B_k}\right] \quad (3.2)$$
$$(1-\alpha_k-\beta_k)\left[\frac{PDPH^{C_k}-1}{C_k}\right]$$

k=0 - Hicks Neutral

F - Factors

QL - Local output

QM - Message toll output

QP - Toll private line output

Although at this stage it might appear that (too) many new parameters have been added to the model, this turned out not to cause any problem during estimation. In particular, there was no significant loss in the explanatory power of the model when many of the variables were preassigned a value of zero.

Share Equation for Inputs and Outputs

Logarithmic differentiation of the cost function with respect to inputs and the application of Shephard's Lemma (1953) leads to the following factor share equations:

$$\frac{\partial \ \ell n \ COST}{\partial \ \ell n \ X_i} = S_i = C_i + \sum_j C_{ij} \ \ell n \ X_j \qquad (3.3)$$

$i \ \varepsilon(w,r,v)$; S_i = cost share of factor i

$j \ \varepsilon(w,r,v,QL,QM,QP,T_F)$

Similarily, assuming that the cross-elasticities between service outputs are zero and since profit maximization with respect to message toll and private line services implies $MR_k = p_k(1+\varepsilon^{-1}_k) = MC_k$, it is possible to write the marginal revenue 'share' equation as:

$$\frac{p_k(1+\varepsilon_k^{-1}) \cdot q_k}{COST} = C_k + \sum_j C_{kj} \ \ell n \ X_j \qquad (3.4)$$

$k \ \varepsilon(QM,QP)$

$j \ \varepsilon(w,r,v,QL,QM,QP,T_k)$

Restrictions Arising from Economic Theory

The cost function is assumed to arise from the process of minimizing the cost of producing a given vector of outputs subject to a production function constraint. The minimization guarantees that the cost function will be a (non-strictly) concave function of factor prices. The non-strictness arises from the fact that the cost function must also be homogeneous of degree 1 in factor prices. In terms of the cost function introduced in equation 3.1, the following parameter restrictions are implied by homogeneity:

$$\sum_i C_i = 1 \qquad i \quad \varepsilon(w,r,v)$$

$$\sum_j C_{ij} = 0 \qquad i,j \ \varepsilon(w,r,v)$$

$$\sum_i C_{ik} = 0 \qquad i \quad \varepsilon(w,r,v)$$

$$k \quad \varepsilon(QL,QM,QP,T_F)$$

Since these restrictions are equivalent to having the factor cost shares add to unity, in order to estimate the model it is necessary to 'drop' one of the factor share equations during estimation. It is also customary to re-write the restrictions in terms of the coefficients associated with one of the factors - in this case, materials. The coefficients of the materials variables are later calculated along with their standard errors.

Well-Behavedness Properties

It was noted above that the cost function must be (weakly) concave in factor prices. This is not guaranteed by parameter restrictions and must be verified at each data point. Similarily the second-order necessary and sufficient conditions corresponding to the assumption of profit maximization with respect to message toll and private line services must be verified at each data point. These latter conditions require that the Hessian matrix of the profit function in QM and QP be negative definitive or equivalently, that the profit function is concave in (QM,QP). It must be stressed that if a cost model violates either of the concavity conditions described above, it will have only limited usefulness for policy analysis or forecasting. It is well-known that pure time-series models will provide at least as good fits of the data as economic models and often yield more accurate short-run forecasts.

Characteristics of the Technology

Although the principal goal of the cost model in this paper is as an input to the optimal pricing model, it is also useful to examine some of the characteristics of the production technology. For example, the marginal costs of the services can be evaluated at each data point according to the formula:

$$MC_k = \left(\frac{COST}{X_k}\right)\left(C_k + \sum_i C_{ik} \ \ell n \ X_i\right) \qquad (3.5)$$

$$k \ \varepsilon (QL, QM, QP)$$

$$i \ \varepsilon (QL, QM, QP, w, r, v, T_k)$$

A measure of 'ray' scale elasticity (which is the inverse of the ray cost elasticity) is defined by:

$$RSCALE = \left[\sum_k MC_k \frac{X_k}{COST}\right]^{-1}, \qquad k \ \varepsilon (QL, QM, QP) \qquad (3.6)$$

The RSCALE measure represents the constrained elasticity of outputs with respect to inputs along a ray drawn from the point of evaluation. It is also possible to examine (marginal) cost complementarities by evaluating the expression:

$$\frac{\partial MC_k}{\partial X_i} = \frac{MC_i MC_k}{COST} + \frac{C_{ik} \ COST}{X_i X_k} \qquad i \neq k; \ i,k \ \varepsilon (QL, QM, QP) \qquad (3.7)$$

It has been shown elsewhere (LeBlanc (1979)) that a sufficient local condition for economies of scope to exist is that, at each data point, the cost complementarity terms are significantly negative.

The Simultaneity Issue

Referring back for the moment to the marginal revenue share equation, it will be noted that properties of demand are necessary to estimate these equations. The problem is simplified somewhat if one assumes that the own-price elasticities of demand for message toll and private line services are constant - or that the demand curves are isoelastic. However, the fact remains that the equilibrium condition $MC_k = MR_k$ has the same econometric implications for simultaneity bias as does any demand and supply market model. It is therefore important, in general, to estimate the cost model equations with the demand equations to obtain a simultaneous estimate of the price elasticity.

With this thought in mind, we undertook some initial estimations of the model.[4] The preliminary results showed that even though the cost model was generally stable, the point estimates of the elasticity of demand for message toll and competitive services varied dependant upon the definition of income variables in the demand equations. We therefore decided to estimate the cost model alone using a grid of elasticity values for message toll (-1.2, -1.8) and competitive services (-1.25, -5.0). The range of elasticities was determined by the preliminary estimations. The adoption of this strategy significantly reduced the computational snags and time needed to estimate the model. This approach effectively builds a sensitivity analysis into the model.

4- Estimation of the Model and Properties of the Estimates

Estimation Results

In this section we report on estimation of the cost model given by the cost function, the factor share equations for labour and capital and the marginal revenue share equations for message toll services and toll private line services. We present only the results when the price elasticity of message toll services (E_{QM}) is -1.5 and the elasticity of private line services (E_{QP}) is -2.0. These elasticities effectively represent the mid-point of the ranges of these elasticities and the qualitative results are representative of those obtained at different elasticity values.

Finally, the parameter estimates presented here reflect some nested testing which has taken place. In the first place, we could not reject the hypothesis that the cost share of materials was constant throughout the sample at the level $C_V = .195$. Incorporating this result as prior information meant that only one factor share equation needed to be estimated along with the cost function and marginal revenue share equations. Further testing has shown that the likelihood of the model is not significantly reduced if the restriction that outputs are separable from inputs is imposed. As well, many of the other parameters within the factor share equations were found to be not significantly different from zero. The parameter estimates are presented in Table 3.1. If a parameter does not appear then it is to be assumed that it has been tested and found not significantly different from zero. Some equation by equation results are presented in Table 3.2. The resulting model is Cobb-Douglas in inputs with factor shares adjusting over time due to technical change. The outputs, although separable from inputs, are not Cobb-Douglas. The interactions between outputs are too complicated to be captured by a Cobb-Douglas model.

Properties of the Estimated Cost Model

In this section we report the important characteristics of the estimated cost model and the implications for the underlying production technology.

At each data point, the cost function is a concave function of factor prices as required. As well, the second-order conditions of profit maximization with respect to message toll and competitive services are satisfied at every data point. The model is inconclusive with respect to economies of scope. The scale elasticity measure is significantly greater than 1, the sizes (10^{-3}) and signs of the cost complementarity terms are not such that economies of scope can be deduced for all outputs. There is some trending in the (ray) scale elasticity measure, however this trending reduces towards the end of the sample when the scale elasticity takes a value of 1.52 in 1978. The marginal costs of message toll and toll private line services closely track the marginal revenues of these services. The marginal cost of local shows some upward trending - achieving a value of 1.36 in 1978.[5] Since the factor part of the cost model is effectively Cobb-Douglas, the elasticities of substitution of all factors are unity.

5- Demand Specification

In this section we briefly review the implications of sections 2 and 3 for the demand equations used within the simulation model. We then derive the final forms of the demand equation for simulation purposes.

Given that the utility function is additively separable, it was shown in section 2 that the demands for services would be almost double-log (with constant income and own-price elasticities)

TABLE 3.1

PARAMETER ESTIMATES OF THE COST MODEL

Parameter	Estimate	Standard Error
C_o	6.961	1.185
C_w	.270	.004
C_{wT}	-.183	.018
B_F	1.222	.117
C_r	.535	.004
C_{rT}	.183	.018
C_{QL}	-.846	.375
C_{QLQL}	.296	.067
C_{QLT}	.045	.012
B_{QL}	.589	.145
C_{QM}	.343	.071
C_{QMQM}	.053	.015
C_{QMQL}	-.081	.024
C_{QMT}	.078	.025
B_{QM}	2.005	.484
C_{QP}	.136	.018
C_{QPQP}	.022	.003
C_{QPQL}	-.030	.005
C_{QPT}	.234	.103
C_{QP}	53.892	27.383

TABLE 3.2

EQUATION BY EQUATION SUMMARY

	R^2	Durbin-Watson	Sum Of Squared Residuals
COST FUNCTION	.9998	1.3332	.0024
LABOUR SHARE	.9786	1.3773	.0007
MESSAGE TOLL SHARE	.9343	1.3103	.0002
PRIVATE LINE SHARE	.9891	.9334*	.00001

*There was no systematic pattern to the residuals of this equation suggesting that the size of the D-W statistic was spurious.

if the expenditure share of the service was small. For Quebec and Ontario (the centre of almost all Bell Canada activities) the share of residential local services in total consumer expenditure is approximately .005. Residential toll services yield an even smaller share. Thus, one can feel reasonably confident about the double-log specification for residential local and message toll services.

Local Residential Demand

Guided by equation 2.16, we write the demand for local residential services as:

$$\ell n\left(\frac{QL^R}{POP}\right) = \alpha_0 + \varepsilon_{QL}\ \ell n\left(\frac{P_L^R}{P}\right) + \eta_{QL}\ \ell n\left(\frac{y}{P \cdot POP}\right) + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 \quad (5.1)$$

The variables have the following definitions:

QL^R =local residential service output

P_L^R =price of local residential services

P =consumer price index

y =sum of gross provincial outputs for Quebec and Ontario

POP =population in the Bell Canada territory

D_1 =step variable for introduction of direct distance dialing in 1959

D_2 =step variable for the introduction of the one minute minimum toll call in 1971.

D_3 =step variable for rate centre shifting in Toronto, 1976.

It will be noted that, consistent with equation (2.16), outputs and income are expressed on a "per-person" basis. Gross provincial products are used as a proxy for personal disposable income in each province. Finally, it will be recalled that ε_{QL} (the own-price elasticity) and η_{QL} (the income elasticity) are not free. They are related by the identity:

$$\varepsilon_{QL} = -\sigma\eta_{QL} \tag{5.2}$$

where $\frac{-1}{\sigma}$ is the elasticity of marginal utility of money.

In Section 6, the choice of σ will be discussed in greater detail. It is sufficient to note here that the efficiency-equity price simulations are conditional upon a choice from a range of possible values of σ.

Toll Residential Demand

An unfortunate feature of the toll demand model was the absence of a time series breakdown of residential message toll prices and quantities.[6] In order to derive a message toll residential quantity series from the available message toll residential revenue series, it was assumed that prices were equal to the aggregate message toll price at each data point.

With regard to the estimating equation for toll residential services, it will be recalled that the cost model was to be estimated over a range of aggregate demand elasticities for message toll and private line services. Since we had created the residential toll quantity series by assuming that the residential price series was equivalent to the aggregate price series, it did not seem useful to assume that the residential own-price demand elasticity differed from the price elasticity of demand for the aggregate of message toll services. In addition, since the residential message toll income and own-price demand elasticities are related by:

$$\varepsilon_{QM} = -\sigma\eta_{QM} \tag{5.3}$$

it was not necessary to estimate any parameters of toll residential demand. Once the aggregate price elasticity is specified for a given simulation, ε_{QM} is known and, conditional upon σ, η_{QM} is known as well. Formally, the residential toll demand can be written:

$$\ell n\left(\frac{QM^R}{POP}\right)_t = \alpha(t) + \varepsilon_{QM}\ \ell n\left(\frac{P_{QM}^R}{P}\right)_t + \eta_{QL}\ \ell n\left(\frac{y}{P \cdot POP}\right)_t \qquad (5.4)$$

where QM^R =the quantity of residential message toll services

P_{QM}^R =the price of residential message toll serves (equal to the price of aggregate message toll services)

P,y and POP are defined above.

$\alpha(t)$ is a forcing function which guarantees that at each point in time the right-hand side of (5.4) is equal to the logarithm of historic residential demand per person.

6- Simulation

In this section we begin by noting some additional assumptions which are necessary to simulate the optimal prices. We then work through a conceptual simulation experiment and point out where sensitivity tests are conducted. We conclude the section with a set of simulation results which are presented and analyzed.

Additional assumptions

1) It is necessary to choose a range of values for $\frac{-1}{\sigma}$ or the elasticity of marginal utility of income. Many authors have suggested ranges (see, for example, Baumol (1979), Baumol and Bradford (1970), Fellner (1967), Mera (1969), Powell et al (1968), Sato (1972), and Maital (1973). There appeared to be some early agreement on a value of -1.5. More recently, however, Davies (1980) has studied evidence which suggests much higher values of the elasticity of marginal utility of income. In this paper we have chosen the range (0,-2.0). This range appears consistent with almost all researchers.

2) It is assumed that the logarithm of income is distributed as normal with mean $\overline{\ell n y}$ and variance $\sigma^2_{\ell n y}$. The log normal distribution provides a reasonable description of the distribution of income in Ontario and Quebec. As well, the distributional coefficients R_{QL} and R_{QM} can be evaluated without numerical integration. In particular, if y is lognormally distributed then, for any Θ,

$$\int_0^\infty y^\Theta f(y)dy = \int_0^\infty \exp\left[\Theta \ell n\ y\right] f(y)dy \qquad (6.1)$$
$$= \exp\left[\Theta \overline{\ell n\ y} + .5\Theta^2 \sigma^2_{\ell n y}\right]$$

The mean and variance of the logarithm of household income were calculated for 1961 and 1971 using Statistics Canada data (cat.98-505, 93-749). The variances of the two years were almost the same and we therefore assumed the variance constant at the level .72. However, there was some increase in the mean from 8.1415 in 1961 to 8.6139 in 1971. A complete series of means for 1956-1978 was created by interpolation (1956-1971) and extrapolation (1972-1978) using the growth rates of gross provincial product per household to approximate changes in the mean over time.

3) The last assumption relates to the way in which the cost model information was introduced into the simulation experiments. A time series of marginal costs was obtained for the aggregate series of local and message toll services from the cost model. It was assumed that these point estimates of marginal cost were accurate in a neighbourhood of the quantities at which they were estimated. In the simulations, the point estimates of historic marginal cost were taken as the levels of constant marginal cost. There are two reasons why this assumption is reasonable. In the first place, the historic time series of marginal costs derived from the cost function can be used to show that, for any given year, marginal costs of services are extremely inelastic with respect to both own-service quantities

and other service quantities (cost complementarity). Secondly, if the marginal costs are constant at the historic levels then the historic levels of business message toll services remain optimal even if the residential component of local and message toll services changes. Thus, using the constant marginal cost assumption in a neighbourhood of the data points means that it is not necessary to include a reoptimization of business message toll services in the simulations.[7]

The Simulation Process - A Conceptual Exercise

In this section we present a brief discussion of the flow of the simulation process for any given simulation experiment. In so doing, we hope to facilitate interpretation of the simulation results.

We begin by noting that there was a three dimensional array of initial conditions upon which the simulations were conditional. These conditions arose from the assumptions regarding the ranges of the own-price elasticities of message toll and toll private line services and the elasticity of marginal utility of income. In particular, it will be recalled that ε_{QM} was given the range (-1.2, -1.8), ε_{QP} was given the range (-1.25, -5.0) and $\frac{-1}{\sigma}$ was given the range (0, -2.0). Our benchmark vector $\left(\varepsilon_{QM}, \varepsilon_{QP}, \frac{-1}{\sigma}\right)$ was taken as (-1.5, -2.0, -1.5). In this section, we will work through the simulations of the benchmark case.

Step 1 - Demand Paramters

Given $\left(\varepsilon_{QM}, \varepsilon_{QP}, -\frac{1}{\sigma}\right) \equiv (-1.5, -2.0, -1.5)$, we begin by using equation (5.3) to calculate the income elasticity of demand for residential message toll services. In this case we arrive at the value $\eta_{QM}=2.25$. Taking ε_{QM} and η_{QM} we use (5.4) to solve for $\alpha(t)$ for each series year (1956-1978).

We next use the fact that $\frac{-1}{\sigma} = -1.5$ in (5.2) and estimate the local residential remand equation (5.1) subject to (5.2) to obtain estimates of the price and income elasticities of local residential demand. In this case we obtain $\varepsilon_{QL}=-.445$ and $\eta_{QL}=.668$.

Step 2 - Distributional Coefficients

Given that the set $\left(\varepsilon_{QM}, \eta_{QM}, \varepsilon_{QP}, \varepsilon_{QL}, \eta_{QL}, \frac{-1}{\sigma}\right)$ is known, it is possible to use the information on the means and variances of the lognormal distribution of income along with equations (6.1) and (2.19) to compute the distributional coefficients for 1956 to 1978. In Table 6.1 the distributional coefficients corresponding to the benchmark parameter set $\left(\varepsilon_{QM}, \eta_{QM}, \varepsilon_{QP}, \varepsilon_{QL}, \eta_{QL}, \frac{-1}{\sigma}\right) = (-1.5, 2.25, -.445, .668, -1.5)$ are displayed.

Step 3

In this final stage of the simulation process, we compute the optimal efficiency-equity prices for residential local and message toll services. The parameter set of Step 1 is augmented by the distributional coefficients (for each year) computed in Step 2. The cost model of section 3 is then estimated conditional upon the parameter set values for ε_{QM} and ε_{QP}. The estimated cost function and the demand equations of section 5 are then combined (subject to the constant marginal cost restriction) to define the profit function of equation (2.20). The yearly efficiency-equity prices result as a simultaneous solution to equations (2.20) and (2.6) where the constraining profit level (Π_0) for any year was the historic level of profits. The results are presented with the optimal efficiency-equity prices expressed as a ratio of the historic prices.

For the Ramsey case the same procedure is followed except that the influence of distributional considerations is removed by the

TABLE 6.1

VALUES OF THE DISTRIBUTIONAL COEFFICIENT* (x 10^{-5})

BENCHMARK CASE

$(\varepsilon_{QM}, \eta_{QM}, \varepsilon_{QP}, \varepsilon_{QL}, \eta_{QL}, \frac{-1}{\sigma}) \equiv (-1.5, 2.25. -2.0, -.445, .668, -1.5)$

	R_{QL}	R_{QM}
1956	.5564	.1008
1962	.5195	.10941
1967	.3546	.0642
1972	.2383	.0431
1978	.1089	.0197

*Calculated using equations (6.1) and (2.19) preassignment $R_{QL} \equiv 1 \equiv R_{QM}$. It will be noted, however, that the Ramsey case is still conditional upon the choice of $\frac{-1}{\sigma}$ through the demand elasticities.

The Simulation Results

In this section we present the simulation results for the benchmark case. The effects of changes in the benchmark parameterization are then discussed. It might be helpful for the reader to refer back to Figure 2.1 and the related discussion.

The benchmark results are presented in Table 6.2 for the Ramsey case and a range of assumptions about the elasticity of marginal utility of income.

One very noteworthy result is that large movements away from the historic levels for the price and quantity of local residential do not appear to be desireable when equity considerations are introduced. Not surprisingly, the Ramsey (regressive) case calls for the largest movements. However, maximum movements in price and quantity of residential local services of about 2% are desireable when the elasticity of marginal utility of income is less than or equal to -1.5. As discussed earlier, this range and hence these percentage changes seem to be consistent with most of the equity-weighting economics literature.

Rather larger movements away from the historic quantities and prices for residential message toll services appear desireable. As in the local case, the greater the equity weighting, the closer the optimal prices and quantities are to the historic levels. It is useful to study these results more carefully for 1978 since these results are the most relevant for the present. Depending upon equity considerations ($\frac{-1}{\sigma}$ = -1.5 or -2.0) it is optimal to lower

TABLE 6.2

BENCHMARK EQUITY - EFFICIENCY SIMULATION RESULTS

$\varepsilon_{QP} = -2.0, \quad \varepsilon_{QM} = -1.5$

	Historic* Prices and Quantities and Marginal Costs		Ratio of Optimal Equity-Efficiency Prices and Corresponding Quantities to Historic* Prices and Quantities			
	P_{QL}^{R}	MC_{QL}	RAMSEY**	$-\frac{1}{\sigma} = -1.0$	$-\frac{1}{\sigma} = -1.5$	$-\frac{1}{\sigma} = -2.0$
1956	.9032	.3268	1.1912	1.0516	1.0080	1.0003
1962	.9872	.4819	1.2242	1.0624	1.0096	1.0004
1967	1.0000	.5501	1.2764	1.0786	1.0119	1.0006
1972	1.0529	.7608	1.3871	1.1156	1.0175	1.0008
1978	1.3399	1.3717	1.5841	1.1843	1.0285	1.0013
	R_{QL}					
1956	103.590		.9250	.9720	.9965	.9999
1962	142.704		.9139	.9664	.9957	.9998
1967	195.921		.8970	.9582	.9948	.9998
1972	267.854		.8644	.9401	.9923	.9997
1978	373.393		.8148	.9090	.9876	.9995

TABLE 6.2 (continued)

BENCHMARK EQUITY - EFFICIENCY SIMULATION RESULTS

$\varepsilon_{QP} = -2.0, \quad \varepsilon_{QM} = -1.5$

Historic* Prices and Quantities and Marginal Costs			Ratio of Optimal Equity-Efficiency Prices and Corresponding Quantities to Historic* Prices and Quantities			
	P^R_{QM}	MC_{QL}	RAMSEY**	$-\frac{1}{\sigma} = -1.0$	$-\frac{1}{\sigma} = -1.5$	$-\frac{1}{\sigma} = -2.0$
1956	1.0650	.3628	.4295	.6219	.8119	.9658
1962	1.0414	.3503	.4094	.5908	.7890	.9498
1967	1.0000	.3223	.3879	.5562	.7505	.9093
1972	1.1019	.3599	.3807	.5385	.7448	.9175
1978	1.3437	.4551	.3785	.5235	.7463	.9433
	Q^R_M					
1956	32.9018		3.5533	2.0390	1.3670	1.0537
1962	56.5391		3.8176	2.2023	1.4271	1.0804
1967	97.2000		4.1401	2.4111	1.5383	1.1535
1972	171.629		4.2568	2.5308	1.5555	1.1379
1978	364.492		4.2952	3.6400	1.5508	1.0916

*In all cases historic prices were used for comparison purposes. As well, historic quantities of residential message toll services were used. Because the demand equation for residential local services was estimated, the fitted values of this equation were used as the historic values for comparison purposes.

**As mentioned in the text, the Ramsey prices are determined when the distributional coefficients are constrained to equal 1. The demand parameters for the Ramsey case are conditional upon $-\frac{1}{\sigma} = -1.5$.

P^R_{QL}, P^R_{QM} refer to the prices of residential local and message toll services respectively.

Q^R_L, Q^R_M refer respectively to the residential quantities of local and message toll services.

MC_{QL}, MC_{QM} are the marginal costs.

residential toll prices between 26 and 6 percent. It must be noted as well that since residential and business message toll quantities are about equal in 1978, the percentage change in total message toll quantities would be equal to approximately one half of the percentage change in residential message toll services.

In terms of the graph presented in Figure 2.1, the simulation results suggest that the iso-profit contours (such as Π_0) are quite flat in a rather large neighbourhood of the historic prices. Thus, it is possible to have large movements in the price of residential message toll services without significant movements in the price of residential local services.

Sensitivity Results

The sensitivity of the results to changes in the benchmark ε_{QP} and ε_{QM} was also investigated.

When the elasticity of toll private line services was allowed to vary, there was almost no change in the simulated optimal efficiency-equity prices for residential local and message toll services. The explanation of this result lies in the fact that cost complementarities between toll private line services and local and message toll services were negligable.

The sensitivity of the results to changes in the elasticity of message toll services is presented in Table 6.3 for the benchmark case. It will be noted that the less elastic is the demand for message toll, the greater are the price and quantity movements. The explanation of this result is straightforward. Examining equation (5.3) we note that, ceteris paribus, the smaller the price elasticity of demand for residential message toll services, the smaller is the income elasticity of demand for residential message

TABLE 6.3

EFFICIENCY-EQUITY PRICES

SENSITIVITY ANALYSIS FOR ε_{QM}

1967

$\varepsilon_{QP} = -2.0$ $\quad -\frac{1}{\sigma} = -1.5$

ε_{QM}	P_{QL}^{R}	QL^{R}	P_{QM}^{R}	QM^{R}
-1.2	1.0579	191.074	.4298	267.809
-1.5	1.0119	194.893	.7505	149.524
-1.8	1.0026	195.693	.8753	123.342
Historic Value	1.0000	195.921	1.0000	97.200

toll services. However, as the message toll income elasticity decreases, so does the spread between the distributional coefficients for local and message toll services. Thus, as the absolute size of ε_{QM} decreases the induced movement in optimal prices is away from historic levels and towards the Ramsey prices.

Conclusions

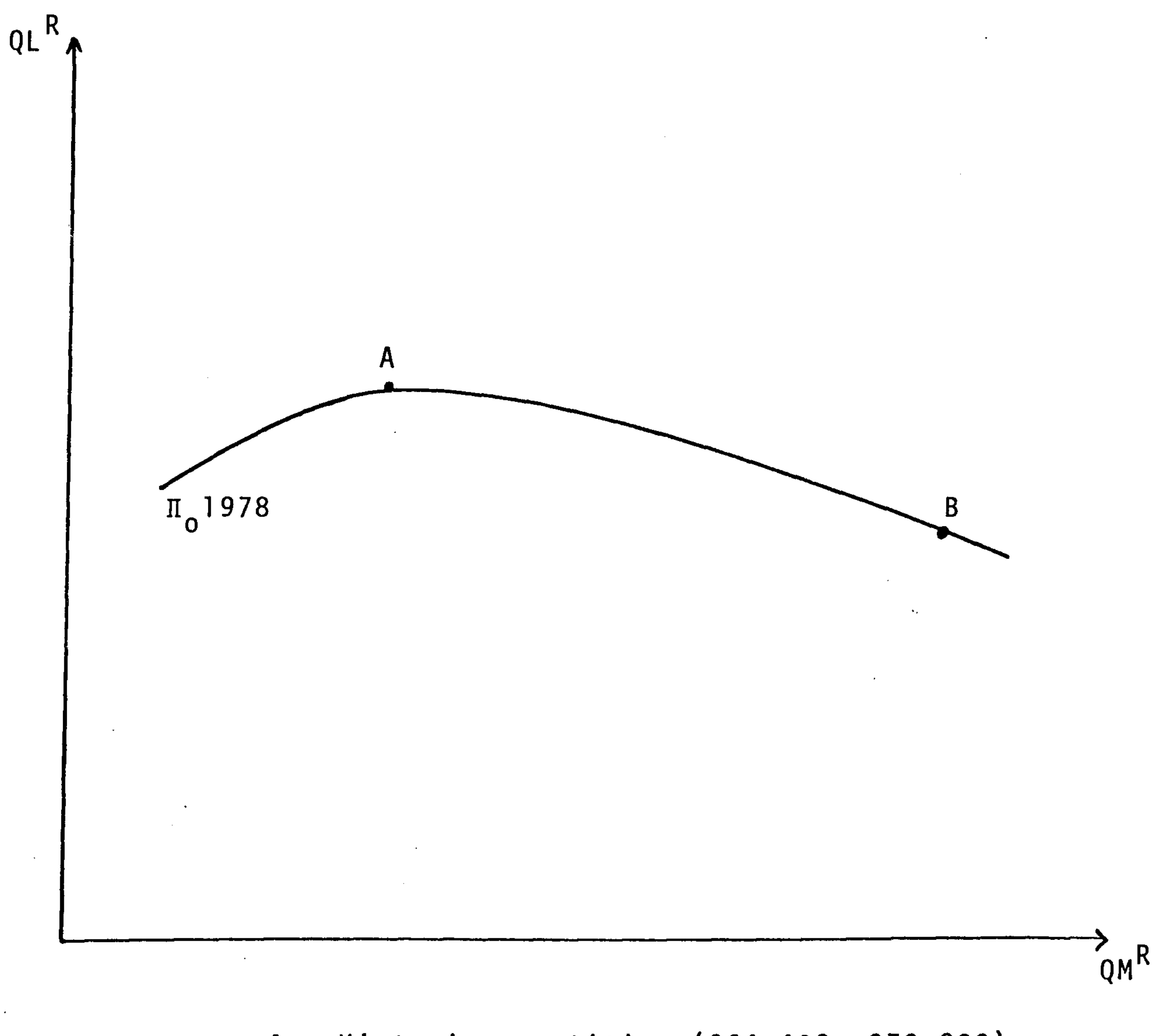
The analysis presented in this paper has considered some of the issues which must be faced in practice if one is interested in introducing both efficiency and equity criteria into the pricing decision of a regulated communications carrier. The results suggest that some adjustment in the prices (and hence quantities) of residential local and message toll services would be desirable. However, the results do not say how these optimal prices should be introduced. It is only reasonable to expect that there would have to be some gradual transition towards any new optimum. It is interesting to note that given the model advanced in this paper, it follows that if prices adjust to the optimal prices along the iso-profit constraint then following this transition path results in the welfare of residential consumers of the services increasing and the greatest increase in welfare will correspond to the first adjustments.

This result is demonstrated with reference to Figure 7.1. The locus drawn in Figure 7.1 depicts the iso-profit locus in residential output space for the year 1978. The curve drawn here is the dual to the iso-profit locus in price space (Π_0) drawn in figure 2.1. Duality insures that the historic output vector (A) occurs at a maximum of the isoprofit locus in output space and at a minimum of the isoprofit locus in price space. The Feldstein optimal point in price space is given by E whereas in output space it is given by B. Figure 7.1 is drawn under the benchmark assumption that $(\varepsilon_{QM}, \varepsilon_{QP}, -\frac{1}{\sigma}) = (-1.5, -2.0, -1.5)$.

Expressed in output space, the adoption of optimal efficiency-equity prices would involve the transition from point A to point B. We will assume that the transition follows the iso-profit locus from A to B and develope the conditions under which consumer welfare is

FIGURE 7.1

RESIDENTIAL OUTPUT SPACE ISO-PROFIT LOCUS FOR 1978



A = Historic quantities (364.492, 373.393)

B = Feldstein optimal quantities (565.254, 368.763) a concave function along the iso-profit locus and therefore the conditions under which the welfare increments associated with successive equal movements along the transition path are decreasing.

To begin, we note that the iso-profit can be written in the general form:

$$QL^R = \Pi\,(QM^R) \tag{7.1}$$

Using a prime (') to denote differentiation, we have

$\Pi'(QM^R) < o$ and $\Pi''(QM^R) \leq o$ in the range AB.

Welfare is in general given by:

$$W = W(QL^R, QM^R) \tag{7.2}$$

Along the iso-profit constraint, welfare can be written:

$$W = W(\Pi\,(QM^R),\ QM^R) \tag{7.3}$$

We therefore wish to develop the conditions under which welfare given by (7.3) is concave in QM^R. Differentiating the welfare function twice we obtain:

$$W''(QM^R) = W_{11}(\Pi')^2 + 2W_{12}\Pi' + W_{22} + W_1\Pi'' \tag{7.4}$$

where, in a standard way, superscript primes refer to total derivatives and subscript numbers refer to partial derivatives with respect to the arguments of the welfare function.

It will be recalled that our simulation results are developed in terms of a welfare function which, in terms of quantities, is additive with the properties $W_1, W_2 > o$, $W_{11}, W_{22} < o$ and $W_{12} = 0$. Thus, in terms of our model, the right-hand side of (7.4) is negative and the welfare function is concave over the transition path AB. The interpretation of this result is that, in terms of our model, even though one may be reluctant (for whatever reason) to enforce a movement all the way from A to B, one can remain confident that the first of M equal sized movements along the adjustment path will supply the greatest welfare improvement to residential users of local and message toll services.

FOOTNOTES

1) This assumption is consistent with any cross-subsidization goals of regulation whereby profits from message toll services can be used to defray losses incurred in the provision of local services. Given the jointness of production it is extremely difficult (if not impossible) to disentangle the extent to which message toll services subsidize local or any other service. In both 1978 and 1980, Bell was awarded the requested increase in intra-Bell long-distance rates.

2) The fact that we do not include a rate-of-return constraint can be justified using both general and Bell - specific arguments. Considering the general arguments first, we can find no empirical A-J study which provides any strong support for the A-J hypothesis. Modelling and measurement errors are simply too large. With regard to Bell, it would appear that historically, rates-of-return have not been strongly policed by regulatory agencies. Using a production function approach, we have shown elsewhere (Breslaw, Corbo and Smith (1979a)) that there appear to be no significant effects arising from rate-of-return regulation. Fuss and Waverman (1980) have examined the same question using a second-order translog cost function. The Fuss-Waverman results suggest that regulatory constraints are not consistent with the data. This latter result must be tempered with the realization that the second-order translog cost specification of Fuss and Waverman is also inconsistent with the existence of a binding regulatory constraint. Breslaw, Corbo and Smith (1979a) have shown that the minimum order of reasonable approximation to the cost function of a regulated firm is three.

3) Although capacity utilization questions may be important, no data were available to adjust the flow of services from measured capital stock. In this paper, it is assumed that in each year the flow of capital services in proportional to the capital stock with a factor of proportionality equal to 1.

4) The TSP version of the non-linear SURE estimator was used to estimate the cost model. Given that some data on the measures of technology were not available prior to 1956, the model was estimated for the period 1956-1978.

5) Properties of the marginal cost function and the estimated values will be discussed in greater detail in the sections on optimal pricing which follow.

6) The Canadian Department of Communications is currently engaged in deriving these price series.

7) For comparison purposes, we did compute the optimal efficiency-equity prices in the case where marginal cost was variable and no reoptimization was undertaken for business message toll and toll private line. Since the elasticity of the marginal cost functions were never identically zero, there were some movements away from the constant marginal cost solution prices. However, the movements were not large and they tended to re-enforce the movements away from the historic prices resulting from the constant marginal cost case. As well, there was no important change in the pattern of results when the isoprofit constraint was replaced by a iso-rate-of-return constraint.

APPENDIX 1

DATA

The data used in this paper were obtained from the following sources:

- -Bell Canada, Annual Charts 1935-1978
- -Statistics Canada CAT 13-213
- -Ontario Ministry of the Treasury, Quarterly Time Series 1947-75
- -Quebec Ministry of Industry and Commerce, Revenues et Depenses 1946-70
- -Interrogatories in CRTC (Canada) hearings:
  - -BELL (NAPO) 1 FEB 80 - 727
  - -BELL (CAC) 3 APR 80 - 511
- -Department of Communications (personal communication)
- -Denny _et al_ (1979).

APPENDIX 1 (continued)

Factors:

L Labour, adjusted for quality, excluding construction

K Capital, total average net stock, constant $1967

M Materials, Divisia index of materials, revenue taxes and uncollectables constant $1967

w wages, (Employee expenses and labour tax) ÷ L

r cost of capital, Hall-Jorgenson derivation, real rate-of-return 3.5%

v price, raw materials, Divisia index (see M)

Technologies:

NEW Crossbar and electronic switching as a % of central office equipment

ACCESS % phones with access to DDD (see Denny et al (1979)) for initial derivation of this series)

PDPH % dial phones

Services:

QL	Local service + miscellaneous + directory assistance	Const$1967
QM	Message toll output, including WATS (DIVISIA)	Const$1967
QP	Other toll, excluding WATS	Const$1967
QL^R	Local residential services	Const$1967
QL^B	Local business services	Const$1967
QM^R	Message toll services - residential	Const$1967
QM^B	Message toll services - business	Const$1967

Similar series in P reflect corresponding prices

Other

GPP Real gross provincial product, Ontario and Quebec

POP Population, Bell Territory

DISCUSSION OF
"AN ECONOMIC ANALYSIS OF
MEASURED SERVICE OPTIONS"

D.O. TURTON
CANADIAN PACIFIC

Question

Compared to optional flat rate (FR) - measured service (MS) billing of local exchange customers, what are the aggregate welfare implications of an option to bill customers on an ex post basis allowing them to pay the minimum of the FR charge and the MS charge plus a premium.

Conclusion of Paper

- The answer will depend on three factors:

  1. The response of demand distribution to the change in billing method (i.e. from optional FR-MS to ex post billing);

  2. The size of the premium that customers would be asked to pay in order to have the ex post billing option; and

  3. The parameters of the FR versus the MS billing methods, (namely F versus ρ and P).

- It is not possible to give a definitive answer to the question as either an increase or a decrease in aggregate welfare is possible.

- There are five strict conditions for an improvement in welfare to occur; violation of one or more of these conditions could result in a loss of aggregate welfare.

  (i) The premium payable for the benefit of ex post billing must exceed the incremental administration cost associated with offering this option;

  (ii) The measured service (MS) option is priced at its marginal cost;

  (iii) The flat rate (FR) charge under optional FR-MS billing must be such that total revenue from FR customers equals total cost of providing service to them;

  (iv) Average usage of flat rate customers facing optional FR-MS billing must be greater than average usage of flat rate customers facing either optional FR-MS billing or ex post billing; and

  (v) The number of customer-periods billed on the basis of the MS tariff must be larger under the ex post option than under the optional FR-MS option.

The first three of these conditions are related to common cost associated with providing local exchange service as well as toll service. As a result, any cost allocated to different groups of local exchange customers will be arbitrary, notwithstanding the fact that these costs can be estimated objectively.

It would appear, therefore, that no decisive answer could be had for the question of the relative merit of optional FR-MS billing versus ex post billing.

DISCUSSION OF

"THE REGULATORY PROCESS UNDER

PARTIAL INFORMATION"

D.O. TURTON

CANADIAN PACIFIC

Issue

There are several approaches in the literature on the subject of optimal pricing for a regulated natural monopoly. Two of these are the Ramsey pricing approach and the Vogelsang and Finsinger (V-F) algorithm. The authors propose an alternative algorithm which, they argue, overcomes some of the limitations of the V-F approach.

Comments

It appears that the proposed model addresses itself to the tendency of the V-F procedure to focus on the regulated utility, with an increasing cost function, to experience spells of potentially severe losses; however it does not address the more serious limitations of a statis analysis.

One of the most important aspect of the issue of optimal pricing for a regulated natural monopoly is the time dimension. In practice optimal pricing can best be described over a multiple of periods, therefore, a multiperiod model would seem to be more appropriate to dealing with this problem.

In addition, it seems that the procedure the authors suggest does not depend on the technology employed in production. As a result the prices the algorithm generate would not be constrained by operational efficiency. The model therefore appears to focus only on the distributional issue.

In order to assess the aggregate welfare implications of a pricing system the regulator may well desire to focus on both production and consumption efficiency.

DISCUSSION OF "WELFARE OPTIMAL SUBSIDY-FREE PRICES UNDER A REGULATED MONOPOLY"

D.C. TURTON
CANADIAN PACIFIC

Question

What are the welfare differences between cross-subsidized Ramsey pricing and a subsidy-free rate structure.

Comments

Relative to subsidy free prices, cross-subsidized Ramsey prices are both Pareto non-comparable and Koopman's efficient non-comparable under indicated assumptions. As a result the author uses a compensation rule to choose which alternative was socially preferable.

Using a two-output welfare optimal subsidy free model, the author comes up with some insightful findings:

1. a subsidy-free price structure is equivalent to a subsidy-free Ramsey price structure.

2. if Ramsey prices lead to cross-subsidization, then the otherwise Ramsey subsidized service would cover its incremental costs while the otherwise Ramsey subsidizing services would cover its stand alone costs.

3. if a profit constraint is binding, economies of scope are a necessary condition for the existence of subsidy-free prices.

4. If diseconomies of scope prevails then provision of service by a multi-output firm is Pareto inferior to provision on a stand alone basis.

5. If legal barriers to entry are removed, cross-subsidized Ramsey prices can encourage inefficient competition for the subsidizing services and can discourage fair competition.

6. Thus, subsidy free prices may not completely eliminate the possibility of non-optimal market structure; however, it is a necessary condition to the sustainability of a natural monopoly or to the existence of fair competition.

Comment

As a cartoon model this analysis provides some very interesting insights. The question that the paper poses seems to be highly relevant from a practical standpoint and, despite the simplifying assumptions, the insights developed should provide regulators with an improved understanding of the issues presented. Of particular interest is the result that a subsidy-free rate structure may not be Pareto superior to cross-subsidized Ramsey pricing.

DISCUSSION OF "EFFICIENCY, EQUITY AND REGULATION A MODEL OF BELL CANADA"

D.O. TURTON

CANADIAN PACIFIC

Issue

A fundamental goal of public policy in Canada with respect to the telecommunications industry is the development and operation of an efficient industry providing services at economic and equitable rates. The authors investigate the efficiency-equity issues underlying this policy from the standpoint of the regulated utility.

Comments

The analysis in this paper is very courageous in that where the unknown is encountered the authors use appropriate proxies to overcome such limitations. The results, however, are conclusions that should not be generalized. The conclusions are therefore only indicative of what the policy authorities should consider in regulating pricing of the natural monopoly subject to a rate of return constraint.

The result suggested by the study, namely that some price adjustments in residential local and toll services seem desirable, appears to be consistent with the crude pricing rule in microeconomic theory: that is, revenue maximization of a natural monopoly or an industry can be achieved by employing a pricing policy that results in the price elasticity of demand drifting towards unity.

WELFARE IMPLICATIONS

THE WELFARE EFFECTS OF A REGULATORY CONSTRAINT: A PRODUCTIVE EFFICIENCY APPROACH

W.E. DIEWERT

University of British Columbia

1. Introduction

The aim of this paper is to present a methodology for evaluating the welfare losses due to monopolistic behavior (or the behavior induced by regulatory constraints) in an industry such as telecommunications where there is a "natural" tendency towards monopoly.

It is generally suspected that regulated utilities are unwilling (or more realistically, are unable) to sell their products at prices equal to marginal costs. Thus the question arises: how can we calculate the loss of welfare which is induced by noncompetitive pricing behavior in a regulated industry?

A first approach to this problem is the following one. Assume that we know preferences and technology and that we have a socially desirable state in mind (a Pareto optimal state, where no consumer's welfare can be improved without reducing someone else's welfare), and compare the welfare of each consumer group in the socially desirable situation with the observed equilibrium welfares. The problem with this approach is that its informational requirements are very high.

A second approach would be to assume that we only have information on the technology of the regulated firm and the rest of the economy and that we are given a socially desirable set of producer prices. Now maximize the net value of output for the entire economy, including regulated and unregulated firms. Compare this "optimal" value of net output with the net output generated by the economy's observed aggregate net output vector, where these outputs and inputs are evaluated at the socially desirable prices.

The second approach is the approach taken in this paper. The main result of the present paper is the quadratic approximation to the measure of output loss defined above which is defined by formula (45) below. In section 6 below, we make some guesses at various elasticities and markups, and we evaluate the resulting output loss. Our estimates of the output loss range from 5 to 40 percent of GNP (see Table 2 below). These are gigantic loss estimates compared to the rather small estimates of deadweight loss that other researchers have obtained using Harberger's [1971] methodology in other contexts. Our large loss estimates are due to the fact that we have used the Averch-Johnson [1962] model of the regulated firm, so that there can be losses due to both monopolistic behavior on output markets as well as noncompetitive behavior on input markets due to the overcapitalization phenomenon.

We now present a more detailed summary of the paper.

In many monopolistic industries, governments impose a rate of return constraint, presumably in order to reduce the welfare losses due to monopolistic behavior. Averch and Johnson [1962] and others[1] established that the effect of the regulatory constraint is to encourage the regulated firm to use "too much" capital, but at the same time, monopolistic pricing behavior will still occur. In section 2 below, we review briefly the microeconomics of the Averch-Johnson model.

There have been surprisingly few studies of the welfare gains or losses due to the regulation of monopolistic behavior. Sheshinski [1971] studies the welfare effects of regulation in a simple general equilibrium model with one consumer and one monopolistic production sector (and no competitive sector). Bailey [1973; pp. 107-109] points out that the welfare loss in a general equilibrium model where the monopolist behaves according to the Averch-Johnson model consists of two components: (i) a loss of productive efficiency due to the over use of capital in the regulated sector and

(ii) a loss of exchange efficiency due to the fact that the regulated monopolist still sells his output at a price greater than marginal cost. Bailey also presents a nice geometric analysis illustrating the two components of welfare loss. However, she does not present a mathematical formula which could show how the welfare loss depends on the magnitude of the distortions on output and input markets and various elasticities of supply and demand. In sections 3 and 4 below, we present some simple formula which enable one to calculate the approximate welfare loss.

In section 3, we take a pure productive efficiency approach to the evaluation of the loss of output induced by regulating a monopoly. The methodology developed in this section should be useful in other models where producers face different prices for the same input or output; e.g., the welfare losses due to sectoral specific taxes or subsidies could be calculated using the material in this section.

The formula developed in section 3 enables us to calculate the loss of productive efficiency due to overcapitalization in the regulated sector, but it does not enable us to calculate the loss of exchange efficiency due to monopolistic pricing behavior. The formula developed in sector 4 enables us to calculate both components of welfare loss. However, the resulting loss measure depends on prices and thus is no longer a pure efficiency loss measure

In section 5, we study how our approximate welfare loss measure changes as we change various elasticities and markups.

In section 6, we specialize the formula developed in section 4 to the case of one monopolistic output and one capital good. Section 7 presents a geometric analysis of this case.

Section 8 concludes with a discussion of some of the weaknesses of our model.

2. The Averch Johnson Model

In this section, we outline a multiple output, multiple input version of the Averch-Johnson [1962] model of regulated monopolistic behavior.

It is assumed that the regulated monopolist attempts to maximize profits subject to three constraints: (i) the outputs and inputs chosen must be technologically feasible, (ii) the output vector supplied must be demanded, and (iii) the profits earned by the firm must satisfy a "fair rate of return" constraint.

Specifically, the regulated monopolist attempts to solve the following constrained profit maximization problem:

$$\max_{y,x,k,p}\Big\{ p \cdot y - w \cdot x - r \cdot k : \text{ (i) } (y,x,k) \in S , \text{ (ii) } p = g(y) , \\ \text{(iii) } p \cdot y - w \cdot x - r \cdot k \leq e \cdot k \Big\} \quad \ldots .(1)$$

where $y \geq 0_I^2$ denotes an I dimensional vector of outputs produced by the regulated monopolist, $p >> 0_I$ is a positive vector of output prices that the monopolist faces, $x \geq 0_J$ is a J dimensional vector of variable inputs (labor and intermediate goods) used by the firm, $w >> 0_J$ is a positive vector of variable input prices that the firm faces, $k \geq 0_N$ is an N dimensional vector of variable capital goods used by the firm and $r >> 0_N$ is the corresponding vector of competitive rental prices that the firm faces, S is the technologically feasible set of outputs and inputs that can be produced and utilized by the firm, $p = g(y)$ means $p_1 = g_1(y)$, $p_2 = g_2(y)$, $\cdots, p_I = g_I(y)$ where g_i is the ith inverse demand function that the firm faces, and $e \equiv (e_1, e_2, \cdots, e_N)^T$ where $e_n \geq 0$ is the excess profit rate that the regulated monopolist is allowed to earn on the nth type of capital for $n = 1,2,\cdots,N$.[3]

Define the firm's <u>variable cost function</u> C as

$$C(w,y,k) \equiv \min_x\{ w\cdot x : (y,x,k) \in S \} \quad . \qquad \ldots .(2)$$

It is easy to see that the second set of constraints in (1) can be substituted into the objective function of (1) and the vector of variables p can be eliminated from the resulting maximization problem. It is also possible to show[4] that the first set of constraints and the vector of variables x can be eliminated from (1) if the term $w\cdot x$ in the third constraint and the objective function of (1) is replaced by the variable cost function $C(w,y,k)$. Thus (1) is equivalent to the following constrained maximization problem:

$$\max_{y,k} \left\{ g(y)\cdot y - C(w,y,k) - r\cdot k : g(y)\cdot y - C(w,y,k) - r\cdot k \leq e\cdot k \right\} . \ . \ .(3)$$

Denote a solution to (3) by (y^*,k^*).[5] Assume $y^* \gg 0_I$, $k^* \gg 0_N$ and $C(w,y^*,k^*)$ is differentiable with respect to the components of y and k. If the regulatory constraint in (3) is not binding, then the first order necessary conditions for (3) are:

$$g(y) + \nabla g(y^*)y^* - \nabla_y C(w,y^*,k^*) = 0_I \qquad \ldots .(4)$$

$$- \nabla_k C(w,y^*,k^*) - r = 0_N \ , \qquad \ldots .(5)$$

where

$$\nabla g(y^*) \equiv [\nabla g_1(y^*),\nabla g_2(y^*),\cdots,\nabla g_I(y^*)]^T,$$

$$\nabla g_i(y^*) \equiv [\partial g_i(y^*)/\partial y_1,\cdots,\partial g_i(y^*)/\partial y_I]^T \quad \text{for} \quad i = 1,2,\cdots,I,$$

$$\nabla_y C(w,y^*,k^*) \equiv [\partial C(w,y^*,k^*)/\partial y_1,\cdots,\partial C(w,y^*,k^*)/\partial y_I]^T,$$

and

$$\nabla_k C(w,y^*,k^*) \equiv [\partial C(w,y^*,k^*)/\partial k_1,\cdots,\partial C(w,y^*,k^*)/\partial k_N]^T..$$

we note that the variable cost function $C(w,y,k)$ will be increasing in the components of y and decreasing in the components of k, so that $\nabla_y C(w,y^*,k^*) \gg 0_I$ is a positive vector of marginal costs and $\nabla_k C(w,y^*,k^*) \ll 0_N$ is a negative vector of decreases in variable cost due to marginal increases in the components of capital. Define the optimal producer's selling price for outputs vector as $p^* \equiv g(y^*)$ and the optimal markup vector as

$$m^* \equiv -\nabla g(y^*)\, y^* \quad . \qquad \ldots .(6)$$

Usually, $m^* \geq 0_I$ and if the firm behaves competitively in output markets, $m^* \equiv 0_I$. Now rewrite the first order conditions (4) and (5) as

$$p^* - m^* = \nabla_y C(w,y^*,k^*) \qquad \ldots .(7)$$

$$r^* = -\nabla_k C(w,y^*,k^*) \quad . \qquad \ldots .(8)$$

If the firm behaves competitively, (7) and (8) would hold with $m^* = 0_I$.

Conditions (4)-(8) were contingent on the nonbindingness of the regulatory constraint in (3). If the components of the e vector are large enough, this will be the case. Now assume that the regulatory constraint is binding. Upon introducing a Lagrange multiplier λ for the constraint in (3), we find that the first order necessary conditions for the resulting constrained maximization problem can be simplified to the following conditions.

$$p^* - m^* = \nabla_y C(w,y^*,k^*) \qquad \ldots .(9)$$

$$r^* - \frac{\lambda^* e}{1-\lambda^*} = -\nabla_k C(w,y^*,k^*) \qquad \ldots . (10)$$

$$e \cdot k^* = p^* \cdot y^* - C(w,y^*,k^*) - r \cdot k^* \qquad \ldots . (11)$$

where $y^* \gg 0_I$ and $k^* \gg 0_N$ are a solution to (3), $p^* \equiv g(y^*)$, $m^* \equiv - \nabla g(y^*)y^*$ is the optimal markup vector and $\lambda^* \geq 0$ is the optimal Lagrange multiplier.

By adapting a technique due to Bailey [1973; pp. 72-75], one can show[6] that $\lambda^* \leq \min_n\{r_n/(r_n + e_n) : n = 1,2,\cdots,N\} < 1$, if any $e_n > 0$.

Equations (9) and (10) show that the behavior of a regulated monopolist is identical to the behavior of a competitive producer who faces output prices $p^* - m^*$ (instead of p^*) and capital rental prices $r - \lambda^*(1-\lambda^*)^{-1}e$ (instead of r). If $p^* - m^* \ll p^*$ and $r - \lambda^*(1-\lambda^*)^{-1}e \ll r$, then there is a presumption that the regulated monopolist will produce "too little" output and utilize "too much" capital.

In the following sections, we will try to make precise what we mean by the last assertion.

3. An Efficiency Approach to the Evaluation of Output Loss

An important implication of the Averch-Johnson model of firm behavior developed in the previous section is that producers in the regulated sector will face a lower effective capital services rental price vector than producers in the competitive sector (assuming that the regulatory constraint is effective). In this section, we develop an approximate measure of the loss of output which an economy will suffer when it allows two groups of producers to behave as if they are facing different prices for the same good.

We introduce a more convenient notation at this point. Let us suppose that there are $M+1$ goods in the economy that two groups of producers either both produce or use as inputs. We index outputs with a positive sign and inputs with a negative sign. Assume that both producers face the same price for the first good but they may face different prices for the other goods. Let f^1 and f^2 denote the production functions for the two sectors; i.e., we have $z_0^i = f^i(z^i)$, $i = 1,2$ where z_0^i is the maximum amount of the first good that can be produced[7] given that sector i produces the net output vector $z^i \equiv (z_1^i, z_2^i, \cdots, z_M^i)$ of other goods.[8]

Consider the following constrained maximization problem:

$$\max_{z^1, z^2} \left\{ f^1(z^1) + f^2(z^2) \; : \; z^1 + z^2 = \bar{z} \right\} , \qquad \ldots .(12)$$

where $\bar{z}$ is a fixed vector of net outputs. Thus, what we are trying to do in (12) is to maximize the net production of good 0[9] across the two sectors given that the two sectors have to produce the net output vector $\bar{z} \equiv (\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_M)^T$ for the other M goods. Solving (12) will ensure that the resulting $M+1$ dimensional net output vector for the economy is efficient.

Using the constraints in (12) to eliminate z^2, we find that the constrained maximization problem (12) is equivalent to the following unconstrained maximization problem:

$$\max_{z^1} \{ f^1(z^1) + f^2(\bar{z} - z^1) \} . \quad \ldots .(13)$$

Let z^{1*} denote a solution to (13), so that z^{1*} and $z^{2*} \equiv \bar{z} - z^{1*}$ are solutions to (12). Assuming that f^i is differentiable in a neighborhood around z^{i*}, the first order necessary conditions for (13) are:

$$\nabla_{z^1} f^1(z^{1*}) - \nabla_{z^2} f^2(\bar{z} - z^{1*}) = 0_M \quad \ldots .(14)$$

where

$$\nabla_{z^1} f^1(z^{1*}) \equiv [\partial f^1(z^{1*})/\partial z_1, \cdots, \partial f^1(z^{1*})/\partial z_M]^T$$

and

$$\nabla_{z^2} f^2(\bar{z} - z^{1*}) \equiv [\partial f^2(z^{2*})/\partial z_1, \cdots, \partial f^2(z^{2*})/\partial z_M]^T .$$

We shall also assume that the second order sufficient conditions for (13) hold at z^{1*}; i.e., we assume that

$$[\nabla^2_{z^1 z^1} f^1(z^{1*}) + \nabla^2_{z^2 z^2} f^2(\bar{z} - z^{1*})] \text{ is negative definite}, \quad \ldots .(15)$$

where

$$\nabla^2_{z^1 z^1} f^1(z^{1*}) \equiv [\partial^2 f^1(z^{1*})/\partial z_i \partial z_j]$$

and

$$\nabla^2_{z^2 z^2} f^2(\bar{z} - z^{1*}) = [\partial^2 f^2(z^{2*})/\partial z_i \partial z_j]$$

are matrices of second order partial derivatives of f^1 and f^2 evaluated at z^{1*} and $z^{2*} \equiv \bar{z} - z^{1*}$ respectively.

We also define the vector of competitive producer prices p^* as

$$p^* \equiv -\nabla_{z^2} f^2(z^{2*})$$

$$= -\nabla_{z^1} f^1(z^{1*}) \quad \text{using (14) and } z^{2*} = \bar{z} - z^{1*}. \qquad \ldots .(16)$$

Now we are ready to introduce a distortions vector ξ. We assume that producers in sector 2 face prices $(1, p_1, p_2, \cdots, p_M) \equiv (1, p^T)$, (with good 0 as the numeraire good which has price 1) while producers in sector 1 face prices $(1, p_1 + \xi_1 t, p_2 + \xi_2 t, \cdots, p_M + \xi_M t) \equiv (1, p^T + \xi^T t)$, where the scalar variable t equals 1. Our complete model of producer behavior can be represented by the following system of equations:

$$p = -\nabla_{z^2} f^2(z^2) \qquad \ldots .(17)$$

$$p + \xi t = -\nabla_{z^1} f^1(z^1) \qquad \ldots .(18)$$

$$\bar{z} = z^1 + z^2 . \qquad \ldots .(19)$$

Note that the first order partial derivatives of f^1 and f^2 are negative since z^1 and z^2 are net output vectors. (17)-(19) are to be regarded as 3M simultaneous equations in the 3M+1 variables p, z^1, z^2 and t. In fact, we can substitute (17) and (19) into (18), eliminating p and z^2 and we end up with the following system in z^1 and t:

$$-\nabla_{z^2} f^2(\bar{z} - z^1) + \xi t = -\nabla_{z^1} f^1(z^1) . \qquad \ldots .(20)$$

We apply the implicit function theorem to solve for z^1 as a function of t around $t = 0$.[10] We find that the vector of first order derivatives of $z^1(t)$ with respect to t evaluated at $t = 0$ is:

$$\nabla_t z^1(0) = -[\nabla^2_{z^1 z^1} f^1(z^{1*}) + \nabla^2_{z^2 z^2} f^2(z^{2*})]^{-1} \xi \quad \ldots \quad ....(21)$$

Now we are ready to calculate the welfare loss due to the distortions within the production sector. Since we are holding fixed the net output of all goods except good 0, it is reasonable to take as our measure of welfare the net output of good 0; i.e., define the welfare function

$$W(t) \equiv f^1(z^1(t)) + f^2(\bar{z} - z^1(t)) \quad . \quad ....(22)$$

We can obtain an approximate measure of welfare change by taking a second order Taylor Series expansion of $W(t)$ around $t = 0$. Thus we calculate $W'(0)$ and $W''(0)$ below. Differentiate (22) with respect to t and we obtain:

$$W'(t) = [\nabla_{z^1} f^1(z^1(t)) - \nabla_{z^2} f^2(\bar{z} - z^1(t))] \cdot \nabla_t z^1(t)$$

$$= t\xi \cdot \nabla_t z^1(t) \qquad \text{using (20)}$$

$$\therefore \; W'(0) = 0 \quad \text{and} \quad ....(23)$$

$$W''(0) = -\xi \cdot \nabla_t z^1(0) \quad ....(24)$$

$$= \xi^T[\nabla^2 f^1(z^{1*}) + \nabla^2 f^2(z^{2*})]^{-1} \xi \qquad \text{using (21)}$$

$$< 0 \quad \text{if} \quad \xi \neq 0_M \qquad \text{using (15).}$$

Thus an approximate measure of the loss of output (of good 0) due to producers facing prices $p + \xi$ in sector 1 and prices $p \equiv -\nabla_{z^2} f^2(\bar{z} - z^1(1))$ in sector 2 is

$$\Delta W \equiv (1/2)\, W''(0) = \xi^T[\nabla^2 f^1(z^{1*}) + \nabla^2 f^2(z^{2*})]^{-1}\, \xi / 2 \quad \ldots\ldots (26)$$

Thus the approximate loss depends on ξ, a vector of differences in prices that the two sectors face and the Hessian matrices of second order partial derivatives of the two production functions f^1 and f^2 evaluated at the points z^{1*} and z^{2*} respectively where z^{1*} and z^{2*} are a solution to the constrained maximization problem (12).[11]

What are the informational requirements for evaluating (26) for an actual economy with distortions within the production sector? First, we require that there be at least one good where both sets of producers face the same price (good 0). This will be the good by which we measure the loss of output due to the distortions. Next, we require that we can measure the difference in prices that the two sets of producers face. This is the ξ vector. Next, we observe the actual net supply vectors for the two sectors for goods 1 to M, $\tilde{z}^1$ and $\tilde{z}^2$, and calculate the total net supply vector $\bar{z} \equiv \tilde{z}^1 + \tilde{z}^2$. Recall that goods $0,1,\cdots,M$ are the goods that <u>both</u> sectors produce (or demand); goods produced or demanded by sector 1 that are not produced or utilized by sector 2 are held constant. If we actually know the production functions f^1 and f^2, then we can simply solve (12), compute $f^1(z^{1*}) + f^2(z^{2*})$ and compare this last number with the actual (distorted equilibrium) net output of good 0, $f^1(\tilde{z}^1) + f^2(\tilde{z}^2)$; i.e., there is no need to resort to the approximation (26) at all. However, if we have only local information about f^1 and f^2 (the usual case), then we approximate $\nabla^2 f^1(z^{1*})$ by $\nabla^2 f^1(\tilde{z}^1)$ and $\nabla^2 f^2(z^{2*})$ by $\nabla^2 f^2(\tilde{z}^2)$, and the approximate loss measure (26) becomes

$$\Delta\tilde{W} \equiv \xi^T [\nabla^2 f^1(\tilde{z}^1) + \nabla^2 f^2(\tilde{z}^2)]^{-1} \xi / 2 \quad . \qquad ....(27)$$

Thus the approximate loss measure (27) can readily be evaluated using only local information on the technology and information on the size of the price distortions between the two sectors, ξ.

The idea of measuring the economic loss due to price distortions as the physical amount of one commodity which can be thrown away compared to an optimal situation while holding constant the net production of other commodities can be traced back to Allais [1943],]1973], who in turn attributes it to Pareto. Debreu [1951], [1954] has also popularized a variant of this idea, the coefficient of resource utilization.

Using (24), the loss measure (26) becomes

$$\Delta W = -\xi \cdot \nabla_t z^1(0)$$

$$= -\Sigma_m \Delta p_m \Delta z^1_m \qquad ....(28)$$

where $\xi_m \equiv \Delta p_m$ is the difference in the mth price that the two sectors face and $\Delta z^1_m \equiv \partial z^1_m(0)/\partial t$ is (approximately) the change in the net supply of good m by sector 1 induced by the distortion; i.e., Δz^1_m is approximately equal to $z^1_m(1) - z^1_m(0) = \tilde{z}^1_m - z^{1*}_m$. The approximate loss formula (28) bears a family resemblance to the "triangle formula" developed by Allais [1943; p. 616], Hotelling [1938; p. 254], Boiteux [1951; p. 130], Debreu [1954; p. 20] and Harberger [1964; p. 70], [1971; p. 790].[12]

Let us review what assumptions are required in order to derive the approximate loss formula (26). First, we require that z^{1*} and z^{2*} be a solution to (12). This is not a restrictive assumption. Secondly, we require that f^1 and f^2 be twice continuously differentiable in a neighborhood of z^{1*} and z^{2*} respectively, so that z^{1*} is an interior solution to (13). Finally, we require that the matrix in (15) be negative definite (instead of being merely negative semidefinite). In particular, note that no global assumptions about the curvature properties of the production functions f^1 and f^2 are required. Thus, the production functions need not be concave, and they can exhibit increasing returns to scale.[13]

We could use (26) or (27) in order to calculate the loss of real output due to the effective regulation of a monopolist. We could add a competitive sector to a regulated monopolist sector, whose behavior is described by equations (1) and (10) of the previous section. We would take 1 labor good as good 0. The components of the ξ vector would be zero if the good were a type of labor or a competitive output produced by the competitive sector, $-m_i^1$ if the good were the ith monopoly output good, and $-\lambda^*(1-\lambda^*)^{-1}e_n$ if the good were the nth type of capital. However, the above substitutions would not lead to a correct use of the model developed on this section. The problem is that the outputs produced by the regulated monopolist and the outputs produced by the competitive sector are completely separate, and thus they would have to be held fixed in the maximization problem (12). Thus, only the capital distortions would enter the loss formula (26). In particular, if the regulatory constraint were not binding, then the loss formula (26) correctly applied to the present situation would yield a zero loss (since in this case $\lambda^* = 0$ and hence $\xi = 0$). As we noted in the introduction, the pure efficiency loss measure developed in this section does not enable us to calculate the loss of exchange efficiency due to monopolistic pricing behavior. Thus, in the next section, we develop a formula that will enable us to calculate both components of welfare loss.

4. A Producer Price Approach to the Evaluation of Output Loss

We revert to the notation used in section 2 with one exception: the vector w, which stood for a vector of wage rates and intermediate input prices that the regulated monopolist faces, now includes in addition, the prices of outputs produced by the competitive sector.

We assume that $S^c = \{(x,k)\}$ is the set of technologically feasible outputs and inputs that is producible by the competitive sector: x is a J dimensional vector of net outputs (outputs are indexed with a positive sign, inputs with a negative sign) while $k \geq 0_N$ is a vector of capital inputs. Let $w \gg 0_J$ be a vector of output and input prices that the competitive sector faces (omitting capital input prices $r \gg 0_N$) and define the competitive sector's variable profit function π as[14]

$$\pi(w,k) \equiv \max_x \{w \cdot x : (x,k) \in S^c\} \quad . \qquad \ldots .(29)$$

Let $\bar{k} \gg 0_N$ denote a vector of the amounts of capital available to be distributed between the two sectors, $k^1 \geq 0_N$ denotes the vector of capital services used by the regulated monopolist sector, and $k^2 \geq 0_N$ denotes the vector of capital services used by the competitive sector. We assume that a planner wishes to produce outputs and allocate inputs between the two sectors so as to maximize net value added subject to aggregate capital constraints, where sector 1's outputs are priced at the producer price vector $p \gg 0_I$ and other outputs and noncapital inputs are priced at $w \gg 0_J$; i.e., the planner wishes to solve the following constrained maximization problem:[15]

$$\max_{y,k^1,k^2} \{p \cdot y - C(w,y,k^1) + \pi(w,k^2) : k^1 + k^2 = \bar{k}\} . \qquad \ldots .(30)$$

Using the constraint in (30) in order to eliminate k^2, we find that (30) is equivalent to the following unconstrained maximization problem in y and k^1:

$$\max_{y,k^1} \{p \cdot y - C(w,y,k^1) + \pi(w,\bar{k} - k^1)\} . \qquad \ldots .(31)$$

We suppose that a solution $y^* \gg 0_I$, $k^{1*} \gg 0_N$ to (31) exists, with $k^{2*} \equiv \bar{k} - k^{1*} \gg 0_N$ and that C and π are twice continuously differentiable with respect to y, k^1, k^2 in a neighbourhood of (w,y^*,k^{1*}) and (w,k^{2*}) respectively. The first order necessary conditions for (31) are:

$$p - \nabla_y C(w,y^*,k^{1*}) = 0_I \qquad \ldots .(32)$$

$$- \nabla_{k^1} C(w,y^*,k^{1*}) - \nabla_{k^2} \pi(w,k^{2*}) = 0_N . \qquad \ldots .(33)$$

We shall also assume that the following second order sufficient conditions for y^*, k^{1*} to solve (31) are satisfied:

$$\begin{bmatrix} -\nabla^2_{yy} C(w,y^*,k^{1*}), & -\nabla^2_{yk} C(w,y^*,k^{1*}) \\ -\nabla^2_{ky} C(w,y^*,k^{1*}), & -\nabla^2_{kk} C(w,y^*,k^{1*}) + \nabla^2_{kk}\pi(w,k^{2*}) \end{bmatrix} \text{ is a negative definite matrix.} \qquad \ldots .(34)$$

We now introduce distortions into the above model. Recall equations (9) and (10). Define $\tau \equiv \lambda^*(1-\lambda^*)^{-1}e$. Our model of producer behavior can be summarized by the following system of equations:

$$p - mt = \nabla_y C(w,y,k^1) \quad , \qquad \text{....(35)}$$

$$r - \tau t = -\nabla_k C(w,y,k^1) \quad , \qquad \text{....(36)}$$

$$r = \nabla_k \pi(w,k^2) \quad , \qquad \text{....(37)}$$

$$k^1 + k^2 = \bar{k} \qquad \text{....(38)}$$

where p, w and $\bar{k}$ are fixed, and y, k^1, k^2 and r are to be regarded as functions of the scalar t. When $t=1$, (35) and (36) correspond to (9) and (10) (without the asterisks). Let us substitute (37) and (38) into (35) and (36), eliminating k^2 and r. We obtain the following equivalent system of equations in y, k^1 and t:

$$p - mt = \nabla_y C(w,y,k^1) \qquad \text{....(39)}$$

$$\nabla_k \pi(w,\bar{k}-k^1) - \tau t = -\nabla_k C(w,y,k^1) \quad . \qquad \text{....(40)}$$

Assumption (34) allows us to solve (39)-(40) implicitly for $y=y(t)$, $k^1=k^1(t)$ in a neighbourhood around $t=0$. The vectors of derivatives of these implicit functions evaluated at $t=0$ are defined by:

$$\begin{bmatrix} \nabla_t y(0) \\ \nabla_t k^1(0) \end{bmatrix} = A^{-1} \begin{bmatrix} m \\ -\tau \end{bmatrix} \qquad \text{....(41)}$$

where the negative definite matrix A is defined by (34).

Now we are ready to evaluate the losses due to noncompetitive pricing behavior within the production sector. We take our welfare function to be the net value of outputs produced minus (noncapital) inputs used by the production sector, evaluating outputs and inputs at the social prices p and w. Thus welfare regarded as a function of t is

$$W(t) \equiv p \cdot y(t) - C\big(w, y(t), k^1(t)\big) + \pi\big(w, \bar{k} - k^1(t)\big) . \qquad \ldots .(42)$$

Differentiate (42) with respect to t and we obtain:

$$\begin{aligned} W'(t) &= p \cdot \nabla_t y(t) - \nabla_y C\big(w, y(t), k^1(t)\big) \cdot \nabla_t y(t) \\ &\quad - \nabla_{k^1} C\big(w, y(t), k^1(t)\big) \cdot \nabla_t k^1(t) - \nabla_{k^2} \pi\big(w, k^2(t)\big) \cdot \nabla_t k^1(t) \\ &= t[m \cdot \nabla_t y(t) - \tau \cdot \nabla_t k^1(t)] \quad \text{using (39)-(40)} \end{aligned}$$

$$\therefore \; W'(0) = 0 \qquad \ldots .(43)$$

and

$$\begin{aligned} W''(0) &= m \cdot \nabla_t y(0) - \tau \cdot \nabla_t k^1(0) \\ &= [m^T, -\tau^T] \, A^{-1} \begin{pmatrix} m \\ -\tau \end{pmatrix} \quad \text{using (41)} \qquad \ldots .(44) \\ &< 0 \quad \text{if} \quad m \neq 0_I \;\text{ or }\; \tau \neq 0_N \;\text{ using (34).} \end{aligned}$$

Thus an approximate measure of the loss of output due to the regulated monopolist taking prices $p - m$ for outputs and $r - \tau$ for capital inputs instead of p and $r \equiv \nabla_{k^2} \pi\big(w, \bar{k} - k^1(1)\big)$ respectively is

$$\Delta W \equiv [W(0) + W'(0) \cdot 1 + (\tfrac{1}{2} W''(0)1^2] - W(0)$$

$$= (1/2)[m^T, -\tau^T]\, A^{-1} \begin{pmatrix} m \\ -\tau \end{pmatrix} \text{ using (43) and (44)} \qquad \ldots .(45)$$

where the negative definite matrix A is defined by (34). Thus if $m \neq 0_I$ or $\tau \neq 0_N$ (or both), then under our assumptions, the approximate loss will be negative.

Let $y(1)$, $k^1(1)$, and $r(1)$ denote the solution to (35)-(38) when $t = 1$; i.e., this corresponds to the distorted (observable) equilibrium for the economy. If we have a global knowledge of the monopolist's cost function C and the competitive sector's variable profit function π, then we can simply solve (30) with $\bar{k} \equiv k^1(1) + k^2(1)$. The solution for (30) will yield $W(0)$. $W(1)$ can be evaluated using definition (42) and the observed (distorted) equilibrium values for the economy. Thus we can calculate $W(1) - W(0)$ directly, and there is no need to resort to the approximation (45). If we have only a local knowledge of C and π, then approximate the matrix A by

$$\tilde{A} \equiv \begin{bmatrix} -\nabla^2_{yy}C(w,y(1),k^1(1)), & -\nabla^2_{yk}C(w,y(1),k^1(1)) \\ -\nabla^2_{ky}C(w,y(1),k^1(1)), & -\nabla^2_{kk}C(w,y(1),k^1(1)) + \nabla^2_{kk}\pi(w,k^2(1)) \end{bmatrix} \quad . .(46)$$

and replace A in (45) by $\tilde{A}$ defined by (46). Thus the approximate loss can be calculated using only local information on the technology and information on the monopolistic markup vector m and the implicit capital subsidy vector $\tau \equiv \lambda^*(1-\lambda^*)^{-1}e$.

5. The Determinants of the Output Loss Due to Distortions

Since the change in welfare due to the price distortions in the production sector is always negative, let us define the nonnegative loss L as $-\Delta W$, where ΔW is defined by (45); i.e., define

$$L \equiv -\Delta W = -\tfrac{1}{2}[m^T, -\tau^T] \begin{pmatrix} -\nabla^2_{yy}C, & -\nabla^2_{yk}C \\ -\nabla^2_{ky}C, & -\nabla^2_{kk}C + \nabla^2_{kk}\pi \end{pmatrix}^{-1} \begin{pmatrix} m \\ -\tau \end{pmatrix} \geq 0. \quad \ldots (47)$$

Denote the ij^{th} element of the matrix of second order derivatives $\nabla^2_{yy}C(w,y^*,k^{1*})$ as α_{ij}, the in^{th} element of $\nabla^2_{yk}C(w,y^*,k^{1*})$ as β_{in}, the mn^{th} element of $\nabla^2_{kk}C(w,y^*,k^{1*})$ as γ_{mn} and the mnth element of $\nabla^2_{kk}\pi(w,k^{2*})$ as δ_{mn}. From (47), it is evident that the loss of output L is a function of the distortion vectors m and τ and the α_{ij}, β_{in}, γ_{mn}, and δ_{mn}. We now attempt to determine how L changes as these parameters change.

First, it is evident if all of the distortions m and τ increase by a common proportionality factor λ, then the loss will increase by the square of the proportionality factor; i.e.,

$$L(\lambda m, \lambda\tau, \alpha_{ij}, \beta_{in}, \gamma_{mn}, \delta_{mn}) = \lambda^2 L(m, \tau, \alpha_{ij}, \beta_{in}, \gamma_{mn}, \delta_{mn}). \quad \ldots (48)$$

If only one component of m or τ increases, the situation is not so clear cut. Differentiating (47) with respect to m_i, the i^{th} component of the markup vector m, yields[16] (where e_i is a unit vector with a one in component i):

$$\partial L/\partial m_i = -[e_i^T, 0_N^T]\, A^{-1} \begin{pmatrix} m \\ -\tau \end{pmatrix} \quad \ldots (49)$$

$$= -dy_i(0)/dt \qquad \text{using (41)}. \quad \ldots (50)$$

"Usually," $dy_i(0)/dt$ will be negative, so "usually" the loss will increase as the i^{th} monopolistic markup m_i increases. If $\tau = 0_N$, and there is only one nonzero monopolistic markup (the i^{th} so that $m_i > 0$ while $m_j = 0$ for $j \neq i$), then

$$dy_i(0)/dt \;=\; [e_i^T,\, 0_N^T]\, A^{-1}\begin{pmatrix} e_i \\ 0_N \end{pmatrix} m_i \;<\; 0 \qquad \ldots .(51)$$

since A is a negative definite matrix by (34) and hence the i^{th} diagonal element of A^{-1} is also negative. Thus, in the case where there is only one distortion ($m_i > 0$), we can rigorously prove that $\partial L/\partial m_i > 0$.

Differentiating (47) with respect to τ_n, the n^{th} component of the implicit capital services subsidy vector, yields (where e_n is a unit vector with a one in component n):

$$\partial L/\partial \tau_n \;=\; [0_I^T,\, e_n^T]\, A^{-1}\begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$= \; dk_n^1(0)/dt \qquad \text{using (41).} \qquad \ldots .(52)$$

"Usually," $dk_n^1(0)/dt$ will be positive, so "usually" the loss will increase as the n^{th} capital subsidy τ_n increases. However, we can only rigorously prove the following proposition: if $m = 0_I$ and $\tau = e_n\tau_n$ where $\tau_n > 0$ so that there is only one nonzero subsidy, then

$$dk_n^1(0)/dt \;=\; [0_I^T,\, e_n^T]\, A^{-1}\begin{pmatrix} 0_I \\ e_n \end{pmatrix} (-\tau_n) \;>\; 0 \qquad \ldots .(53)$$

so that $\partial L/\partial \tau_n > 0$ under these conditions also.

We now attempt to evaluate the change in L due to a change in α_{ij}. First note that $\alpha_{ij} = \alpha_{ji}$ and

$$\alpha_{ij} \equiv \partial^2 C(w,y^*,k^{1*})/\partial y_i \partial y_j$$

$$= \partial p_i(w,y^*,k^{1*})/\partial y_j \qquad \ldots .(54)$$

where $p_i(w,y^*,k^{1*}) \equiv \partial C(w,y^*,k^{1*})/\partial y_i$ is the marginal cost of producing an additional unit of the i^{th} output. Thus α_{ij} is the change in this marginal cost due to a marginal increase in output j.[17] Differentiate L with respect to α_{ij} and we find[18]

$$\partial L/\partial \alpha_{ij} = -\frac{1}{2}[m^T,-\tau^T]\, A^{-1}\begin{pmatrix} e_i e_j^T + e_j e_i^T, & 0_{I\times N} \\ 0_{N\times T}, & 0_{N\times N}\end{pmatrix} A^{-1}\begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$= -\frac{dy_i(0)}{dt}\frac{dy_j(0)}{dt} \qquad \text{using (41).} \qquad \ldots .(55)$$

If $i=j$, (55) implies that the loss will decrease as α_{ii} increases; i.e.,

$$\partial L/\partial \alpha_{ii} = -[dy_i(0)/dt]^2 \leq 0 \quad \text{for} \quad i = 1,2,\cdots,I. \qquad \ldots .(56)$$

Since C is assumed to be twice continuously differentiable at (w,y^*,k^{1*}), $\gamma_{mn} = \gamma_{nm}$ where

$$\gamma_{mn} \equiv \partial^2 C(w,y^*,k^{1*})/\partial k_m \partial k_n$$

$$= -\partial r_m^1(w,y^*,k^{1*})/\partial k_n \qquad \ldots .(57)$$

where

$$r_m^1(w,y^*,k^{1*}) \equiv -\partial C(w,y^*,k^{1*})/\partial k_m$$

is the competitive rental price for the m^{th} type of capital service used in sector 1. Differentiation of L with respect to γ_{mn} yields

$$\partial L/\partial\gamma_{mn} = -\frac{1}{2}[m^T, -\tau^T]\, A^{-1}\begin{pmatrix} 0_{I\times I}, & 0_{I\times N} \\ 0_{N\times I}, & e_m e_n^T + e_n e_m^T \end{pmatrix} A^{-1}\begin{bmatrix} m \\ -\tau \end{bmatrix}$$

$$= -\frac{dk_m(0)}{dt}\;\frac{dk_n(0)}{dt} \qquad \text{using (41)} . \qquad \ldots .(58)$$

If $m = n$, (58) implies that the loss will decrease as γ_{nn} increases; i.e.,

$$\partial L/\partial\gamma_{nn} = -\,[dk_n(0)/dt]^2 \leq 0 \quad \text{for} \quad n = 1,2,\cdots,N. \qquad \ldots .(59)$$

Note that

$$\beta_{in} \equiv \partial^2 C(w,y^*,k^{1*})/\partial y_i \partial k_n$$

$$= \partial p_i(w,y^*,k^{1*})/\partial k_n$$

$$= -\partial r_n^1(w,y^*,k^{1*})/\partial y_i$$

since $p_i(w,y^*,k^{1*}) \equiv \partial C(w,y^*,k^{1*})/\partial y_i$, $r_n^1(w,y^*,k^{1*}) \equiv -\partial C(w,y^*,k^{1*})/\partial k_n$ and $\partial^2 C/\partial y_i \partial k_n = \partial^2 C/\partial k_n \partial y_i$. Differentiation of L with respect to β_{in} yields

$$\partial L/\partial\beta_{in} = -\frac{1}{2}[m^T, -\tau^T]\, A^{-1}\begin{pmatrix} 0_{I\times I}, & e_i e_n^T \\ e_n e_i^T, & 0_{N\times N} \end{pmatrix} A^{-1}\begin{bmatrix} m \\ -\tau \end{bmatrix}$$

$$= -\frac{dy_i(0)}{dt}\;\frac{dk_n^1(0)}{dt} \qquad \text{using (41).} \qquad \ldots .(61)$$

Finally, note that $\delta_{mn} = \delta_{nm}$ where

$$\delta_{mn} \equiv \partial^2\pi(w,k^{2*})/\partial k_m \partial k_n$$

$$= \partial r_m^2(w,k^{2*})/\partial k_n$$

where $r_m^2(w,k^{2*}) \equiv \partial\pi(w,k^{2*})/\partial k_m$ is the competitive rental price for the m^{th} type of capital service used in sector 2. Differentiation of L with respect to δ_{mn} yields

$$\partial L/\partial\delta_{mn} = \frac{1}{2}[m^T, -\tau^T] A^{-1} \begin{pmatrix} 0_{I\times I}, & 0_{I\times N} \\ 0_{N\times I}, & e_m e_n^T + e_n e_m^T \end{pmatrix} A^{-1} \begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$= \frac{dk_m(0)}{dt} \frac{dk_n(0)}{dt} . \quad \quad ....(62)$$

If $m = n$, (62) implies that the loss will increase as δ_{nn} increases;[19] i.e.,

$$\partial L/\partial\delta_{nn} = [\delta k_m(0)/dt]^2 > 0 \quad \text{for} \quad n = 1,2,\cdots,N. \quad ....(63)$$

We have established without ambiguity how L changes as α_{ii}, γ_{nn} and δ_{nn} increase; however, our results are ambiguous with respect to increases in α_{ij} for $i \neq j$, β_{in}, γ_{mn}, and δ_{mn} for $m \neq n$. It is possible to derive unambiguous results for the latter parameters if we introduce the concept of a compensated change. For example, let us consider the case where we increase α_{ij} for $i \neq j$. If the technology set in sector 1, S, exhibited constant returns to scale, we could not increase α_{ij} by itself, since the matrix of α's, α say, and the matrix of β's, β say, must satisfy the restrictions[20]

$$\alpha y^* + \beta k^{1*} = 0_I . \quad \quad ....(64)$$

Thus for a <u>compensated</u> increase in α_{ij}, we simultaneously decrease α_{ii} and α_{jj} so that (64) would continue to hold if it happened to be true initially. Thus, if we let α_{ii}, α_{jj}, and α_{ij} be functions of a parameter u, we have:[21]

$$\alpha_{ij}(u) \equiv \text{initial value} + u$$

$$\alpha_{ii}(u) \equiv \text{initial value} - (y_j^*/y_i^*)u \quad \text{and}$$

$$\alpha_{jj}(u) \equiv \text{initial value} - (y_i^*/y_j^*)u \quad . \qquad \ldots .(65)$$

Now use equations (65) and define the derivative of L with respect to a compensated increase in α_{ij} for $i \neq j$ as $\partial L/\partial\alpha_{ij}\big|_c \equiv \partial L/\partial u$. Substituting (65) into (47) and differentiating yields

$$\partial L/\partial\alpha_{ij}\big|_c \equiv -\frac{1}{2}\,[m^T, -\tau^T]\; A^{-1}\begin{pmatrix} M_{ij}, & 0_{I\times N} \\ 0_{N\times I}, & 0_{N\times N}\end{pmatrix} A^{-1}\begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$\geq 0 \qquad \text{for} \qquad 1 \leq i < j \leq I \qquad \ldots .(66)$$

since the matrix $M_{ij} \equiv [(-y_j^*/y_i^*)e_ie_i^T - (y_i^*/y_j^*)e_je_j^T + e_ie_j^T + e_je_i^T]$ is negative semidefinite.

We can define analogous compensated increases in β_{in} and $\gamma_{mn}(m \neq n)$. We find that

$$\partial L/\partial\beta_{in}\big|_c \equiv -\frac{1}{2}\,[m^T, -\tau^T]\; A^{-1}\begin{pmatrix} -(k_n^{1*}/y_i^*)e_ie_i^T, & e_ie_n^T \\ e_ne_i^T & , -(y_i^*/k_n^{1*})e_ne_n^T\end{pmatrix} A^{-1}\begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$\geq 0 \qquad \text{for} \qquad i = 1,2,\cdots,I \text{ and } n = 1,2,\cdots,N, \quad \ldots(67)$$

and

$$\partial L/\partial \gamma_{mn}\Big|_c \equiv -\frac{1}{2}[m^T, -\tau^T]A^{-1}\begin{bmatrix} 0_{I\times I}, & 0_{I\times N} \\ 0_{I\times N}, & -(k_n^{1*}/k_m^{1*})e_m e_m^T + e_m e_n^T + e_n e_m^T - (k_m^{1*}/k_n^{1*})e_n e_n^T \end{bmatrix} A^{-1}\begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$\geq 0 \quad \text{for} \quad 1 \leq m < n \leq N . \qquad \ldots .(68)$$

Finally, we can define an analogous compensated increase in δ_{mn} for $m \neq n$. We find that

$$\partial L/\partial \delta_{mn}\Big|_c \equiv \frac{1}{2}[m^T, -\tau^T]A^{-1}\begin{bmatrix} 0_{I\times I}, & 0_{I\times N} \\ 0_{I\times N}, & -(k_n^{2*}/k_m^{2*})e_m e_m^T + e_m e_n^T + e_n e_m^T - (k_m^{2*}/k_n^{1*})e_n e_n^T \end{bmatrix} A^{-1}\begin{pmatrix} m \\ -\tau \end{pmatrix}$$

$$\leq 0 \quad \text{for} \quad 1 \leq m < n \leq N . \qquad \ldots .(69)$$

Thus, compensated increases in α_{ij}, β_{in}, and γ_{mn} lead to increased losses, while compensated increases in δ_{mn} lead to decreased losses.

In the following section, we specialize the general loss formula (47) to the case of one monopolistic output and one transferable capital input, so that we can say more about the determinants of the welfare loss.

6. The One Monopolistic Output, One Transferable Capital Case

First, consider formula (41) when there is only one output in the monopolistic sector ($I = 1$) and there is only one capital good which is used in both industries ($K=1$).[22] In this case, the matrices of coefficients α, β, γ, and δ which we defined in the previous section reduce to scalars, which we will denote by α, β, γ, and δ in this section; i.e.,

$$\alpha \equiv \partial^2 C(w,y^*,k^{1*})/\partial y \partial y, \quad \beta \equiv \partial^2 C(w,y^*,k^{1*})/\partial y \partial k^1$$

$$\gamma \equiv \partial^2 C(w,y^*,k^{1*})/\partial k^1 \partial k^2, \quad \delta \equiv \partial^2 \pi(w,k^{2*})/\partial k^2 \partial k^2 \,. \qquad \ldots .(70).$$

Now (41) may be rewritten as (m and τ are now scalars):

$$\begin{bmatrix} dy(0)dt \\ dk^1(0)dt \end{bmatrix} = - \begin{bmatrix} \alpha, & \beta \\ \beta, & \gamma-\delta \end{bmatrix}^{-1} \begin{bmatrix} m \\ -\tau \end{bmatrix}$$

$$= \frac{-1}{\alpha(\gamma-\delta) - \beta^2} \begin{bmatrix} \gamma-\delta, & -\beta \\ -\beta, & \alpha \end{bmatrix} \begin{bmatrix} m \\ -\tau \end{bmatrix} \,. \qquad \ldots .(71)$$

From (37), we may define the competitive sector's rental price for the tradable capital good as $r^2(t) \equiv \partial\pi(w,k^2(t))/\partial k^2 = -\partial\pi(w,\bar{k}-k^1(t))/\partial k^2$ where we have also used (38): $k^2(t) \equiv \bar{k} - k^1(t)$. Thus we have

$$dr^2(0)/dt = -\delta \; dk^1(k)/dt \qquad \ldots .(72)$$

and

$$dk^2(0)/dt = -dk^1(0)/dt \,. \qquad \ldots .(73)$$

We may also use Hicks [1946; pp. 312-13)[23] Aggregation Theorem in order to aggreagate the x goods associated with the w vector; i.e., define the aggregate net output X as

$$X(t) \equiv \pi(w,k^2(t)) - C(w,y(t),k^1(t)) \quad . \qquad \ldots .(74)$$

Then upon differentiating (74), we obtain

$$\frac{dX(t)}{dt} = [-\frac{\partial\pi}{\partial k^2} - \frac{\partial C}{\partial k^1}]\frac{dk^1(t)}{dt} - \frac{\partial C}{\partial y}\frac{dy(t)}{dt}$$

$$= [-t\tau]\frac{dk^1(t)}{dt} - [p - mt]\frac{dy(t)}{dt} \quad \text{using (39) and (40)}$$

$$\therefore \frac{dX(0)}{dt} = -p\frac{dy(0)}{dt} \qquad \ldots .(75)$$

where p is now the (scalar) price of output for the monopolistic sector (it is the constant revenue per unit output that the monopolist receives and should not be confused with the marginal cost function $p(w,y,k^1) \equiv \partial C(w,y,k^1)/\partial y$).

Our negative definiteness assumption (34) translates into the following restrictions on α, β, γ, and δ:

$$\alpha > 0 \ , \qquad \ldots .(76)$$

$$\gamma - \delta > 0 \ , \quad \text{and} \qquad \ldots ..(77)$$

$$D \equiv \alpha(\gamma - \delta) - \beta^2 > 0 \ . \qquad \ldots .(78)$$

We shall make the following additional assumption:

$\delta \leq 0$.[24] . . . .(79)

What about the sign of $\beta \equiv \partial^2 C(w,y^*,k^{1*})/\partial y \partial k^1 = \partial p(w,y^*,k^{1*})/\partial k^1 = \partial^2 C(w,y^*,k^{1*})/\partial k^1 \partial y = -\partial r^1(w,y^*,k^{1*})/\partial y$? If the monopolist's technology set S is not "unusual," then β will be negative.[25] However, theoretically, β could be positive- so below we will distinguish the "usual" and "unusual" technology cases for the monopolist, corresponding to $\beta < 0$ and $\beta > 0$ respectively.

The table below indicates the direction of change in y (the monopolist's output), k^1 (transferable capital used by the monopolist), k^2 (transferable capital used by the competitive sector), $r \equiv r^2$ (the competitive rental rate for transferable capital) and X (net output of other goods minus labour input) as t increases, evaluated at an initial productively efficient equilibrium (i.e., one that is a solution to the maximization problem (30)). In addition to letting β be positive or negative, we consider three cases. Case (1) is the pure monopoly case, where there is either no rate of return regulation or it is ineffective ($m > 0$ and $\tau = 0$ in this case); Case (2) is the pure overcapitalized case where the monopolist faces effective rate of return regulation and a perfectly elastic demand curve ($m = 0$ and $\tau > 0$ in this case); Case (3) is a general regulated monopoly case where both m (the monopolist's markup) and τ (the effective subsidy placed on transferable capital by the monopolist) are positive.

TABLE 1

Comparative Statics for a Regulated Monopoly

Derivative	Pure Monopoly Case $m>0$,	$\tau=0$	Pure Overcapitalized Case $m=0$,	$\tau>0$	General Case $m>0$,	$\tau>0$
	Usual $\beta<0$	Unusual $\beta>0$	Usual $\beta>0$	Unusual $\beta>0$	Usual $\beta<0$	Unusual $\beta>0$
$dy(0)/dt$	<0	<0	>0	<0	?	<0
$dk^1(0)/dt$	<0	>0	>0	>0	?	>0
$dk^2(0)/dt$	<0	<0	<0	<0	?	<0
$dr(0)/dt$	≤ 0	≥ 0	≥ 0	≥ 0	?	≥ 0
$dX(0)/dt$	>0	>0	<0	>0	?	>0

From Table 1, we see that in the pure monopoly case where the monopolist's technology is not unusual, the monopolist's output and capital input decrease, the competitive rental price of capital does not increase and the competitive output (minus labour input) aggregate X increases, compared to the "efficient" equilibrium (at least for sufficiently small distortions). On the other hand, in the pure overcapitalized case with a usual technology, the comparative statics results are precisely opposite to the pure monopoly case. Thus in the general case with $m>0$, $\tau>0$, and $\beta<0$ (the usual technology case), if m is large relative to τ, the comparative statics will follow the pure monopoly case, while if τ is large relative to m, the comparative statics will follow the pure overcapitalized case.

We now evaluate the loss formula (47) when $I = 1$ and $N = 1$. Using definitions (70), we find that the loss is

$$L = \frac{1}{2}[m, -\tau]\begin{bmatrix}\alpha, & \beta \\ \beta, & \gamma-\delta\end{bmatrix}^{-1}\begin{bmatrix}m \\ -\tau\end{bmatrix}$$

$$= \frac{1}{2}\frac{(\gamma-\delta)m^2 + 2\beta m\tau + \alpha\tau^2}{\alpha(\gamma-\delta) - \beta^2}$$

$$> 0 \quad \text{if} \quad m \neq 0 \quad \text{or} \quad \tau \neq 0 \quad \text{using (76)-(78)} \quad \ldots(80)$$

It is more useful to translate the loss formula (80) into an elasticity framework. Let us express the loss as a fraction of $C(w,y^*,k^{1*})$, the initial (efficient equilibrium) variable cost for sector 1; i.e., define $\hat{L}$ by

$$\hat{L} \equiv L / C(w,y^*,k^{1*}) \quad \ldots(81)$$

where L is defined by (80).

Recall the definitions of α, β, and γ (see (54), (60) and (57) respectively). Define comparable elasticities as follows:

$$\hat{\alpha} \equiv \partial \ell n\, p(w,y^*,k^{1*})/\partial \ell n\, y = y^* \alpha/p^* \quad \ldots(82)$$

where

$$p^* \equiv p(w,y^*,k^{1*}) \equiv \partial C(w,y^*,k^{1*})/\partial y ,$$

$$\hat{\beta} \equiv \partial \ell n\, p(w,y^*,k^{1*})/\partial \ell n\, k^1 = k^{1*}\beta/p^* , \quad \ldots(83)$$

and

$$\hat{\gamma} \equiv -\partial \ell\, r^1(w,y^*,k^{1*})/\partial \ell n\, k^1 = k^{1*}\gamma/r^{1*} \quad \ldots .(84)$$

where $r^{1*} \equiv r^1(w,y^*,k^{1*}) \equiv -\partial C(w,y^*,k^{1*})/\partial k^1$. Let us also convert the ad valorem markup and capital subsidy variables, m and τ, into percentage variables; i.e., define

$$\hat{m} \equiv m/p^* \; ; \quad \hat{\tau} \equiv \tau/r^{1*} \; . \quad \ldots .(85)$$

Note that at the efficient equilibrium, total revenue in sector 1 (the monopolist's sector) can be defined as p^*y^*. The initial efficient equilibrium capital rental total is $r^{1*}k^{1*}$ and the initial wage plus intermediate materials bill is $C(w,y^*,k^{1*})$. Define the following ratios:

$$s_x \equiv C(w,y^*,k^{1*})/p^*y^* \; ; \quad s_k \equiv r^{1*}k^{1*}/p^*y^* \; . \quad \ldots .(86)$$

If the technology exhibits constant returns to scale in sector 1, then s_x will be labour's share and s_k will be capital's share of total revenue; i.e., with constant returns,

$$s_x + s_k = 1 \; . \quad \ldots .(87)$$

However, in the case of a general technology, (87) need not hold. Let us define ρ to be the percentage increase in variable costs due to a one percent increase in y and k, evaluated at the efficient equilibrium point; i.e., define

$$\rho \equiv \left. \frac{\partial \ell n\, C(w,\lambda y^*,\lambda k^{1*})}{\partial \lambda} \right|_{\lambda=1}$$

$$= [y^* \frac{\partial C(w,y^*,k^{1*})}{\partial y} + k^{1*} \frac{\partial C(w,y^*,k^{1*})}{\partial k^1}] / C(w,y^*,k^{1*})$$

$$= [y^* p^* - k^{1*} r^{1*}] / C(w,y^*,k^{1*}) \quad . \qquad \ldots .(88)$$

If $\rho < 1$, then the technology for sector 1 exhibits local increasing returns to scale, while if $\rho > 1$, the technology exhibits local decreasing returns to scale around the efficient equilibrium point. If we multiply both sides of (88) through by $C(w,y^*,k^{1*})/p^*y^*$, we find that (88) is equivalent to

$$\rho s_x = 1 - s_k \quad . \qquad \ldots .(89)$$

Note that if the technology set S exhibits constant returns to scale (i.e., S is a cone so that $C(w,y,k^1)$ is linearly homogeneous in y,k^1, then $\rho = 1$ and (89) reduces to (87).

We have not introduced an elasticity for δ, since we are going to assume that $\delta = 0$. [26]

Assuming that $\delta = 0$, substituting (80), (82)-(86) into (81) yields the following formula for $\hat{L} \equiv L/C$:

$$\hat{L} = \frac{1}{2} s_k s_x^{-1} [\hat{m}, -\hat{\tau}] \begin{bmatrix} s_k \hat{\alpha}, & \hat{\beta} \\ \hat{\beta} & , \hat{\gamma} \end{bmatrix}^{-1} \begin{bmatrix} \hat{m} \\ -\hat{\tau} \end{bmatrix}$$

$$= s_k \left\{ \hat{\gamma}\hat{m}^2 + 2\hat{\beta}\hat{m}\hat{\tau} + s_k\hat{\alpha}\hat{\tau}^2 \right\} / 2s_x \left\{ s_k\hat{\alpha}\hat{\gamma} - \hat{\beta}^2 \right\} . \qquad \ldots .(90)$$

In Table 2 below, we evaluate the loss $\hat{L}$ for various parameter values.[27]

We table ρ as well as s_x and s_k, where the three parameters satisfy (89). Recall that $\rho < 1$ corresponds to locally increasing returns to scale for the monopolist.

TABLE 2

The Loss $\hat{L}$ For Various Parameter Values

s_x	s_k	ρ	$\hat{m}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{L}$
1/2	1/2	1	1/5	1/5	1/2	1/2	-1/4	8.0%
1/2	3/5	4/5	1/5	1/5	1/2	1/2	-1/4	8.2%
2/5	1/2	5/4	1/5	1/5	1/2	1/2	-1/4	10.0%
1/2	1/2	1	1/5	1/5	1/2	1/2	0	12.0%
1/2	1/2	1	1/5	1/5	1/2	3/2	-1/2	12.0%
1/2	3/5	4/5	1/5	1/5	1/2	1	-1/2	14.4%
1/2	3/5	4/5	1/5	1/5	1	1	-1/4	4.9%
1/2	3/5	4/5	1/5	0	1/2	1/2	-1/4	16.0%
1/2	3/5	4/5	0	1/5	1/2	1/2	-1/4	16.5%
1/2	1/2	1	1/5	1/5	1/2	1/2	1/4	40.0%
1/2	3/5	4/5	1/5	1/5	1/2	1/2	1/4	35.7%

Inspection of (90) shows that $\hat{L}$ is homogeneous of degree 2 in $\hat{m},\hat{\tau}$, homogeneous of degree minus one in s_x and homogeneous of degree minus one in $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$. Thus if we double $\hat{m}$ and $\hat{\tau}$ in Table 2, the corresponding $\hat{L}$ will increase by a factor of four. On the other hand, if we double s_x (and half the corresponding ρ so that (89) remains true), then the corresponding $\hat{L}$ will fall to one half of the tabled level. Finally, if we divide $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ by one half, then the corresponding $\hat{L}$ will double.

Inspection of Table 2 and use of the above homogeneity properties of $\hat{L}$ suggest that: (i) the loss becomes larger the smaller are the elasticities $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, (ii) the loss becomes larger the larger are the distortions $\hat{m}$ and $\hat{\tau}$, (iii) the loss becomes larger the smaller s_x (or the larger ρ) becomes, (iv) the loss becomes larger as either one of the distortions $\hat{m}$ or $\hat{\tau}$ decrease to zero if we are in the usual technology case with $\hat{\beta} < 0$, and (v) the loss becomes larger as $\hat{\beta}$ increases from negative to positive values; i.e., as we change from the usual technology case to the unusual technology case.

Before turning to the geometry of the one output, one capital case, it is useful to develop a more familiar economic interpretation for the "usual" and "unusual" technology cases. However, it is necessary to make an additional assumption:[28]

$$\gamma \equiv \partial^2 C(w,y^*,k^{1*})/\partial k^1 \partial k^1 > 0 \quad . \qquad \text{....(91)}$$

Let $k^1(y)$ be the solution to the following cost minimization problem.

$$\min_{k^1} \{ C(w,y,k^1) + r^* k^1 \} \quad . \qquad \text{....(92)}$$

The first and second order conditions for (92) are

$$\partial C(w,y,k^1(y))/\partial k^1 + r^* = 0 \quad ; \qquad \text{....(93)}$$

$$\partial^2 C(w,y,k^1(y))/\partial k^1 \partial k^1 > 0 \quad . \qquad \text{....(94)}$$

When $y = y^*$, $k^{1*} \equiv k^1(y)$ satisfies (93) (recall (36) with $t = 0$), and (94) is satisfied by assumption (91). The Implicit Function Theorem guarantees the existence of the solution $k^1(y)$ to (92) for all y in a neighbourhood of y^*, and we also have

$$\frac{\partial k^1(y^*)}{\partial y} = -\frac{\partial^2 C(w,y^*,k^{1*})}{\partial k^1 \partial y} \frac{\partial^2 C(w,y^*,k^{1*})}{\partial k^1 \partial k^1} = -\beta\gamma \quad . \qquad \ldots .(95)$$

Thus since $\gamma > 0$ by (91), $-\beta \geq 0$ if and only if $\partial k^1(y^*)/\partial y \geq 0$; i.e., if and only if capital is not an inferior input in a neighbourhood of the initial equilibrium in the monopolistic sector.

7. The Geometry of the One Monopolistic Output, One Transferable Capital Case

Consider the constrained maximization problem (30) with $I = 1$ (one monopolistic output) and $N = 1$ (one transferable capital good). We hold w fixed throughout, and thus we may apply Hick's Aggregation Theorem and define the aggregate net output of sector 2 as

$$X^2 \equiv \pi\,(w, k^2) \quad . \qquad \ldots .(96)$$

We also define the aggregate net non capital input used in sector 1 as

$$X^1 \equiv C\,(w, y, k^1) \qquad \ldots .(97)$$

and aggregate (nonmonopolistic) output for the economy as

$$X \equiv X^2 - X^1 \quad . \qquad \ldots .(98)$$

The aggregate good X can be taken to be the numeraire good, and thus the price p in (30) is to be interpreted as the price of the monopolistic output relative to this numeraire good.

Our first task is to show how the economy's feasible set of outputs (y,X) can be geometrically derived from a knowledge of the π and C functions.

Turn to Panel (i) of Figure 1. The curved line represents the locus of (X^2, k^2) points that are consistent with (96); i.e., it is simply sector 2's production function.

Consider the family of curved lines in panel (iv) of Figure 1. There are curved lines drawn for output levels $y = 1, 2, \cdots, 7$. For each y, the corresponding line represents the set of (X^1, k^1) that are consistent with (97); i.e., each curved line corresponds to the set of input combinations (X^1, k^1) that can produce the corresponding output level y and thus each curved line is an isoquant for sector 1. We have the X axis pointing downward in panel (iv) and upward in panel (i), but the reason for this will be explained later.

Define the marginal product of capital in sector 2 as

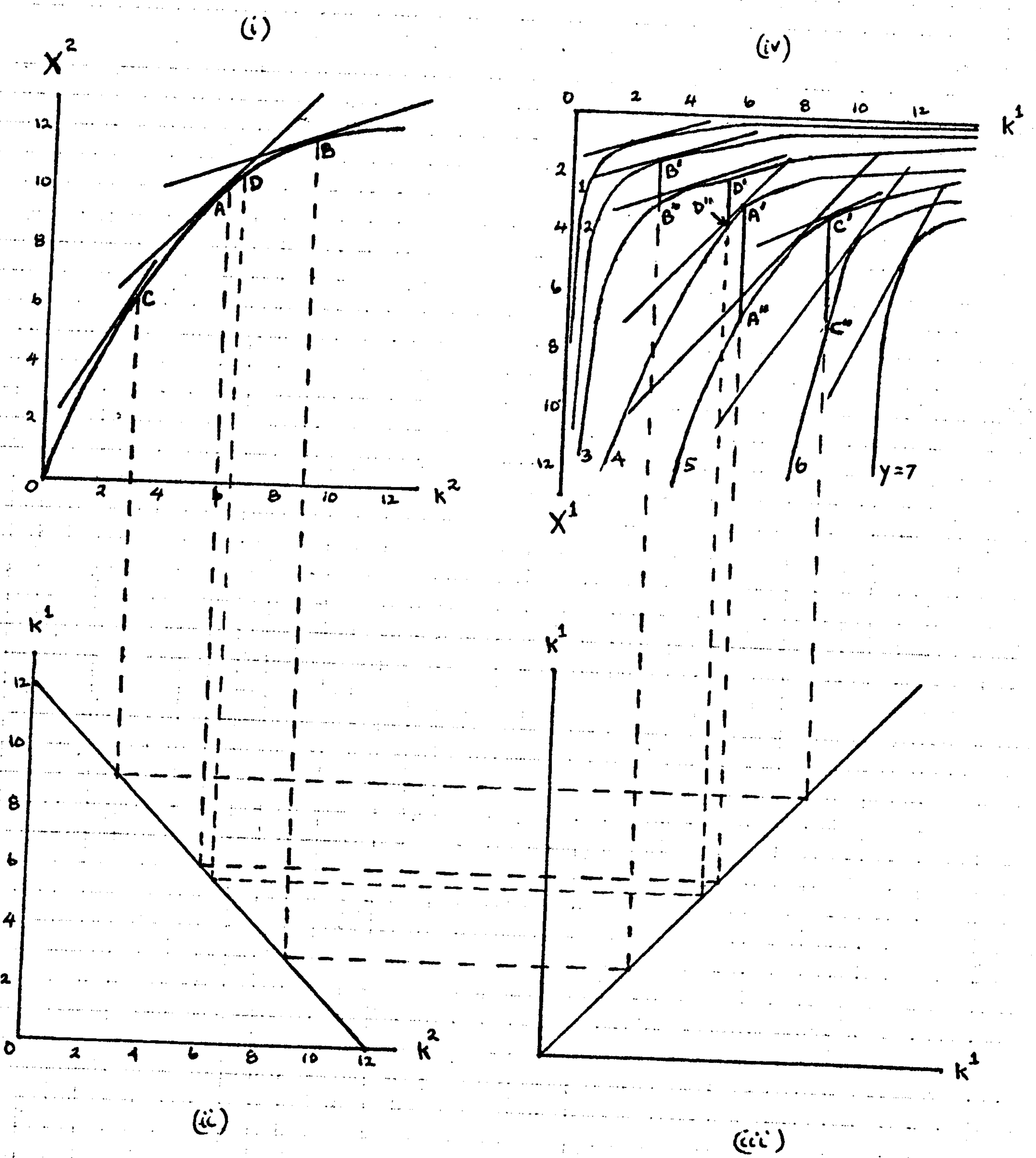
$$r^2(k^2) \equiv \partial\pi\,(w, k^2)/\partial k^2 \qquad \ldots.(99)$$

and the marginal rate of substitution of capital for aggregate non capital input in sector 1 as[29]

$$r^1(y, k^1) \equiv -\partial C\,(w, y, k^1)/\partial k^1 \;. \qquad \ldots.(100)$$

We know from (33), that efficient (y, k^1, k^2) points are characterized by

Figure 1



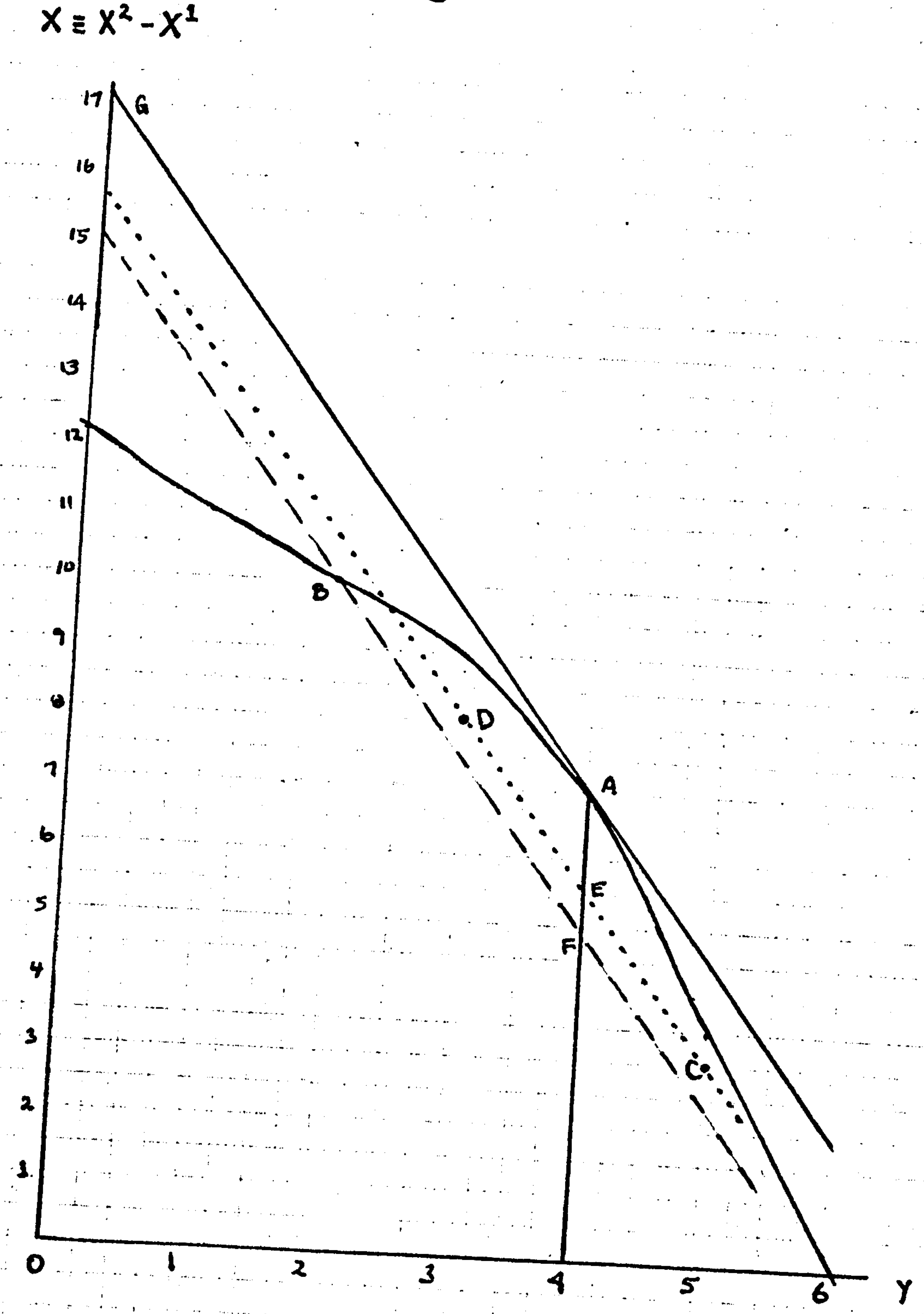
$$r^1(y,k^1) \quad = \quad r^2(k^2) \quad . \qquad \ldots .(101)$$

Panel (ii) of Figure 1 represents the capital constraint on the economy: $k^1 + k^2 = \overline{k}$. Panel (iii) is simply an auxiliary diagram, which represents the equation $k^1 = k^1$. The basic data for the economy are the curved lines in panels (i) and (iv) and the capital constraint in panel (ii).

Start with a point on the curved line in panel (i), say the point A, which corresponds to $X^2 = 10$, $k^2 = 6$. Follow the dashed line from A down to panel (ii) and read off how much capital will be available for sector 1 (k^1 turns out to be 6 also) and follow the dashed line over to panel (iii) and up to panel (iv) up to the point A'. Note that the slopes of the straight lines tangent to the curves through A and A' are equal. This tangency corresponds to the equality in (101) and thus we have found an efficient point. At the point A', $y = 4$, $X^1 = 3$ and $k^1 = 6$. Thus $y = 4$ and $X = X^2 - X^1 = 10 - 3 = 7$ is an efficient output combination, and it appears as the point A in Figure 2.

Now pick another point on the production function in panel (i), say B and follow the dashed line to the point B' in panel (iv) of Figure 1. Again, note that the slopes at B and B' are equal, so we have found another efficient point (point B in Figure 2). Repeat this process for all k^2 between 0 and 12, and we trace out the curved line in Figure 2. This is the efficient set of outputs for the economy. To pick out a particular point on this efficient set, we need to know the price of y relative to X, p. We have chosen p so that the point A on Figure 2 is chosen (p is equal to minus the slope of the tangent line through the point A).

Figure 2


$X \equiv X^2 - X^1$
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
Y
G
B
D
A
E
F
C

From (32), p is also equal to $\partial C(w,y^*,k^{1*})/\partial y = \partial X^1(y^*,k^{1*})$ = the change in X^1 due to a small change in y, holding k^1 constant. Thus p is approximately equal to the length of the line joining the points A' and A".

The point B in Figure 2 corresponds to the B, B' equilibrium in Figure 1. Note that the corresponding p is approximately equal to B'B" and thus the monopolist's markup m is approximately A'A" - B"B" = 1. Thus the B equilibrium in Figure 2 is a pure monopoly equilibrium and the loss of output to the economy ΔW is equal to AF and the percentage loss L_B is AF/GO = 2/17 = 11.8%.

Consider now a pure overcapitalization equilibrium with $m = 0$ and $\tau = 1$. The solution turns out to be the CC' equilibrium in Figure 1. Compare the slope of the tangent line at C' to the slope of the tangent line at C. The difference in slopes is equal to $\tau = r^2 - r^1 = 1$. Note also that A'A" is equal to C'C" so that both the A and the C equilibrium have the same selling price for the monopolistic good. This pure overcapitalization equilibrium corresponds to the point C in Figure 2. Note that C is in the interior of the locus of (X,y) efficient output points, and the percentage loss of output associated with the C equilibrium L_C is AE/GO = 1.5/17 = 8.8%.

Now consider a distorted equilibrium with $m = 1$ and $\tau = 1$. This corresponds to the DD' equilibrium in Figure 1. Note that D'D" = B'B" so that the B and D equilibria have the same markup and that the slopes of the tangent lines to D and D' differ by $\tau = 1$. This DD' equilibrium corresponds to the point D in Figure 2, which is also in the interior of the efficient set. Fortuitously, the percentage loss of output associated with this D equilibrium L_D is also equal to AE/GO = 8.8%. Note that the dashed and dotted lines in Figure 2 are parallel to the tangent line at A.

8. Conclusion

Regulated industries are usually regulated because if left unregulated, the industries behave in a monopolistic manner; i.e., there would be a divergence between the marginal cost of producing an output and its selling price. Unfortunately, regulation can often introduce additional distortions without eliminating monopolistic markups, e.g., in the Averch-Johnson model of regulation, the regulated firm will tend to use "too much" capital. In this paper, we have provided a methodology which will enable one to calculate the loss of output when there are distortions within the production sector of an economy or when there is monopolistic pricing behaviour.

The methodological approach presented in section 3 was a pure productive efficiency approach: we asked how much extra output could be produced if we eliminated the distortions within the production sector of an economy, holding constant other outputs and aggregate inputs into the production sector.

In section 4, we observed that the approach developed in section 3 was not entirely satisfactory for evaluating the losses due to monopolistic pricing behaviour. Thus in section 4, we abandoned the pure productive efficiency approach in favour of an approach that depended on producer prices as well.

In section 5, we showed how the loss measure introduced in section 4 depended on various parameters, while sections 6 and 7 considered the one monopolistic output, one transferable capital case from an algebraic and geometric point of view.

One of the main advantages of our methodology for evaluating the losses due to distortions within the production sector is that it requires information on technology only; information on consumer's preferences is not required. Furthermore, if we are willing to approximate the loss quadratically, we do not even require global information on the technology. All that is required are estimates of various production elasticities and the distortions (the monopolistic markups, implicit capital subsidies, etc.); i.e., all that is required is <u>local</u> information on the technology.

However, our approach has some disadvantages as well: (i) if the returns to scale in sector 1 are very large, then the negative definiteness assumption (34) may not be satisfied, (ii) our approach considers only the loss of output induced by distortions within the production sector compared to an optimal point, and does not consider possible losses of utility induced by the distortions,[30] (iii) if the distortions are very large, then approximating the unobserved matrix A defined by (34) by the observable matrix $\tilde{A}$ defined by (46) may yield a rather poor approximation, (iv) we have assumed that the rest of the production sector is competitive, a rather dubious assumption, (v) although we have reduced the informational requirements that a full blown general equilibrium approach would require, our informational requirements on the technology are still high, and (vi) our approach is essentially static.[31]

In addition to the above criticisms, one could also criticize the present paper on the grounds that it does not provide answers to the following fundamental questions: (i) how do we determine the socially desirable producer prices (or more generally, the socially desirably Pareto optimal point), and (ii) given that we can define the optimal state, how can we induce producers to get to it?

For answers to the above questions, the reader will have to turn to the other papers presented at this conference.[32] However, once the socially desirable producer prices have been determined, the present paper offers a useful approach to comparing the losses of real output generated by different regulatory schemes.

Footnotes

[1]See Bailey [1973] for a nice review of the literature with many extensions.

[2]Notation: $y \geq 0_I$ means each component of the I dimensional vector y is nonnegative, $y \gg 0_I$ means each component is positive, and $p \cdot y$ (or $p^T y) \equiv \Sigma_{i=1}^{I} p_i y_i$ denotes the inner product of the (column) vectors p and y.

[3]The firm is allowed to expense the n^{th} type of capital services at a rental price of $r_n + e_n \equiv s_n$ where s_n is a "fair rate of return" rental rate set by the regulatory authority.

[4]See Theorem 1 in Diewert [1981a].

[5]If $C(w,y^*,k^*)$ is differentiable with respect to the components of w, then the optimal input vector is $x^* \equiv \nabla_w C(w,y^*,k^*) \equiv [\partial C(w,y^*,k^*)/\partial w_1, \cdots, \partial C(w,y^*,k^*)/\partial w_J]^T$ (Shephard's [1953][1970] Lemma).

[6]See Theorem 2 in Diewert [1981a].

[7]If sector i is using good 0 as an input, then z_0^i is minus the minimum amount of input required to produce the vector of net outputs $z^i \equiv (z_1^i, z_2^i, \cdots, z_M^i)^T$.

[8]Define $f^i(z^i) \equiv -\infty$ if there is no z_0^i such that $(z_0^i, z^i) \in S^i$ where S^i denotes the sector i production possibilities set.

[9]Or minimize the net input required of good 0 in order to produce the net output vector $\bar{z}$.

[10]We can do this because (15) holds.

[11]By assumption (15), the solution to (12) will be locally unique.

[12]For a reconciliation of these various approaches to the measurement of welfare loss due to distortions, see Diewert [1981b].

[13]The degree of returns to scale of f^i evaluated at z can be defined as $\partial \ell n\, f^i(\lambda z)/\partial\lambda\big|_{\lambda=1} = z^T \nabla f^i(z)/f^i(z)$. If this number is greater than 1, then f^i exhibits increasing returns to scale at z.

[14]The concept of the variable profit function is due to Samuelson [1953], who called it a national product function. Properties of π and references to the literature are found in Gorman [1968] and Diewert [1974].

[15]Decentralized price taking competitive producers will also solve (30) if the technology sets S and S^c are both convex. However, the convecity assumption is not required for the analysis which follows.

[16]We have used the well known rule for differentiating a quadratic function of x, say $x^T Bx$, with respect to the components of x: $\nabla_x(x^T Bx) = 2Bx$ where B is a symmetric matrix.

[17]If the technology set S is convex, then McFadden [1978] and Diewert[1978] show that $\begin{bmatrix} \alpha, & \beta \\ \beta^T, & \gamma \end{bmatrix}$ is a positive semidefinite matrix, where α, β and γ

are I by I, I by N and N by N matrices of the α_{ij}, β_{in}, and γ_{mn}, respectively. Also if the technology set S^c is convex, then Diewert [1974] shows that $\delta \equiv [\delta_{mn}]$ is a negative semidefinite matrix. Under these conditions, $\alpha_{ii} \geq 0$ for each i, $\gamma_{nn} \geq 0$ for each n and $\delta_{nn} \leq 0$ for each n. If S exhibits constant returns to scale, then it is easy to show (see Diewert [1974] and McFadden [1978]) that the following restrictions are satisfied: $\begin{pmatrix} \alpha , & \beta \\ \beta^T, & \gamma \end{pmatrix} \begin{pmatrix} y \\ k^{1*} \end{pmatrix} = \begin{pmatrix} 0_I \\ 0_N \end{pmatrix}$.

Similarly, if S^c exhibits constant returns to scale, $\delta k^{2*} = 0_N$.

[18]If a matrix A has elements which are functions of a parameter α, then by differentiating the identity $A(\alpha)A^{-1}(\alpha) = I$, it can be shown that $d\{A^{-1}(\alpha)\}/d\alpha = -A^{-1}(\alpha)\{dA(\alpha)/d\alpha\}A^{-1}(\alpha)$.

[19]If S^c is convex, then $\delta_{nn} \leq 0$ and as δ_{nn} increases, the absolute value of δ_{nn} decreases until $\delta_{nn} = 0$.

[20]See footnote 17.

[21]Our technique is analogous to that used by Sato and Koizumi [1970], except that they use analogues of equations (64) to eliminate the α_{ii} altogether.

[22]Of course, there can be many specific capital goods in each sector that cannot be transfered to the other sector.

[23]See Diewert [1978] for additional references.

[24]This assumption will be satisfied if the competitive sector's production possibilities set S^c is convex. If S^c is a cone, so that

there is constant returns to scale in the competitive sector and there are no specific capital goods in the competitive sector, then $\delta = 0$.

[25]If S is a convex cone, then using footnote 17 and assumption (76), β must be negative.

[26]This is consistent with there being no specific capital goods in the competitive sector and constant returns to scale in the competitive sector. If in fact $\delta < 0$, then we need only adjust γ and γ^* upwards appropriately.

[27]With $\delta = 0$, $s_x > 0$ and $s_k > 0$, the parameters $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ should satisfy the counterparts to the restrictions (76) to (78); i.e., $\hat{\alpha} > 0$, $\hat{\gamma} > 0$ and $s_k\hat{\alpha}\hat{\gamma} - \hat{\beta}^2 > 0$.

[28]If the technology set S is convex, then $\gamma \geq 0$. Assumption (91) corresponds to local strict convexity of $C(w,y,k^1)$ in k^1 around the point (w,y^*,k^{1*}).

[29]From (60), $-\beta \equiv \partial r^1(w,y^*,k^{1*})/\partial y = -\partial^2 C(w,y^*,k^{1*})/\partial k^1 \partial y \equiv -\partial^2 X^1(y^*,k^{1*})/\partial k^1 \partial y$. In panel (iv) of Figure 1, $-\beta > 0$ since as we travel from A' to A", the slope of the tangent line becomes steeper.

[30]Loosely speaking, our approach provides estimates of the loss of producer's surplus but not of consumer's surplus. The problem is that some consumers could gain and some could lose as a result of the distortions, compared to a Pareto optimal state, and thus we would have to make interpersonal comparisons of utility. However, given information on

consumer's preferences, a "total" loss measure that did not make interpersonal comparisons could be derived using Debreu's [1951][1954] coefficient of resource utilization idea. See Diewert [1981b] for the details of this approach in the context of a somewhat different model.

31 However, I would conjecture that the output loss generated by the noncompetitive behavior of a regulated firm would only increase in the context of a dynamic model which made aggregate capital stocks choice variables; i.e., investment would be induced into the wrong industries due to the static distortions.

32 In particular see Warskett and de Fontenay [1981] and the literature they refer to.

THE REGULATION OF TELECOMMUNICATION AND VERTICAL INDUSTRY STRUCTURE

by

Fred M. Westfield
Vanderbilt University

This paper examines some economic consequences of monopoly regulation in a vertical industry structure. Our model consists of a "downstream industry" and an "upstream industry". The downstream industry may be an unregulated monopoly or a regulated monopoly. It can also be considered a "competitive" industry when price is made to equal to marginal cost. A single product (q) is sold at a single price (p) to final consumers, though the analysis may be generalized to allow for discriminatory pricing (multi-part tariffs) and for more than one final product. The "upstream industry" supplies an intermediate product or service (y), when the industry is not vertically integrated, at price s, to the downstream industry which uses y as an input in the production of the downstream product or service. The downstream industry also requires an input (x) which can be substituted, as determined by the production function, for the input y. Input x is sold at price, w, which throughout our analysis remains constant and can thus serve as a numeraire.

The scenario is to examine this vertical structure, first, under the assumption that the upstream and downstream industries are one and the same, that the upstream industry's output, namely y, is produced by the downstream industry. We illuminate the economic consequences of various regulatory rules and evaluate their effects on social waste--on deadweight losses.

We next examine the production chain in the absence of vertical integration. Now upstream output is supplied by monopoly or possibly competitively. Regulation of upstream monopoly will affect the quantities y, and price, s, which, in turn, affect the downstream cost structure and p and q for final product. What are the effects, upstream and downstream, of various regulatory rules? How are deadweight losses affected by various regulations?

As this paper is meant to shed light on regulatory problems for telecommunications, we need to provide some interpretation to the notions upstream and downstream industries. One interpretation is that downstream or final product is, say, "telephone service" while the upstream industry represents, say "equipment supplies." An alternative interpretation is that downstream product represents final output in the national income sense, while the upstream industry represents telephone services which serve as an input in the production of final product. As we are here considering a vertical chain with only two links, one may choose the interpretation. A more elaborate analysis would deal with both interpretations, more than two links in the vertical chain.

THE MODEL

The quantity of final product sold is designated by q. Input quantities required are x and y, and the production function for final product is assumed linear-homogeneous

$$(1) \qquad q = F(x, y)$$

so that the unit-isoquant is given by

$$(2) \qquad 1 = F(x/q, y/q) = F(a, b).$$

Prices for the inputs are assumed to be parametric from the point of view of the purchaser. Consequently, given the price w for x and s for y, cost minimization by producers of final product, together with the homogeneity assumption, determines the input coefficients:

$$(3) \qquad a = a(s/w), \; b = b(s/w),$$

as functions of relative input prices.

(4) $a' \geq 0, b' \leq 0$

because of convex-to-the-origin isoquants. The equalities of (4) remind us that increased relative prices of inputs may at isoquant "kinks" or "corners" result in unchanged input coefficients.

Consider the input y as resource or intermediate good. Its unit cost of production is r, independent of the amount produced or of the vertical structure. Without vertical integration a producer of the final product q cannot also enter the industry producing y. He must purchase at the price s from a supplier who obtains it at a constant unit cost r. With vertical integration, the unit cost of y to the combined firm is also r. In effect s=r. The other input, x, is sold at the constant price w, whether production is vertically integrated or not.

Under these assumptions average and marginal cost of downstream production, λ, do not depend on quantity of final product q. For given prices s of y and w of x

(5) $\lambda = wa(s/w) + sb(s/w) = \lambda(w, s)$,

where a and b satisfy a generalized cost-minimizing isoquant-isocost tangency

(6) $wa' + sb' = 0.$

Demand for final product is

(7) $q = D(p), D' < 0,$

where p is the price of final product. Inverted, this function gives average revenue $p = p(q)$ and total revenue

(8) $R = p(q)q = R(q)$.

Marginal revenue, R' is assumed positive for sufficiently small q, i.e., demand is price elastic at such a point, and the slope $R'' < 0$.

INTEGRATED INDUSTRY

If the industry is vertically integrated the price s of the input y is its cost r. Designating by subscript "o" the equilibrium magnitudes resulting from this organization of production, marginal and average cost of final product is

(9) $\lambda_o = a_o w + b_o r = \lambda(w, r)$.

The input-output ratios a_o and b_o, given by (3) with $s=r$, satisfy the cost-minimization condition (6).

Downstream Price Equals Marginal Cost

If regulation, or else competition set $\bar{p} = \lambda_o$ (marginal cost pricing ing), then output is determined by the market clearing conditions $\bar{q} = D(\bar{p})$. And in the absence of externalities, of price-marginal cost wedges elsewhere, or of inappropriate income distribution, it is a well known theoretical result that the equilibrium configuration is socially optimal. Pareto Optimality implies absence of deadweight losses.

Unregulated Downstream Monopoly

Without regulation integrated monopoly leads to the profit maximizing output q_o determined by the marginal revenue-marginal cost equality

(10) $R'_o = R'(q_o) = \lambda_o$

and monopoly equilibrium price p_o, determined by the average revenue function or from

(11) $p_o = R_o'E_o/(E_o - 1)$,

where $E_o = -d \ln q/d \ln p$, the value of the price elasticity of demand evaluated at q_o, a positive number greater than unity.

Assuming no externalities and no price-marginal cost wedges elsewhere, the triangular area $\frac{1}{2}(p_o-\lambda_o)(\bar{q}-q_o)$ can represent the deadweight loss from integrated monopoly (Figure 1).

Regulated Downstream Monopoly

If the minimum costs λ_o were known to the regulators, they could order the price $\bar{p}$ to equal λ_o and, as we have seen, eliminate the deadweight losses entirely. And by setting a price between p_o and $\bar{p}$ they, in effect, determine the deadweight loss triangle because the regulated firm would find it profitable to produce all output demanded at such a ceiling and to produce this output efficiently at its lowest possible unit cost, λ_o.

But regulators usually are not in a position to specify such a price. They do not know the engineering details required for λ_o. They may have data provided by the firm; but one cannot expect a firm to report data to regulators if to do so is unprofitable and not to do so is legal. The regulators apply a formula related to reported costs, or to some component of reported costs, in order to determine price.

Markup Constraint

The widely studied formula, limiting rate of return and leading to what has been referred to as the Averch-Johnson effect, will be reviewed below. A simpler formula allows price to be determined as a percentage markup, say $m > 1$, on all of a firm's cost components, not merely on some components of cost as does rate of return regulation. Such a formula has the desirable property of reducing the size of the unconstrained monopoly's deadweight loss triangle without causing new deadweight losses, provided m is less than p_o/λ_o (i.e. the constraint is effective) and also provided the price p satisfying the constraint $m\lambda_o = p$ lies on the elastic portion ($E > 1$) of the product's demand curve. The profit maximizing monopolist subject to such markup constraint will combine inputs and produce output at the minimum unit cost λ_o. Price is lowered and ouput increased from the monopoly level according to the demand curve, so as to satisfy the regulatory constraint $m\lambda_o \geq p$ as equality.

It does not seem to be widely appreciated, however, that this regulatory constraint will be responsible for added deadweight losses if p, satisfying $m\lambda_o = p$, would lie on the inelastic portion ($E < 1$) of the demand curve. The profit maximum requires that p,q not lie on the inelastic portion of a demand curve (for which $R'(q) < 0$). It is more profitable to charge the price p_u and sell q_u determined by $R'(q_u) = 0$, at a point of unitary price elasticity. Rather than lowering price to equal $m\lambda_o$, the firm combines inputs inefficiently, increasing technologically unnecessary costs, $aw + br = \lambda > \lambda_o$, to justify the price $p_u = m\lambda$ to the regulatory authorities. The <u>padding</u> of costs in this fashion adds less to total cost than does the higher price (which additional cost allows) adds to toal revenue. These padded costs equal $(\lambda - \lambda_o)q_u = (p_u/m - \lambda_o)q$ and are an offset to reduction in deadweight losses from $\frac{1}{2}(p_o - \lambda_o)(\bar{q} - q_o)$ to $\frac{1}{2}(p_u - \lambda_o)(q_u - q_o)$ that regulation does accomplish.

These results are deduced rigorously by analyzing the maximum of

$$(12) \qquad R - \lambda q$$

where $\lambda = wa + rb \geq \lambda_o$, a and b are free to move along the isoquant $F(a,b) = 1$ and

(13) $R \leq m\lambda q$

(14) $R \leq R(q)$

(15) $\lambda_o \leq \lambda$

$R, q, a, b \geq 0.$

Here (13) is the regulatory constraint, (14) allows the firm to operate "inside" the demand curve and (15) allows for input combinations other than those yielding minimum costs.

Assuming $1 < m < p_o/\lambda_o$, i.e. m effective and viable, and a solution for which R, q, a, b > 0, analysis of the Kuhn-Tucker conditions reveals that (13) and (14) are satisfied in the solution as equalities. However, (15) may be satisfied in the solution as equality or as inequality. When satisfied as equality, $R' \geq 0$. The profit maximum subject to constraint occurs at a point on the demand curve with price elasticity $E \geq 1$. When (15) in the solution is satisfied as inequality, $R' = 0$. The profit maximum occurs at P_u, q_u: on the demand curve at $E = 1$. The maximum profit for the firm is $R(q_u) - \lambda q_u = R(q_u)(1-1/m)$. Note if the firm were to produce any other ouput $q \gtrless q_u$ that satisfies (15), then its profit is $R(q)(1 - 1/m)$, which, because $R''(q) < 0$, is a smaller amount.

All this is summarized in Figure 2. Figure 2a shows the case where the markup is high enough to lead to a solution on the elastic position of the demand curve. There is no cost padding; and the deadweight loss of unregulated monopoly is unambiguously reduced. In Figure 2b the markup on "efficient cost" would lead to a price-output point on the inelastic position of the demand curve. The solution occurs at the price-output point at $E = 1$. While the unregulated monopoly's deadweight loss triangle is reduced in area, there is an increase in deadweight loss, represented by the rectangle of padded costs. In this case there may be a net increase or decrease in deadweight loss as a result of regulation.

Rate of Return Regulation

Where integrated monopoly regulation is not based on markup of all cost, but as markup of some costs, as, in effect, rate of return regulation, analogous deadweight losses result from cost padding when demand is inelastic. But in this case there are also added deadweight losses when demand is elastic. (See [3] and [4]).

The added deadweight losses with elastic demand are clearly illustrated by the Averch-Johnson model. If input y is considered to be "capital", the regulatory constraint is written as

$$p(q) - wa(\gamma/w) - \rho b(\gamma/w) \leq 0,$$

where $\rho > r$ is the allowed rate of return (markup allowed on capital cost) by the regulatory authority and γ is a shadow price of capital.

It is well know that with $R'(q) > 0$ $(E > 1)$, the profit maximum, subject to constraint is given by the solution of

$$(16) \qquad p(q - wa(\gamma/w) - \rho b(\gamma/w) = 0.$$

$$(17) \qquad R'F_y = \gamma < r$$

$$(18) \qquad R'F_x = w \quad ,$$

where $\gamma < r$ implies that the regulatory constraint is binding.

Because of the assumed linear homogenous production function (Euler's theorem), one obtains

$$R'(q) = wa(\gamma/w) + \gamma b(\gamma/w)$$

where, because $\rho > r$ and $\gamma < r$, $R'(q) > \lambda_o$. Calling the profit maximizing downstream product price and output p_A and q_A, respectively, it can be proven that $\bar{p} < p_A < p_O$, $\bar{q} > q_A > q_o$. As output q_A could be produced at unit cost λ_O, but is because of the regulation induced substitution of capital (b) for non-capital (a) produced at unit cost $Z_A = wa(\gamma/w) + rb(\gamma/w) > \lambda_o$, there is a regulation induced deadweight loss $(Z_A - \lambda_O)q_A$ which offsets the reduction in deadweight loss due to lower price and increased output $[q_A - q_O][\frac{1}{2}(P_A - p_O) + (p_A - \bar{p})]$ Figure 3 illustrates all this.

In sum, even where demand is elastic regulations may be a source of net increase in deadweight losses. Moreover, where demand is inelastic, cost padding will occur also under rate of return regulation. Such added costs are deadweight losses.

The analysis of efficiency does suppose that r and w equal social marginal costs. If, for example, r is above the social marginal cost of "capital" then the regulation induced change in input proportions favoring capital may actually reduce deadweight losses. The so called Averch-Johnson distortion may then improve input proportions as price of output is reduced. On the other hand, if w is higher than social marginal cost of non-capital, then the Averch-Johnson effect further distorts input proportions.

We now turn to examine the effects of regulation when vertical integration is absent.

NON-INTEGRATED INDUSTRY

Upstream Marginal Cost Pricing

If upstream price s of the intermediate product y equals marginal cost r then outcomes, with downstream regulation or without, would be the same with vertical integration or without. Only direct technological interaction between the separate vertical links, i.e. external technological economies or diseconomies, would upset this conclusion. This case does not require further attention.

Downstream Marginal Cost Pricing

Suppose downstream competition, or else downstream regulation, sets downstream price equal to marginal cost recorded by the downstream firm, then some deadweight losses will generally occur when s exceeds r. The deadweight losses derive from two sources. The first is the higher level of downstream marginal cost and resulting higher downstream price that restricts downstream output. The second is technical substitution against y because of its higher price. The second effect would, of course, not occur if the downstream technology does not permit substitution of x for y or if the price of y were to rise in the same proportion as the price of x. The first effect is responsible for a deadweight loss triangle under the demand curve for downstream product, the second, for a deadweight loss rectangle that shows the socially excessive cost from socially inefficient factor proportions as the downstream firm minimizes its input cost based on the price ratio s/w rather than the lower social marginal cost ratio r/w.

Unregulated Upstream Monopoly

Because of downstream marginal cost pricing

(19) $p(q) = \lambda(w,s) \equiv wa(s/w) + sb(s/w)$.

Price w is unchanged at its marginal costs but the price s of y is now set by the upstream monopolist maximizing profit

(20) $\phi = (s-r)bq$.

As s is raised by the monopoly, downstream competitive firms substitute against the input y, if technology allows, lowering b, raising

a. As long as substitution against y is not complete so that b is positive, increases in b shift marginal cost λ upward. That is

$$(21) \quad d\lambda/ds = a' + (s/w)b' + b = b > 0.$$

$$(22) \quad dq/ds = (d\lambda/ds)/p' = b/p'.$$

If substitution against y is total, $d\lambda/ds = b = 0$ and $dq/ds = 0$.

The rate of change in upstream monopoly profit in response to change in price s is

$$(23) \quad d\phi/ds = (s-r)(bdq/ds + qdb/ds + bq)$$

As input proportions adjust to price s for least cost and final product price increases as it must, one obtains, substituting for dq/ds from (22),

$$(24) \quad d\phi/ds = bq \left\{ 1 - \frac{s-r}{s} \left[\frac{bs}{p} \quad \frac{-p}{p'q} + \frac{-db}{ds} \cdot \frac{s}{b}\right] \right\}$$

or

$$(25) \quad d\phi/ds = bq \left\{ 1 - [(s-r)/s][\beta E + e] \right\},$$

where: bq is the quantity of y sold; $(s - r)/s$ is the profit per unit of y expressed as percentage of the selling price; $bs/p=\beta$ is the "relative share" of the input y, also equal to $1 - \alpha$, where $\alpha = wa/p$ is the relative share of x; E is the price elasticity of final demand; and $e = -(db/ds)(s/b)$ is the price elasticity of demand for input y, final output q held constant--of the output-compensated demand for y. It can be shown that $e = \alpha\sigma$, where σ is the elasticity of substitution.

For upstream monopoly profit to take on its maximum value, call it ϕ_1, $d\phi/ds = 0$. This is the first-order condition for the interior maximum. It determines the price s_1 for the quantity y_1, the input coefficients a_1 and b_1, and the downstream price and quantity p_1 and q_1. The subscript "1" indicates equilibrium values when production is not integrated--

upstream monopoly maxmizing profit, downstream pricetaking firms combining inputs using prices s, w to minimize costs, product price equalling marginal cost and market for product cleared. The condition is equivalent to the requirement that (derived) marginal revenue, $MR = s(1 - 1/\eta)$, marginal to the mutatis mutandis derived demand curve (AR), final product sold at marginal cost price, is equal to marginal cost, $MC = r$, of y. This follows at once from the fact that $\beta E + e = \beta E + \alpha\sigma$ is the Hicksian price elasticity η of derived demand for y.

The relationship between a downstream marginal cost price $p_1 = \lambda(w,s_1) = \lambda_1$, where s_1 is determined by unregulated upstream monopoly maximizing profit, and unregulated integrated

monopoly price p_o given in (11) has been analyzed elsewhere. The price p_1 may be higher or lower than p_o, depending on whether, for price s^* of y, defined by $p_o = \lambda(w,s^*) = \lambda^*$, (derived) marginal revenue for the seller of y is higher or lower than marginal cost r. Sufficient, but not necessary, for $s_1 > s^*$, $\lambda_1 > \lambda^*$ and, therefore $p_1 > p_o$ is an elasticity of derived demand: $\eta^* = \alpha^*\sigma^* + \beta^*E_o < 1$. As η^* is a weighted mean of σ and E_o, and $E_o > 1$, it is required that $0 < \sigma^* < 1$.

If $p_1 > p_o$ then the area of the "triangle" $\frac{1}{2}(p_1 - \lambda_o)(q_1 - \bar{q})$ is by itself greater than $\frac{1}{2}(p_o - \lambda_o)(q_o - \bar{q})$. And, because of input misallocation, there is here the additional deadweight loss rectangle $[wa(s_1/w) + rb(s_1/w) - \lambda_o]q_1 > 0$. If $p_1 < p_o$ then the area of the "triangle" is smaller. The combined areas of the triangle and the rectangle may be larger or smaller than the deadweight loss triangle for unintegrated monopoly. Which of the two areas is larger will depend on the shape of the downstream demand function and production function.

We conclude that allowing vertically integrated monopoly to exercise monopoly power may lead to better resource allocation than downstream marginal cost pricing and unregulated upstream monopoly. Indeed, integrated monopoly must lead to superior resource allocation when the derived demand for upstream ouput is inelastic at price s^*.

Upstream Markup Regulation

As unregulated upstream monopoly operates on the elastic portion of its demand curve, a restriction on upstream markup, combined with downstream marginal cost pricing, will cause a reduction in the upstream price, in the downstream marginal cost and, therefore, in the downstream marginal cost price. If the allowed markup, say n, is sufficiently high and/or derived demand elasticity $\eta > 1$, cost padding is not profitable. Assuming the constraint effective, deadweight loss is reduced because $s = nr < s_1$ implies downstream price $p = \lambda < \lambda_1$, and so the downstream loss triangle shrinks. In addition, social unit cost $wa(s/w) + rb(s/w)$ is lowered as upstream price reduction induces socially desirable substitution of y for x; but the increased output may cause the area of the deadweight loss rectangle to increase or decrease depending on elasticities. (Figure 4).

Consider now the situation where $\eta < 1$ for $nr = s$. As demonstrated earlier such inelastic demand makes cost padding profitable. The maximum profit maximizing upstream price is then $s_v = nr' > nr$ given by $\eta = 1$, and $(r' - r)$ is the padding--costs that need to be found by the upstream firm to justify price s_v to the regulators. The regulatory process thus induces deadweight losses. Downstream price is lowered from p_1 to $p_v = \lambda(w, s_v)$, social unit costs have been lowered from $wa(s_1/w) + rb(s_1/w)$ to $wa(s_v/w) + rb(s_v/w)$ while new social losses $(r' - r)b(s_v/w)q_v$ are incurred as profit income is frittered away on socially unnecessary costs that do not bring social benefits.

How do the results of conrolling the upstream markup compare with those of unregulated integrated monopoly? The answer is obvious if $\eta^* < 1$. Derived demand is then inelastic at price s* for which the downstream marginal cost price is as high as the integrated monopoly price p_o. Thus, $s_v > s^*$ and p, the downstream marginal cost price $p_1 > p \geq p_v > p_o$. With upstream regulation, or without, deadweight losses are greater than they would be for unregulated integrated monopoly. Final product price is higher, quantity is lower; final output is produced at higher social cost; and to top it off, there may be cost padding by the upstream firm if allowed markup n is low.

If $\eta^* > 1$, matters are uncertain. Unregulated upstream monopoly price s_1 may, as we saw, result in downstream marginal cost $\lambda(w,s_1) \gtrless p_o$, depending on whether at price s* upstream marginal cost r is or is not less than upstream monopolist's derived marginal revenue. Thus, an upstream markup restriction can, but need not, lower upstream price to a level $s = nr < s^*$. And as long as $\eta > 1$, there is no cost padding. Further, the downstream marginal cost price, equal to $\lambda(w,nr)$, may be lower than an unregulated price, p_o, of integrated monopoly. But even such a lower downstream price does not insure that deadweight losses are lower. Though the deadweight loss triangle $\frac{1}{2}(p - \lambda_o)(\bar{q} - q)$ is smaller, deadweight losses from input price distortions remain when $b(r/w) > b(nr/w)$, $a(r/w) < a(nr/w)$. These are precisely equal to the difference between profits $[\lambda(w,nr) - wa(r/w) - rb(r/w)]q(p)$, earned by integrated monopoly charging a price $p = \lambda(w,nr)$ and profits $[nrb(nr/w) - rb(nr/w)]q(p)$ earned by markup regulated upstream monopoly charging price $s = nr$. Such "wasted" profits maybe larger than the reduction in the area of the triangle. If they are, unregulated integrated monopoly is the superior alternative; if they are not regulated upstream monopoly with downstream marginal cost pricing is preferable.

As $n \to 1$ and $\eta > 1$ for all $s \geq r$, upstream markup regulation together with downstream marginal cost pricing leads to a Pareto Optimum with zero deadweight losses, just as markup controlled integrated monopoly was seen to lead to Pareto Optimality as $m \to 1$ provided $E > 1$ for $p \geq \lambda_o$.

Downstream and Upstream Markup Regulation

Suppose now that downstream price p is effectively regulated by the markup constraint $p = m\lambda(w,s)$ at the at the same time that upstream price s is effectively regulated by markup constraint $s = nr$.

If in addition final demand elasticity E and derived demand elasticty η are strictly greater than unity for any allowed markup percentages, so that all cost padding is ruled out, then for given m, $p = m\lambda(w,nr)$ approaches $p = m\lambda_o$ as $n \to 1$. This is, of course, the configuration for markup regulated integrated monopoly. Further, for given n, $p = \lambda(w,nr)$ as $m \to 1$. This is in the configuration for downstream marginal cost pricing combined with markup regulated upstream monopoly that we have just now examined. As both n and m together approach unity, a Pareto Optimal solution is approached.

In contrast, if as downstream price p is lowered to satisfy a markup requirement $E = 1$ for $p_u > m\lambda_o$, then this is surely the effective lower limit to markup controlled downstream price. Likewise, if upstream price s is lowered to satisfy a markup requirement, $s_v > nr$ for $\eta = 1$, then s_v is a lower limit to the markup controlled upstream monopoly price.

Which method of control can achieve the lower price for final product? If $p_u < \lambda(w,s_v)$, then $p_u < m\lambda(w,s_v)$. Control of downstream markup cannot attain a price as low as p_u. For $m \to 1$, downstream price approaches $p_v = \lambda(w,s_v)$. If $p_u > \lambda(w,s_v)$ though downstream costs are lowered to a level less than p_u, downstream price cannot be made less than p_u by controlling downstream markup. Thus, if $p_v > p_u$, the lower limit to product price is p_v; if $p_v < p_u$, the lower limit is p_u.

As $\eta = \alpha E + \beta\sigma$, it follows that $p_v > p_u$, $0 < \sigma < 1$; that $p_v < p_u$ implies $\sigma > 1$. One may note that the Cobb-Douglas production function (CES, $\sigma = 1$), $\eta = \alpha E + \beta$. Therefore, $p_v = p_u$.

This does not seem to settle the matter as to which mix of the regulatory parameters minimizes deadweight loss. Where there are lower limits to upstream or downstream prices, because of demand inelasticity, it is clear that the regulatory authority can, in principal, avoid any deadweight losses from cost padding by setting margins not less than $m = p_u/\lambda_o$ and $n = s_v/r$. Furthermore, for a given downstream price objective p that can be achieved with mixtures of markups given by $p = m\lambda(w, nr)$, it is better to reduce n and raise m. This keeps the loss triangle unchanged, but shrinks the loss rectangle. The trade-off is $-(dm/dn)(n/m) = br/\lambda(w,nr) = \beta$. It remains to be investigated, however, how least deadweight loss configurations are related to the Ramsey inverse elasticity rule.

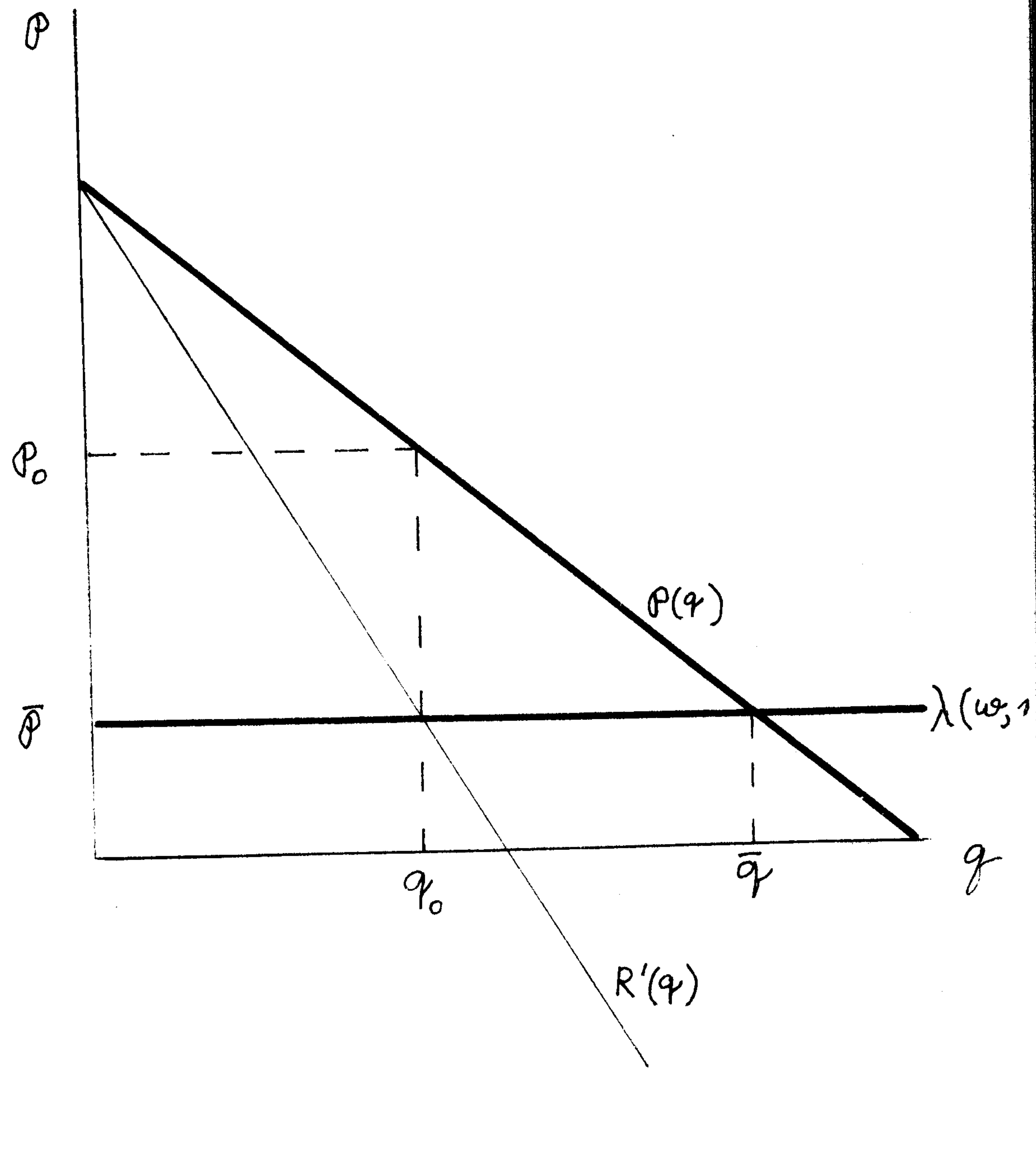
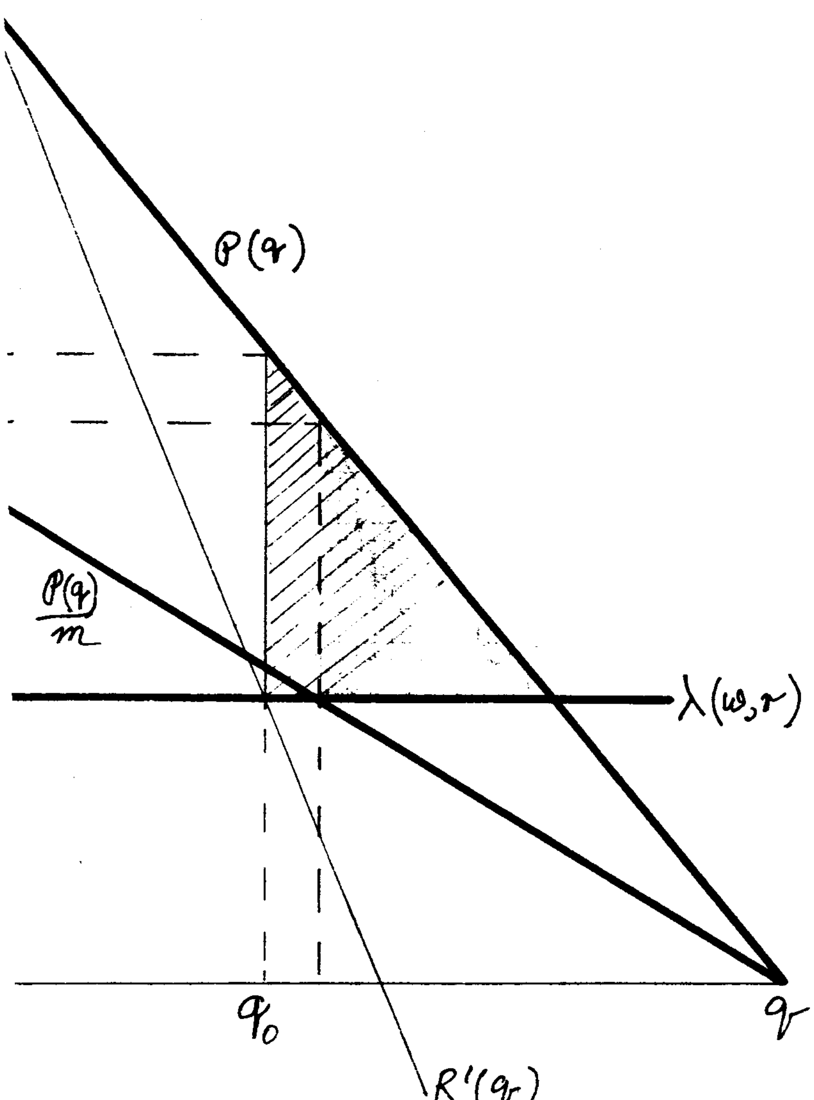


Figure 1



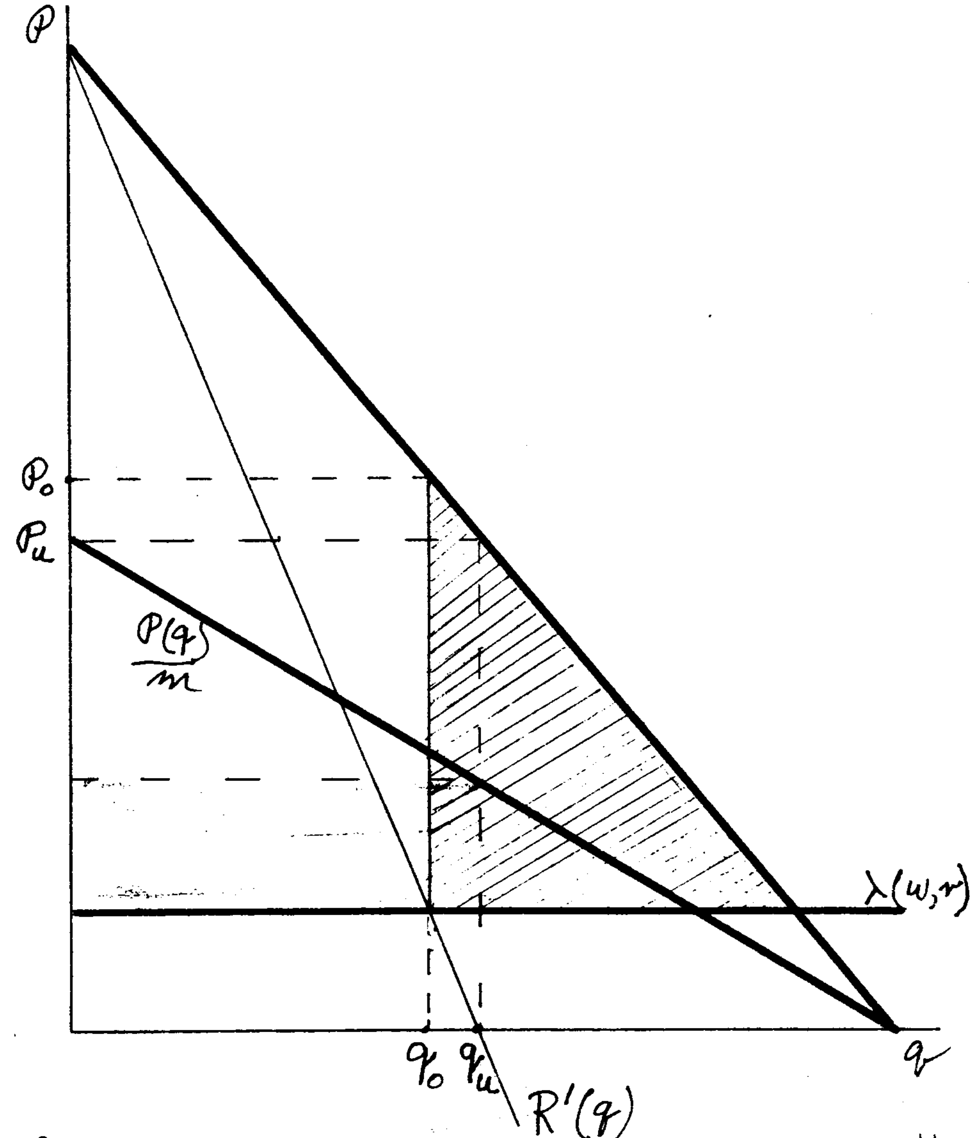


Figure 2

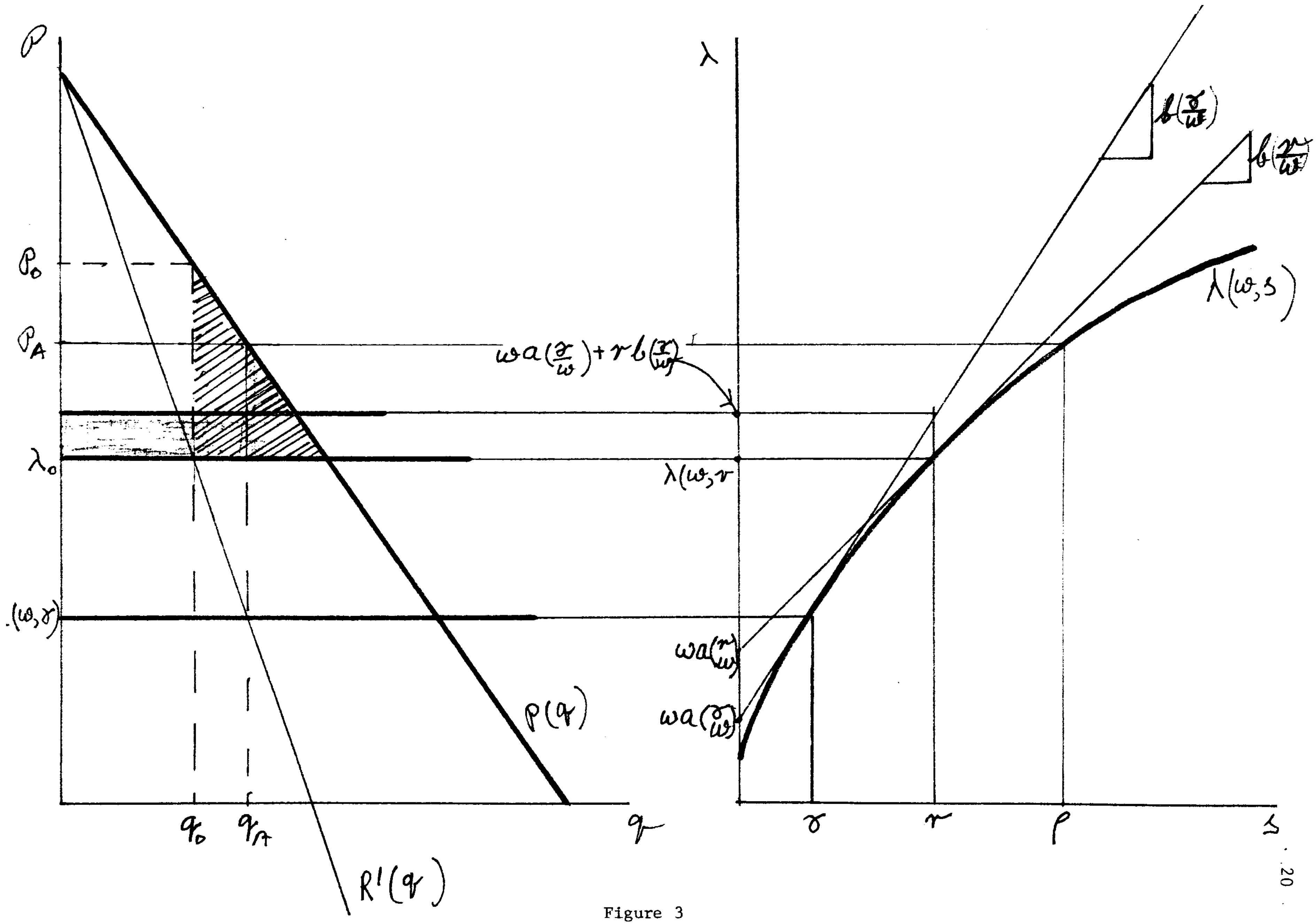

p
p₀
p_A
λ₀
(ω,γ)
q₀
q_A
q
R'(q)
p(q)
λ
b(γ/ω)
b(r/ω)
λ(ω,s)
ωa(γ/ω)+rb(γ/ω)
λ(ω,r
ωa(r/ω)
ωa(γ/ω)
γ
r
ρ
s

Figure 3

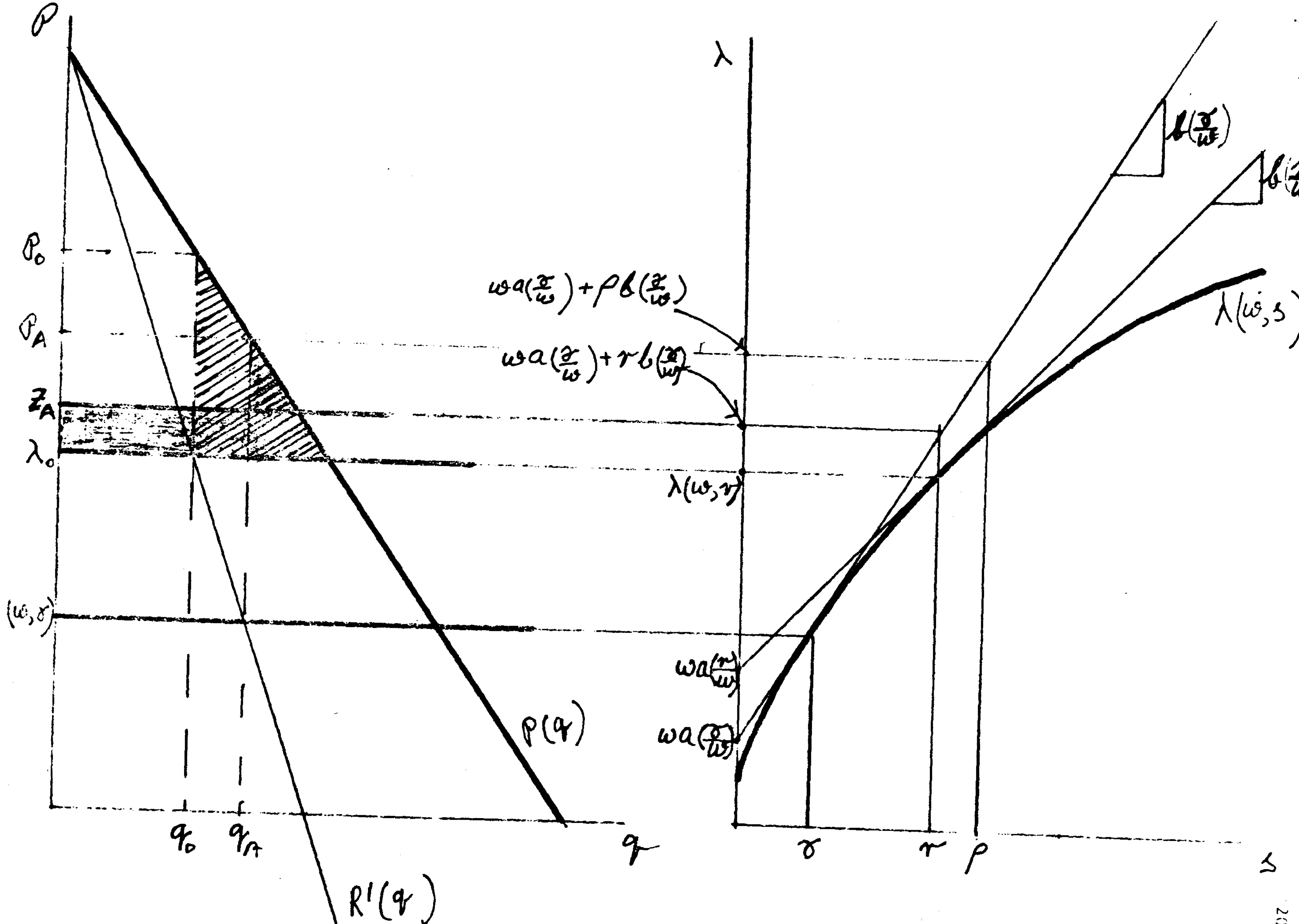


Figure 3

CORRECTED

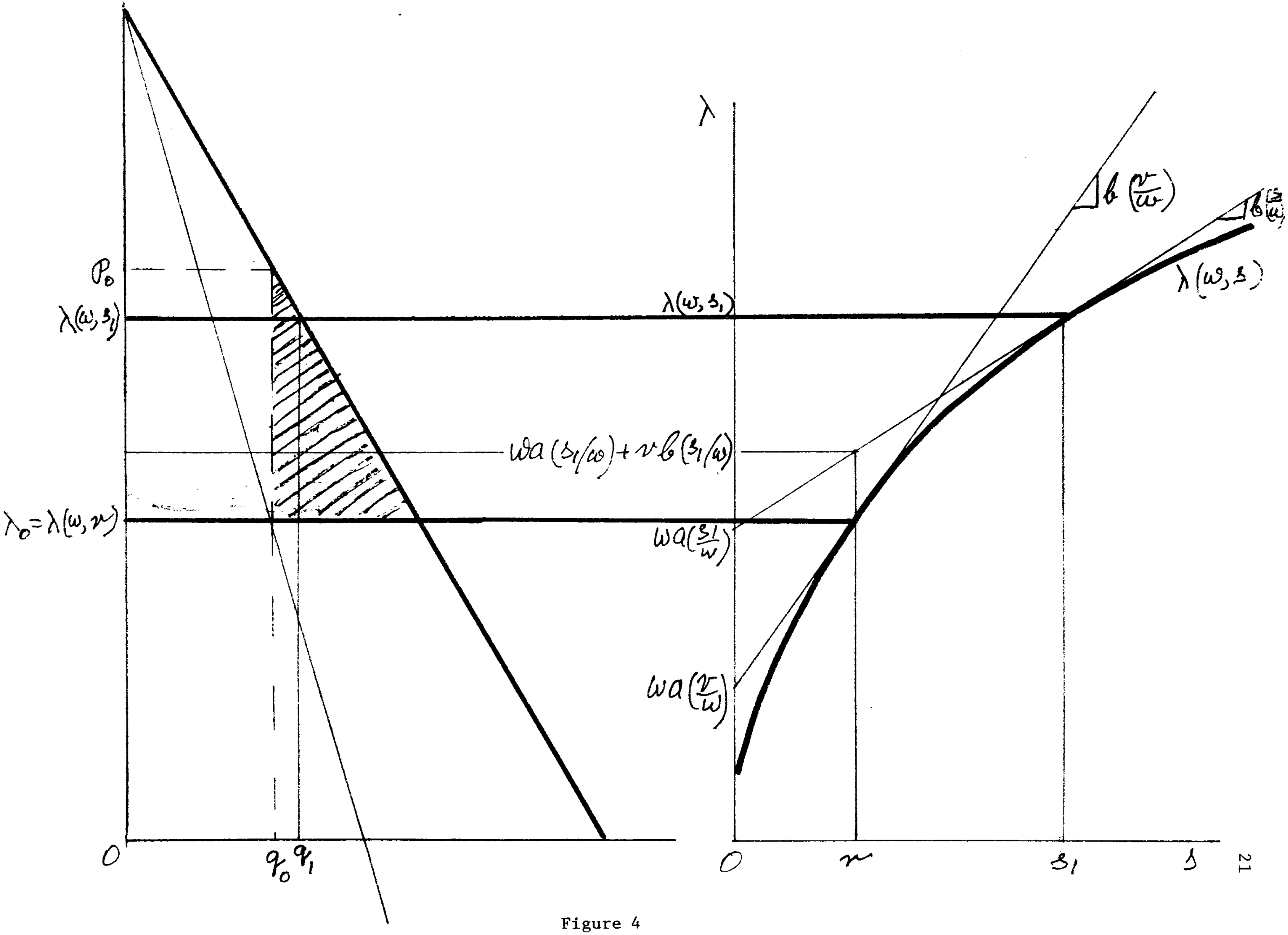

P_0
$\lambda(w, s_1)$
$\lambda_0 = \lambda(w, v)$
O
q_0 q_1
λ
$b\left(\frac{v}{w}\right)$
$\lambda(w, s)$
$wa(s_1/w) + vb(s_1/w)$
$wa\left(\frac{s_1}{w}\right)$
$wa\left(\frac{v}{w}\right)$
v
s_1
s

Figure 4

EMPIRICAL EVALUATION OF CROSS-SUBSIDY TESTS FOR CANADIAN INTERREGIONAL TELECOMMUNICATIONS NETWORK

C. Autin, G. LeBlanc

1. Introduction

The central objective of this paper is the empirical evaluation of some cross-subsidy tests for various telecommunications services (mainly private lines and public message services) in Canada, using the National Planning and Policy Simulation Model (NPPS model).

Any economic problem can be studied from an efficiency as well as from an equity points of view. The question behind the first point is to know if a particular good or service is provided at the cheapest way and if it is not, is there any incentive or policy for reaching this goal? From an equity point of view, one of the problems is to define means (taxation, regulation, etc) that government and/or industry can use in order to ensure that a particular efficient state of the economy is reached. Needless to say, the central problem of cross-subsidization rests on equity grounds since one can define the cross-subsidy concept as follows. A group of economic agents (consumers or producers) obtain certain service(s) from a producing system, for a given period. If the revenues for that service do not cover the value of the corresponding units utilized in the producing system, then some other economic agents must bear the difference.

At the policy level, the problem of cross-subsidization has always be a subject of discussion. And, in the telecommunications industry, it has becoming more crucial, both in Canada and in the United States, given the new institutional arrangements in which the companies must now operate.

As formulated as previously, the cross-subsidization problem is relatively easy to understand. But, once ones becomes more specific, many difficulties are raised: for example, which definition of revenues and costs one must use. The problem is even more complex taking the institutional, economic and technical characteristics of the telecommunications industry into account. Even if problems behind those characteristics will be more fully discussed in the next section, in relation with the cross-subsidy problem, it can immediately be said that each of them contributed to certain difficulties in applying the marginal cost pricing approach. It will also becoming clear why we favor an approach based on game theory in order to evaluate cross-subsidies, if any, for the various services under study.

At a more operational level, every cross-subsidization test necessitates the computation of some revenues and some costs, and to be meaningful those computations must be made at a certain level a desaggregation. One is then forced to use a big machinery. The model we used in order to compute those items is the one developed many years ago by the Department of Communications in collaboration with other parties [(1)]. The description of the model, with its main properties, is the subject of a published paper [(2)]. However, in order that the present paper be self-contained to a certain extent, a brief description of this model is done in Section 3, stressing in particular the way the costs of the services are determined in view of applying the cross-subsidization tests. The costing concept retained for present computations is the incurred cost, both for average and incremental ones, based on the reproduction values.

Section 4 and 5 of the paper presents the theoretical cross-subsidy tests which will be applied and reports on the rempirical results for the various tests performed using the aforementioned model respectively. Two series of tests are performed. In the first series of simulation two tests were applied (stand alone cost test and incremental cost test) for

---

(1) This project has been financed by the Canadian Department of Communications. It is a tripartite effort of: Department of Communications, Sorès Inc. of Montréal and Laboratoire d'économétrie. Mr. G. Henter was the project manager. Professor T. Matuszewski contributed to the conceptualization of the model. Mr. J.-P. Schaak was responsible for the software. However the present authors are sole responsibles for the interpretation given here.

(2) Autin, C., LeBlanc, G. "A National Telecommunications Planning and Policy Simulation Model", in Models and Decision Making in National Economies, J.M.L. Janssen, L.F. Pau, A. Straszak (eds), North-Holland Publishing Company, 1979.

the following pairs of services: public message versus private line, short distance traffic versus long distance toil one, peak versus off-peak demand, regional versus adjacent versus nonadjacent plus U.S. traffics. Note that all simulations are based on the present demands for these services and consequently assume the current usage of the network. The second series of simulation has been done for the public message versus private line services only. But instead of assuming the present demand for these services, like in the first serie of simulation, we formulated different hypotheses about the rate of growth of the demands and also about the allocation of the common costs, especially using the Shapley value as a way of splitting those costs. For these services and combining these hypotheses, four tests have been applied: incremental cost, stand alone cost, generalised incremental cost and finally generally stand alone cost tests.

Finally a last section evaluates the empirical findings, discussed the main weaknesses of our approach and suggests some extensions.

2. Theoretical Bases of Cross-Subsidization Problem

2.1 The Cross-Subsidization Problem in Telecommunications

Loosely speaking, by cross-subsidization, one means that somebody has to pay in full, or in part, for somebody else's consumption of a particular service. This aspect of who has to pay for whom is always present in our society. In the domain of telecommunications, this latter can be interpreted from many points of view. Among the most important is the possibility of financing a service out of profit generated by supplying some other services. Most recently this problem was posed in the following terms by the Telecommunications Committee, Canadian Transport Commission, in its decision of August 15, 1974. It concerns the expenditures in the Construction Program of Bell Canada for increasing the quality of non-urban services. The Committee said:

> "We fully realize, however, that such expenditures would require substantially more revenues from multi-party services to pay for them than the present rate structure would provide, and that such additional revenues would have to come from a new and higher rate structure for multi-party services or from increased rates for other services offered by Bell, or from both." (emphasis added)

The previous example refers to the question of who has to pay for the increased level of quality of the services. But, the problem of potential financing of the competitive services by the monopolistic ones is becoming a crucial question as the telecommunication industry now operates in at least a partial competitive environment. So, it is not surprising to find the following sentence in the Green Paper published by the Government of Canada:

> "If the carriers are to be permitted to offer unregulated services, one of the essential saveguards is that the public interest be taken fully into account in any circumstances where there is a possibility that the subscribers to one service may be subsidized by subscribers to another service, particularly if the latter are the general public."

As a final example, the Federal Communications Commission asserts in Docket no 18128 that one fundamental question in regulation is

> "Whether the rate levels for the (telecommunications) services will subject any person or class of persons to unjust or unreasonable preference or advantage to any person, class of persons or locality, or subject any person, class of persons or locality to any undue or unreasonable prejudice or disadvantage...".

The motive of such an interest comes to light when the F.C.C. contends that

> "The public interest is not generally served by cross-subsidization of any one class of services by any other class of services, or by cross-subsidization of one sub-class of services by any other sub-class within the same class of services."

The problem of cross-subsidization is then so constant preoccupation, from the regulator viewpoint as well as from an industry point of view. As formulated as previously, the cross-subsidization concept appears relatively easy to understand. In fact, the concept is very delicate to capture, and especially in the telecommunications industry, for at least three reasons:

i) There are several definitions of this concept and consequently several tests for evaluating the presence, if any, of cross-subsidies among, say, certain services; a review of some definition and tests is done in the following sub-section;

ii) even if specialists retain a particular definition of cross-subsidy, some difficulties appear at an operational level; for example must we use an embedded or a reproduction marginal cost concept. The precise definitions we employed are described in the next section and the section after the next one;

iii) the usual framework in which cross-subsidization is scrutinized is the marginalist welfare theory. However, the main technical charateristics of the telecommunications industry contribute to difficulties in applying the marginal cost pricing approach. For example, as long term average costs are decreasing, a tariff based on this approach will not recover all costs incurred in providing the services. But as this industry is regulated, it must recover all its costs, implying that certain services must be priced higher than its corresponding marginal cost. Also, the jointness of supply, meaning that it is cheaper to supply a particular service to a group of customers than to supply it separetely, has a consequence that many costs are common and then somewhere one of the marginal costs is meaningless. Finally, the fact that many indivisibilities are present in the industry implies that marginal cost cannot be defined in a unique manner. All these reasons explain why we favor a framework based on game theory concepts in view of studying the cross-subsidy problem. This is the subject of the following sub-section.

The cross-subsidization problem has many policy implications. In effect, if the regulated monopolist can apply a subsidy-free tariff structure, then there is no incentive for a potential competitor to go in the market (if it were allowed to do so). But, if it is found that some markets are financing some others, then many questions arise:

- can the monopolist react, in terms of both quantities and prices, to the entry of a competitor in such a way that the competitor will eventually go to bankruptcy (this problem has been studied among others by Panzar and Willig [12])?

- traditionally, the tariffs applied by the carriers have been based on a value of service approach rather then a cost of service approach. In other words, tariffs were based on the inverse elasticity rule. It is immediate, and it can easily be shown, that such a tariff structure is not subsidy-free. But, there is now a tendency to go to the cost of value approach, i.e. to tag each component of the network. Consequently, is the existence of entry barriers compatible or not with a tariff structure based on this principle?

Even if these policy questions, and many others, are not formally discussed in this paper, it is implicit that there are our main motivations for studying the problem of cross-subsidization in the canadian telecommunications industry.

2.2 The Game Theoretic Approach

2.2.1 Introduction

Common costs and economies of scale render difficult the problem of pricing commodities or services produced by a publicly owned or regulated enterprise. In these circumstances, regulators and policy makers are on the horns of the dilemma which arises between efficiency and equity. In fact, common costs and increasing returns call forth a conflict between welfare maximization and subsidization. Economic efficiency argues strongly against subsidization. However, in many cases, governments think equity is more important than efficiency and decide to introduce subsidization.

One question of equity that is raised is the following: does a certain price structure for a multiservice firm unduly favor the consumers of one service at the expense of the consumers of another service, i.e. do the prices result in cross-subsidization? The thorny problem is to define what we mean exactly by cross-subsidy. If those who receive the benefits of an economic process differ from those who bear the costs, there is subsidization. But to calculate the extent of cross-subsidy, we need a much more precise definition. Many authors have proposed different ones leading to several tests. In this section, we will review the more important of these tests.

2.2.2 One-service Tests

All the tests quoted in this paragraph are made on a cost-revenue causation basis and examine only one service at a time. Let us consider a firm providing n services, the set of which is denoted $N = \{1,2,\ldots,n\}$, with demands q_i and prices p_i. We define a cost function as following: if S is a subset of N, C(S) means the cost of supplying the services in S on a "stand-alone" basis. We make the assumption of economies of joint production, which can be stated:

$$C(S \cup T) \leq C(S) + C(T) \qquad (1)$$

for all subsets S, T of N with $S \cap T = \phi$. This means that supplying services S and T jointly costs no more than supplying these services separately.

Similarly, R(S) represents the revenues gained by the production of this subset of services. We assume the absence of cross-elasticities, that is we assume that quantities demanded are function of their own prices only. We can then write

$$R(S) = \sum_{i \in S} r_i = \sum_{i \in S} p_i \cdot q_i\ (p_i).$$

We suppose that the firm's profits must be zero or, equivalently, that revenues just cover total costs, including cost of capital, i.e.

$$R(N) - C(N) = \pi(N) = 0. \qquad (2)$$

Hazelwood [6] described a way for studying cross-subsidy: "Any... subscriber to the service should be able to obtain extra units of the service if he is willing to pay an amount equal to the cost of providing these units". This gave rise to the Incremental Cost Test (ICT), of which one interpretation can be expressed as follows: the firm's prices $(p_1,\ldots,p_n)$ are subsidy-free if and only if

$$R(i) \geq C(N) - C(N-i), \quad \text{for any } i \text{ in } N. \qquad (3)$$

This means that the revenues from supplying the service i must at least equal the added costs necessary to provide this service.

Another variant has been proposed by D. Gillette which is called the Stand-Alone Tests (SAT): a service of a multiservice enterprise subject to a profit-constraint does not yield a subsidy if the revenues from that service are no greater than the revenues required by a subsidiary firm supplying the same service and bearing the same profit constraint. This can be reformulated in the following terms:

$$R(i) \leq C(i), \quad \text{for all } i \text{ in } N. \tag{4}$$

This is a comparison between single revenues and stand-alone costs. Here, we have to point out that C(i) means the cost incurred in supplying individually the service i while R(i) represents the revenues yielded by the service i in the whole coalition of services.

Zajac [14] propounded yet another approach in saying that "no customer group should pay higher prices than it would pay by itself". The difficulty of this approach is in the meaning of the expression "by itself". Zajac proposed two scenarios. The Scenario 1 Test (SIT) stated that "the price for each service does not exceed the service's price if it were the only service offered". He found these minimal prices by setting the prices of all the services but one so high as completely to choke off their demands; it remains just one demand whose minimal price can be easily determined. In the Scenario 2 Test (S2T), no service disappears since it is taken over by an alternative supplier. The existence of many suppliers who act independently makes that scenario an inefficient one.

The last definition of cross-subsidization, we will recall, is that given by Faulhaber [5]: "If the provision of any commodity (or group of comodities) by a multicommodity enterprise subject to a profit constraint leads to prices for the other commodities no higher than they would pay by themselves, then the price structure is subsidy-free". We can conclude from that definition that subsidy-free prices permit to affirm that the supply of each commodity by the firm is "Pareto superior" to non-provision.

All the above definitions or tests look at each service individually; therefore, cross-subsidization involving groups of services may not be detected by these tests. Such tests would be sufficient if a single service were responsible for all costs or if common costs were joint among all services. But when costs are common to a proper subset of the whole of services, we have to test that subset for cross-subsidy.

2.2.3 Generalized Tests

In this section, we will extend some of the preceding tests so that they take into consideration the several groups of services. Loehman and Whinston [10], [11], in defining incremental costs in the case of joint production, have insisted on the fact that the incremental cost of a service depends upon what it is incremental on. Intuitively, we admit that the incremental cost of a service provided alone may be quite different from the incremental cost of the same service if provided along with other services.

The ICT can be generalized as follows: the firm's prices are subsidy-free if and only if

$$R(S) \geq C(N) - C(N-S), \quad \text{for all subsets } S \text{ of } N. \qquad (5)$$

The extended form of the SAT is: the firm's prices are subsidy-free if and only if

$$R(S) \leq C(S), \quad \text{for all subsets } S \text{ of } N \qquad (6)$$

It is easy to show that the generalized ICT (GICT) is identical in meaning to the generalized SAT (GSAT), by remembering we assumed that the profit constraints must always be satisfied. In fact, if we substract equation 6 from equation 2 and with (6), we have

$$R(N) - R(N-S) \geq C(N) - C(N-S)$$

that is

$$R(N) - R(N-S) = R(S) \geq C(N) - C(N-S), \quad \text{and inversely.}$$

This signifies that in the case of a zero-profit constraint, GICT is equivalent to GSAT, i.e. the set of prices satisfying equation 5 (GICT) is identical with the set of prices satisfying equation 6 (GSAT). It is an interesting result since ICT was not equivalent to SAT (Faulhaber [5] has worked out an example showing this).

In this first section, we mentioned the interplay between subsidization and restricted entry into the market, cross-subsidy is only possible in a market whose entry is constrained. In a free-entry market, subsidization would be prevented by the threat of a new competitor's entry which could underprice. Faulhaber [4] thus proposed to give a more precise meaning to Zajac's proposition that "no customer group should pay higher prices than it would pay by itself". The new modified version would be: "no customer group should pay higher prices than it would pay if there were free entry into the market".

2.2.4 The firm as a cooperative game

Many authors introduced a game theoretic approach for analyzing some economic problems. Apparently, the first author who utilized such an approach in the study of cross-subsidization was Faulhaber [4], [5]. The theory of n-person cooperative games yields an easy recognizable structure for the "game" of cross-subsidy.

Let us assume the same hypotheses as at the beginning of section 2.2 For given demand levels $q_1,\ldots,q_n$, we can view the consumer groups of the services $N = \{1,\ldots,n\}$ as the "players"; the cost function $C(\cdot)$ is the characteristic function, corresponding to the "value" of the game; the vector of revenues $(p_i q_i)$ is the "payoff" vector; finally, the players can form the "coalition" N whose cost is C(N) or many "subcoalitions" S, where $S \subset N$, with costs C(S).

Expression 7, the assumption of economies of joint production,

$$C(S \cup T) \leq C(S) + C(T), \quad \text{for } S, T \subset N, \ S \cap T = \phi \tag{7}$$

which insures that there is a cost incentive toward cooperation, is the condition of subadditivity.

Equation 8, the zero-profit constraint,

$$R(N) - C(N) = \Pi(N) = 0 \tag{8}$$

represents the condition in game theory that the whole value of the game must be shared among the players by way of the payoff vector.

The set of "imputations" of a game is the set of revenues satisfying the zero-profit constraint and

$$R(i) \leq C(i), \quad \text{for all } i \text{ in } N. \tag{9}$$

These are the revenues which cover the total costs and for which each consumer group pays no more than its stand-alone cost. We note that the set of imputations is nothing else than the set of revenues passing the stand-alone test.

The "core" of a game is those imputations for which

$$R(S) \leq C(S), \quad \text{for all } S \subset N. \tag{10}$$

It is the set of revenues covering the total costs and for which no coalition of consumers can pay more than the stand-alone cost of that coalition. Here too we note that the core is nothing else than the set of revenues passing the generalized SAT, that is the generalized ICT.

The reference to the theory of games allows us to apply the results of that theory to the cross-subsidy problem. It is well known from the theory of n-person cooperative games that any game fulfilling the subadditivity condition has a non-empty set of imputations. This implies that, as long as we assume the existence of economies of joint production, there is at least one vector of revenues that passes SAT. This implication is interesting because the hypothesis necessary is not really severe since it corresponds to the notion of a natural monopoly. However, a serious problem crops up with the fact that not every game possesses a core. We will come back to this difficulty later on.

Faulhaber [5] has proved the following very interesting theorem: if we make the assumptions that:

cross-elasticities are zero, i.e. $(\delta q_i/\delta p_j)(p_j/q_i) = 0 \quad i \neq j,$ (11)

the prices are not "perverse", i.e. $\delta\Pi(S)/\delta p_i > 0$

for all i in S (12)

then the core of the preceding game is identical to the set of subsidy-free prices. The theorem signifies that if revenues are in the core of the game, i.e. pass GICT, and if conditions 11 and 12 are satisfied, then no consumer coalition could obtain lower prices. The global coalition N can block all other subcoalitions $S \subset N$. The usefullness of this result as a practical guide stems from the reduction of all cross-subsidization tests to a price test. In the zero-cross-elasticity case, to determine whether prices are subsidy-free or not, we need only calculate revenues and costs of the hypothetical coalition based on the initial fixed price structure and demand levels. There is no need for demand elasticities.

Nevertheless, when cross-elasticities are non-zero, we have to define a more complex game in which the value of the game is now profit and the price vector is the new payoff vector. The profits must be constrained to be non-negative. The core of the new game (the "price" game) is defined as follows: the price vector $p = (p_1,\ldots,p_n)$ belongs to the core if and only if

a) $\Pi(N, p) = 0$,

b) there does not exist a subset $S = \{i_1,\ldots,i_S\}$ and a price vector $p^* = (p^*_{i_1},\ldots,p^*_{i_S})$ such that

1) $\Pi(S, p^*) \geq 0$ for any feasible choice of p^*_k, $k \notin S$,

2) $p^*_j < p_j$ for all $j \in S$

For any price vector in the core of this game, no incentive exists to form other coalition than N to get lower prices. Faulhaber called such prices stable and gave the following interpretation: if the price vector of a regulated firm is stable, then allowing free entry would not induce any consumer group to desert the global coalition, i.e. the prices must be subsidy-free. He propounded another test for cross-subsidization, called the Stability Test (ST): a regulated firm's price vector is subsidy-free if and only if it belongs to the core of the preceding price game.

Under the hypotheses of zero-cross-elasticities, all the above-mentioned tests (GICT, GSAT, SIT, ST) are equivalent. In presence of non-zero cross-elasticities however, these tests are no longer equivalent and the relative stringency of the three relevant tests depends upon the sign of the cross-partial derivatives of the demand relationships, i.e. if the services are substitutes or complements.

2.2.5 Imputation of Incremental Cost

Several methods exist, all arbitrary, to separate common costs. Loehman and Whinston [10], [11] have deduced, from a set of axioms, a meaningful formula of social incremental cost which provides a way of allocating joint costs.

They postulate a service provided from a common facility and distributed to a given set of users. Each user is assumed to face fixed demand. The axioms which they are asked to consider for financing the facility are the following:

1) Charges for use of the facility cover total costs.

2) Charges imputed to one user must be based only on the incremental costs caused by that user and not on the incremental costs of other users.

3) The charge is independent of the ordering of users, i.e. users with equal demands cause the same incremental costs and hence will pay the same charge.

4) The charge is homogeneous of degree one in the incremental costs, i.e. if all prices increase by a multiple, then the charge will also increase by the same multiple.

These axioms are intended to exhibit some equity in supplying a public service and illustrate an approach for making welfare choices without reference to a welfare function.

From these axioms and assuming that n users with fixed positive demands $K_1,\ldots,K_n$ agree to use a collective facility, Loehman and Whinston demonstrate that individual charges for use of the facility are given by the following formula:

$$F(i) = \sum_{\substack{G \subset N \\ i \in G}} \frac{(n-g)!\,(g-1)!}{n!} \{C(G) - C(G-i)\}$$

Where G are subsets of size g of the whole group of users N, and C(G) is the minimum cost in fulfilling demands K_G for the subgroup G. This result signifies that if the supposed users accept the fairness of these axioms and take them as a constitution, they must then also accept the cost-allocation formula F(i).

This pricing system has thus a touch of equity and efficiency since it imposes on each user the need to pay the social incremental costs due to his demands and covers all the costs of supplying a public service. It is also worth noting that the cost-allocation formula derived from the four axioms is the only one that can fulfill all these axioms if we further assume that the function F(i) is twice continuously differentiable for each i (see [11]).

Under the assumptions of perfect competition, the incremental-cost formula shares the costs in the same way as marginal-cost pricing does. However, in the presence of decreasing costs, unlike marginal-cost pricing, the incremental-cost scheme covers the full costs. Moreover, the existence of decreasing costs implies incentives to use and finance a collective facility since $F(i) \leq C(i)$, i.e. a person's charge in a joint facility is no greater than the charge if he had to provide the service by himself.

Those acquainted with game theory will have noticed that the incremental-cost formula is identical to the Shapley value of a game. In fact, Loehman and Whinston [11] have pointed out the parallel between the set of axioms taken by Shapley to derive his formula and the four axioms they used to produce their own scheme. There is a link between the incremental-cost formula and the game theoretic approach which is worth mentioning. If the Shapley value were in the core of the "price" game defined as previously, this would imply that using the incremental-cost scheme for allocating costs, one could thus obtain subsidy-free prices. Unfortunately, such a result is not always true, i.e. the Shapley value does not need to lie in the core. It can be shown that for important class of games, the so-called convex game, first, the core is always non-empty, second it always contains the Shapley value (see Shapley [13]). However, the incremental-cost formula represents a useful scheme for allocating costs in a fair manner whether it yields subsidy-free prices or not.

2.2.6 Game Theoric Determination of a Subsidy-Free Tariff Structure

Assume a multiservice firm offering $N = \{1,2,\ldots,n\}$ services and denote by S a subset (a coalition) of N. Denote also by C(S) the (minimal) cost of supplying the subset S. It will be assumed that the $C(\cdot)$ satisfies the following properties:

a) monotonicity: $T \subset S \rightarrow C(T) \leq C(S)$

b) sub-additivity: $C(S \cup T) \leq C(S) + C(T), \quad S, T \subset N; \quad S \cap T = \phi$

The interpretation of a) is straightforward. The hypothesis b) means that it does not cost more to provide S and T jointly than to provide them separately.

Denote by R(S) the revenues derived from S. Of course $R(\cdot)$ is additive, i.e.

$$R(S) = \sum_{i \in S} r_i$$

where r_i represents revenues derived from providing the service i. Finally, if we denote by $\Pi(\cdot) \triangleq R(\cdot) - C(\cdot)$, the profit function, it follows from the previous hypothesis about $R(\cdot)$ and $C(\cdot)$ that the profit function is super-additive:

$$\Pi(S \cup T) \geq \Pi(S) + \Pi(T), \quad S, \quad T \subset N; \quad S \cap T \neq \phi$$

The main idea behind the determination of subsidy-free tariff structure is that the tariffs must be such that the gains coming from the economy of scale of providing <u>all</u> services at the same time be not destroyed. To achieve this, one must find some imputations $u = (u_1, \ldots, u_n)$ which are in the <u>core</u> of the so-called game, where the core is defined in the following manner

$$\text{Core } (N, \Pi) = \{u \geq 0 \mid \sum_{i \in S} u_i \geq \Pi(S), \sum_{i=1}^{n} u_i = \Pi(N), S \subset N\}$$

So the core is defined as the set of imputations which satisfy the following two constraints: first, the imputation given to any coalition is not less than the profit the coalition can obtain by its own actions, second, the imputation for all the services must add to the maximal profit which the coalition of all services can win. If the core is not empty, it can be obtained by resolving the following standard linear programming problem:

$$\text{Min} \sum_{i=1}^{n} u_i$$

subject to

$$\sum_{i \in S} u_i \geq \Pi(S), \quad \forall S \subset N$$

$$u_i \geqq 0 \qquad i \in \{1, 2, \ldots, n\}$$

Finally, knowing that $\Pi(S) = R(S) - C(S)$, one can then rewrite the first constraints as follows:

$$\sum_{i \in S} u_i \geq R(S) - C(S)$$

Now, if one assumes that the demand for each service is very inelastic to its respective price and that the cross-elasticities of the services are zero, one can redefine the core as follows

$$\text{Core } (N, C) = \{f_i \mid \sum_{i\,S} f_i \le C(S), \quad f_i \le r_i, \quad \sum_{i=1}^{n} f_i = C(N)\}$$

by defining $f_i \triangleq r_i - u_i$ and also $r_i \triangleq t_i q_i(t_i)$. It then follows that one can determine subsidy-free tariffs covering the total costs by setting

$$t_i' = f_i / q_i, \qquad i = 1,2,\ldots,n$$

because

$$\sum_{i \in S} t_i' q_i = \sum_{i \in S} f_i \leqq C(S), \quad S \subset N$$

and

$$\sum_{i=1}^{n} t_i q_i = \sum_{i=1}^{n} f_i = C(N).$$

Of course, it is not an easy work to empirically determine this kind of tariff structure, taking into account the lot of information required to apply the previous approach. Moreover, it is evident that the validity of this approach is weakened by the fact that some hypothesis needed are too strong and, in fact, are certainly not true for certain aspects of the telecommunications industry. Hence, we suggest that efforts be directed to improve the theoretical basis and to acquire simultaneously the necessary data.

3. Brief Review of the NPPS Model (version 1977)

3.1 Objective

The National Policy and Planning Simulation model is a very disaggregated model, closed to the long distance network observed in Canada. It was built to evaluate the financial and economic impact, on each carrier, of scenarios mixing variations on technical, accounting and/or economic variables (see [2]) at the level of perception of managers. Among the main issues behind the modelizing of the TransCanadian network were the costing of various services, the effects of various settlement schemes for splitting the revenues and costs of the interregional activities and the differential effects of methods of accounting.

3.2 Logical Structure

The model is modular with four blocks (fig.3-1):

3.2.1 The Operating Block is inspired by the structure and the operating rules of the national telecommunication network. The data bases are as follows:

a) Traffic data base

i) For the non-switched traffic, point to point circuit requirements for a base period are given.Television and private lines are the only non switched services considered so far.

ii) For switched traffic, point to point offered traffic profiles are provided in Erlangs or C.C.S. for typical days and the traffic can be modulated along the 24 hours.

b) Switching network (S.N.) data base

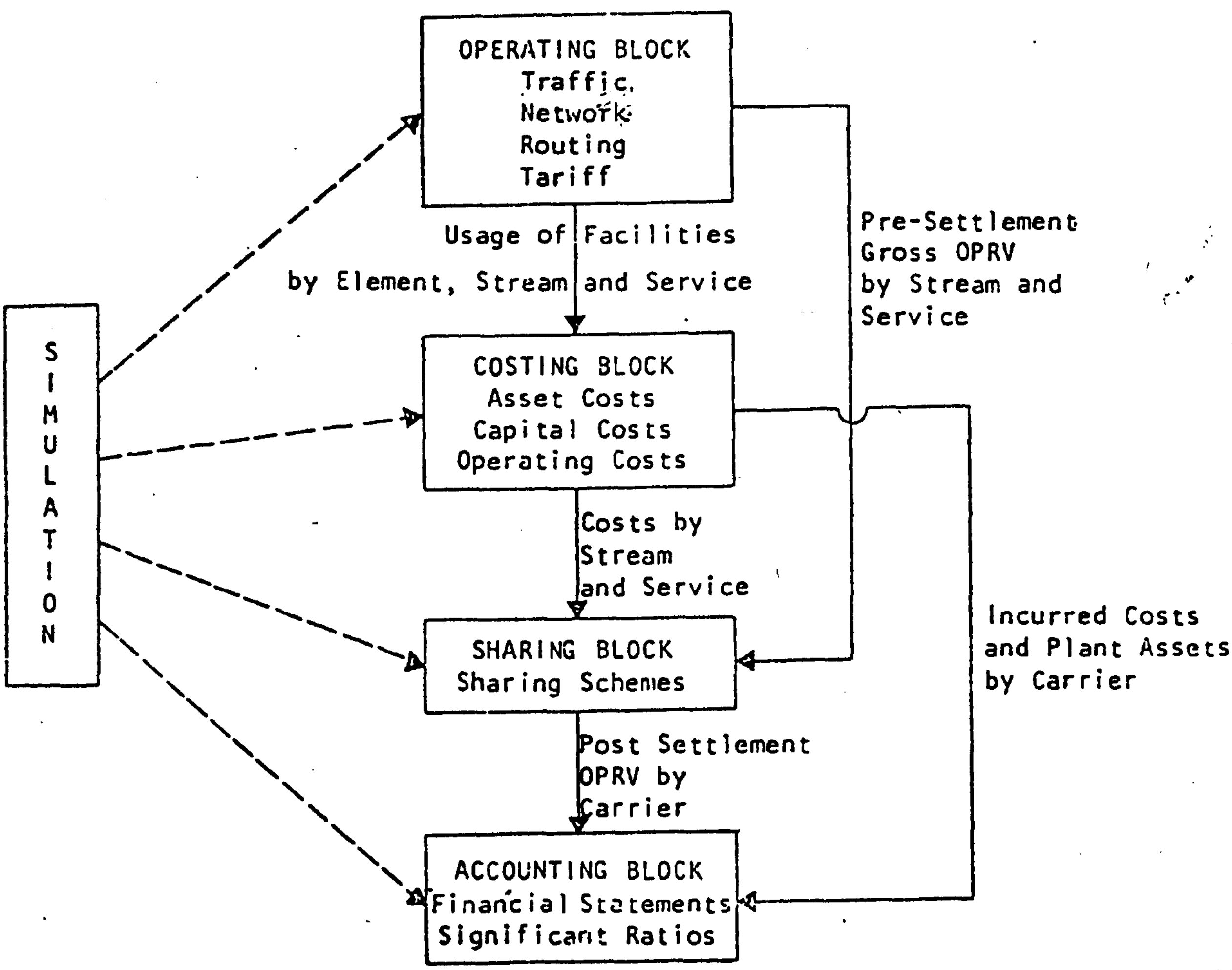
The S.N. is given with its configuration, its hierarchical tree and the rules of overflowing, its quality of service parameters (probability of lost on the ultimate trunk), the number of circuits on each links, the lo-

A NATIONAL TELECOMMUNICATIONS PLANNING

AND POLICY SIMULATION MODEL

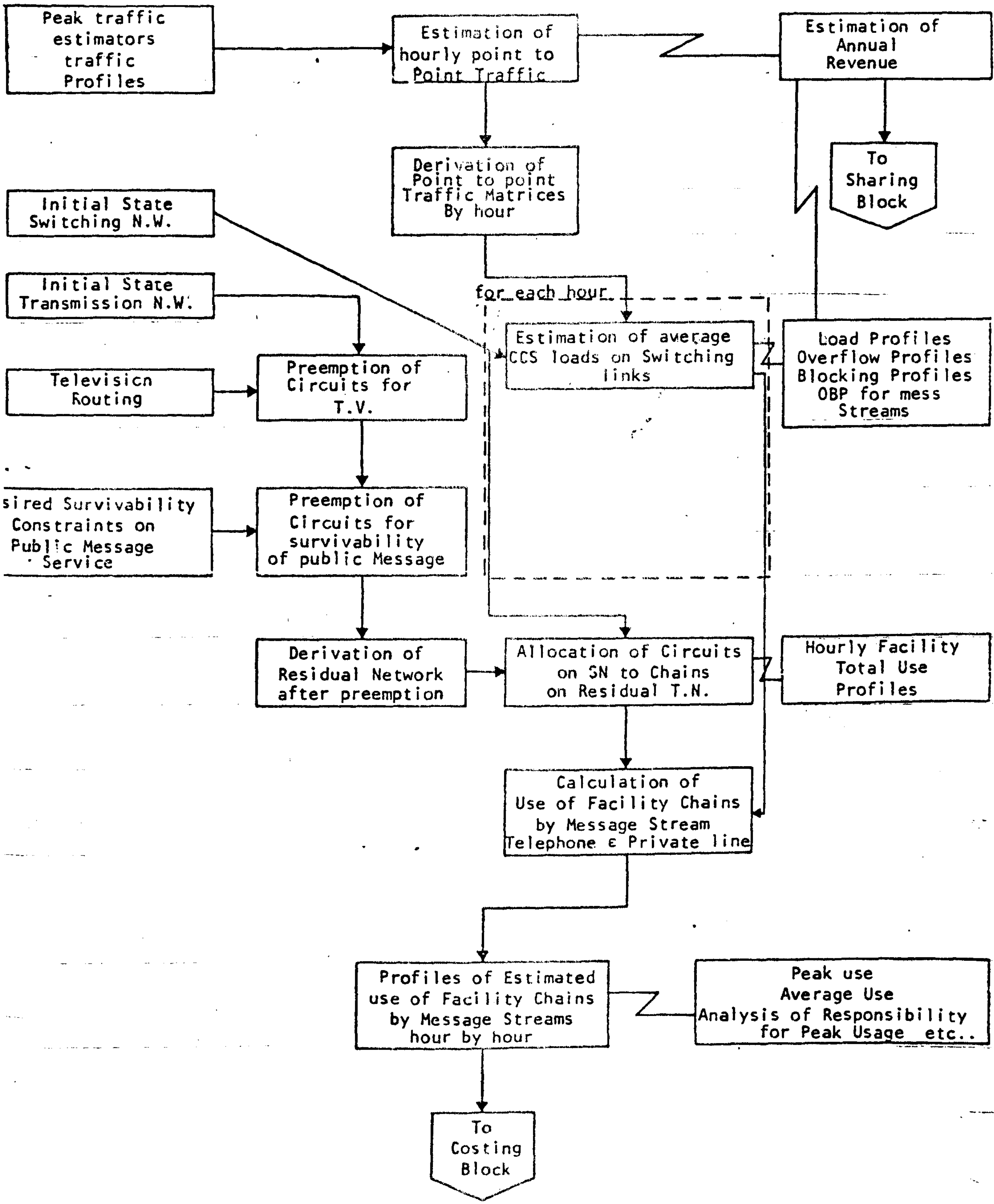
CONCEPTUAL STRUCTURE

FIGURE 3-1



OPERATING BLOCK FLOWCHART

FIGURE 3-2



cation of the switching machines, an ownership tag for each facility. The S.N. model comprises 96 nodes and 373 links.

c) Transmission network (T.N.) data base

The T.N. is given with its configuration, the link capacities (actual and ultimate after completing certain facilities), the ownership tags. The T.N. comprises 219 transmission nodes and 239 transmission links.

d) Tariffs data base is derived from the Trans Canada Telecommunication System (1977).

The 4 algorithms in the Operating block are as follows:

a) The traffic carried on the S.N. is estimated by an algorithm. The expected traffic on a given link is decomposed according to the origin-destination streams by using the Erlang formula for overflow probabilities and Poisson's formula for loss probabilities on the ultimate links. A careful sequence of choices of links ascending and descending the node hierarchy, permits the building-up of loads from different streams on each link for any typical desired time period.

b) If the expansion of the S.N. is needed, an algorithm computes the dimensioning of high usage and ultimate trunks for the new total traffic; then by comparing with the existing circuits, it derives the necessary addition in terms of circuit requirements and switching facilities. The dimensioning utilizes a version of the economic C.C.S. rule which asserts that the number of circuits to handle a given volume of switched traffic should be allocated between the direct and alternate route so that the ratio of the respective marginal costs equal the ratio of the corresponding marginal efficiency.

c) Given the circuit requirements for the adjacent nodes of the S.N. and the circuit requirements for the origin-destination pairs of non-switched traffic, a linear programming algorithm allocates those requirements to routes subject to capacity constraints on transmission facilities. It is worth mentionning that the routes are not enumerated a priori, but searched with the help of the dual variables associated with the capacity constraints at each iteration.

One of the allocating criteria used is the minimizing of the "cost" of the allocated circuits; the unit cost for each facility being derived from the embedded cost (average or incremental). It can be shown that this criteria is equivalent to maximizing the cost of the excess capacity under the hypothesis that there is enough capacity for the given requirements. Formally, the allocation model for the T.N. is:

$S = (S_1,\ldots,S_i,\ldots,S_m)$ the spare capacity vector for all transmission links;

$x_k = (x_{1k},\ldots,x_{jk},\ldots,x_{n_k k})$ the number of circuits carried on the j-th route for the k-th origin-destination pair ($k = 1,2,\ldots,L$);

$C = (C_1,C_i,\ldots,C_m)$ the weight (unit cost most of the times) vector corresponding to the spare capacity vector, generally derived from the asset valuation functions in the Costing Block;

$u = (u_1,\ldots,u_i,\ldots,u_m)$ the capacity of the links of the T.N.;

v_k: the circuit requirement for the k-th pair;

I_m: an identity matrix of order m;

A_k: a boolean matrix indicating the link-in-route membership for the k-th pair, built from the T.N. corrected for routing restriction (if any);

$e_k = (e_{1k},\ldots,e_{n_k k})$ a vector of 1.

So the model is:

$$\text{Max } z = cs$$

$$\text{Subject to: } I_m S + \sum A_k x_k = u$$

$$e_k x_k = v_k, \quad x_k \geq 0, \quad \text{all } k.$$

d) If an expansion-allocation for the T.N. is needed, the preceeding allocation model can be modified as follows:

$$\text{Min } z = d\Delta u$$

$$\text{Subject to: } -I_m \Delta_u + \sum A_k x_k + I_m s = u_o$$

$$e_k x_k = v_k + \Delta v_k$$

$$0 \leq \Delta u \leq \overline{\Delta u}, \quad x_k \geq 0, \quad \text{all } k$$

where Δu: the vector of expansion on the link;

$\overline{\Delta u}$: the vector of upper bounds on the components of Δu;

d: the incremental unit cost coefficient vector, generally derived from the asset valuation functions in the Costing Block;

Δv_k: the additions to the circuit requirements for pair k.

e) For estimating the usage of facilities by traffic streams, the estimates of the average offered and carried traffics on each link of a given S.N. for any period are inputs for an algorithm which, using some separation rules based on the proportions of different traffic streams, compiles the composite traffic between adjacent nodes of the S.N. and by extension the composite usage of switching node equipments. Moreover, through, the circuit requirements, the circuit allocation on the T.N. can be associated with that composite traffic so the usage of any transmission facility can be split accordingly.

3.2.2 The Costing Block (fig. 3-3)

This block is designed firstly, to apply different costing concepts to the physical facilities of the carriers, secondly, to allocate costs to the various services.

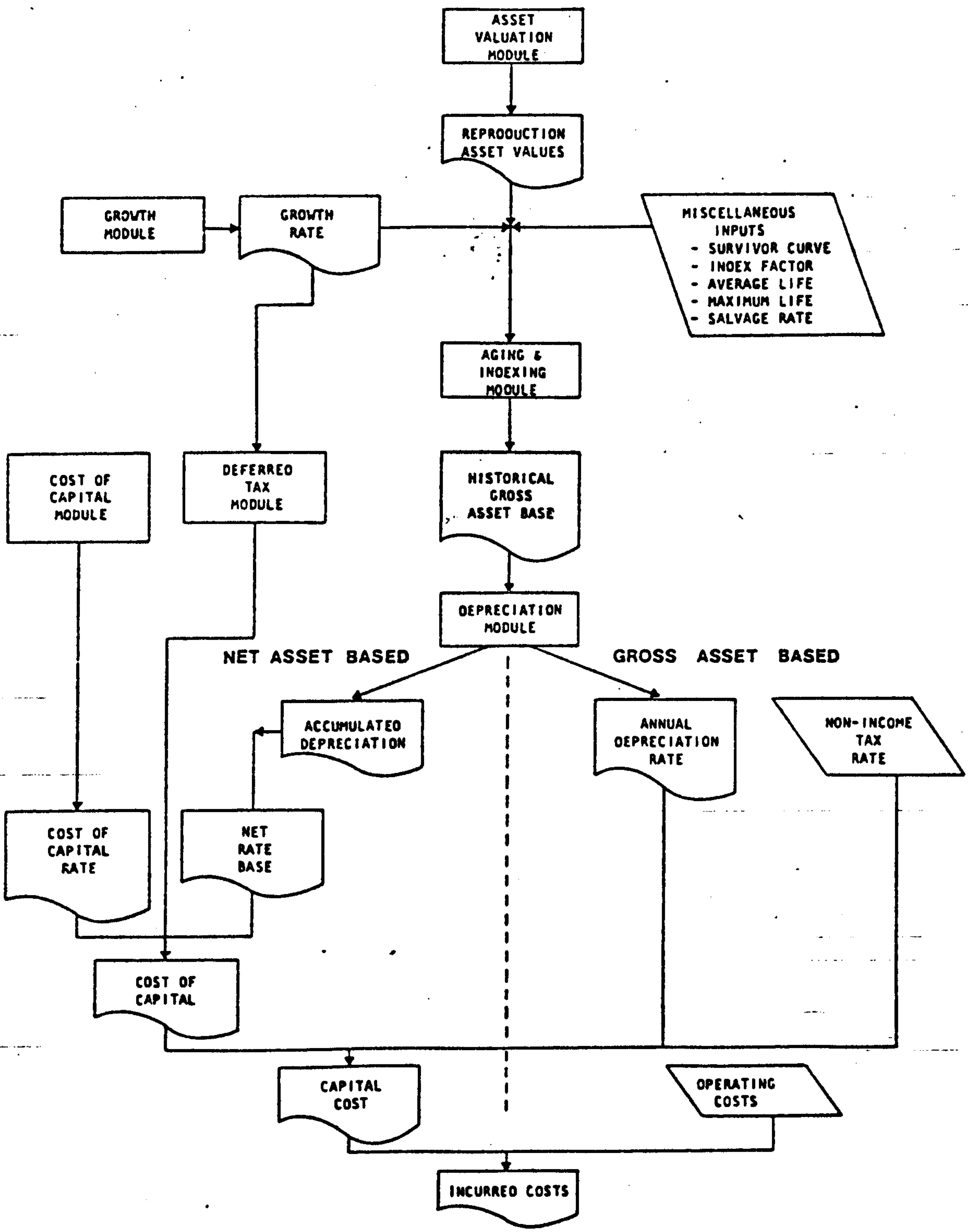
a) The incurred costs of the facilities

By definition the incurred cost a facility is made of the depreciation, the cost of capital, the operating cost and the non-income tax. For each carrier, the incurred costs are computed for the following facilities: switching, transmission, general equipment, building and land. The flowchart in Figure 3 shows the main articulations of the computations. The Asset Valuation module gives beginning-of-a (chosen)-year values for the aforementioned types of plant. The values for general equipment, building and land are taken directly from the financial statements of the carriers; but, for transmission on links and switching equipments at nodes, assets valuation functions are used. Both functions are step functions with important fixed costs at the origin. The functions for transmission differ for stations sheltering regular repeaters, branching or junction repeaters and terminal repeaters. The unit for transmission is the radio frequency channel; for the switching it is the trunk. For branching and terminal repeaters multiplexing costs are added according to a classification of links indicating the type of multiplexing plans and the number of circuits defining the effective capacity of the links; these informations are also used to inflate the circuit requirements to reflect the multiplexing practice: grouping voice circuits and not going to the maximum grouping capacity to avoid interferences.

The Assets Valuation module uses "current" values for "current" technology (it was current around 1976). If reproduction costing is pursued, there is no need for the Aging and Indexing module which applies various methods of depreciation, survival characteristics, growth rates, price and productivity indexing to derive an embedded cost, backward so to speak, in order to simulate the evolution of the gross asset base.

COSTING BLOCK FLOWCHART

FIGURE 3-3



The Cost of Capital module is calculated according to the following equation:

Cost of capital = $\{[RORE \times (1 - DCR)] + (i \times DCR)\}/(i - t)$,

where t = taxe rate, RORE = rate of return on equity,

DCR = debt/capitalization; i = average interest rate on debt.

The last component of the incurred cost, that is the operating cost, includes: maintenance, marketing and commercial, traffic, accounting, engineering, other expenses. These costs are estimated by applying relevant ratios to the asset costs.

b) Costing of services

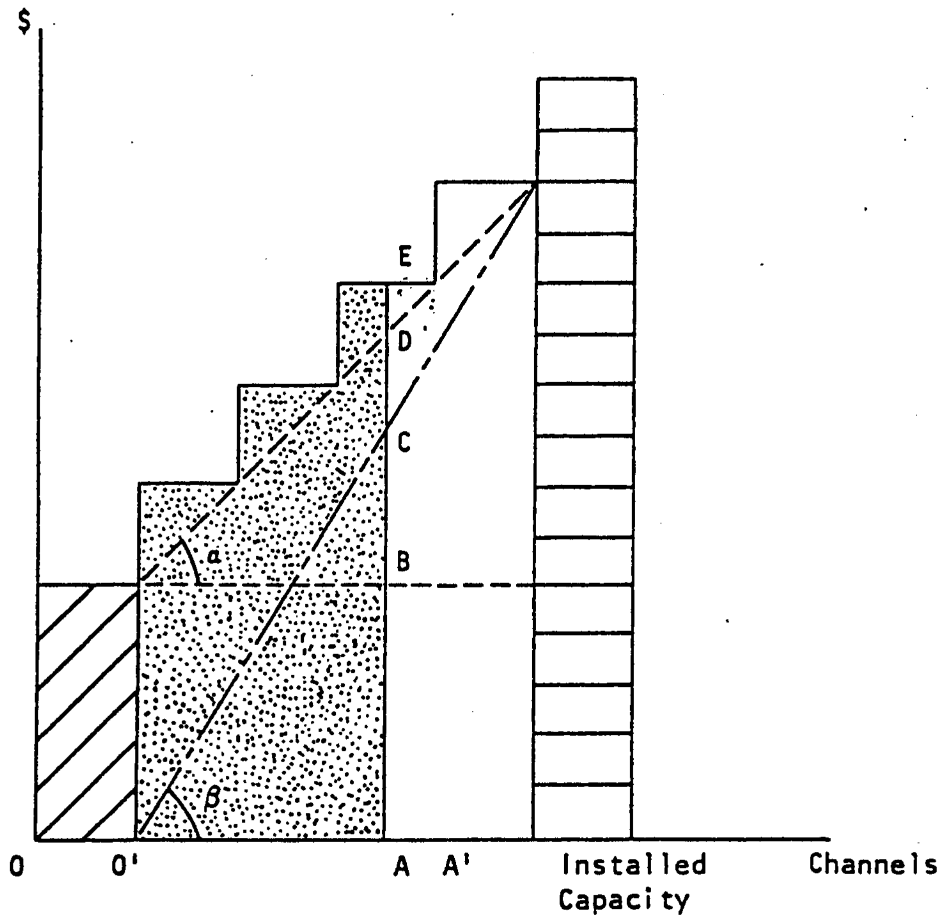
For a switched service any cost (average or incremental) is the summation of the cost due to the switching network plus the cost due to the transmission network. For a non switched service only the transmission cost is involved.

The incremental switching cost is derived from dimensioning the network twice, using the economic C.C.S. rule. Each time, a particular demand configuration is chosen and the difference between the value of equipments for those two demands is the cost looked for.

The incremental transmission cost is computed also by establishing the cost difference of two demand configurations for the entire network. The cost coefficients in the objective functions are derived from the transmission asset valuation functions. Several ways exist for choosing cost coefficients; in other words, several slopes can be drawn on the valuations functions according to the user intentions (for example: how does he want to treat excess capacity?). See figure 3-4.

FIGURE 3-4

Costing of transmission facilities



: Protection channel and infrastructure

: Video dedicated channel(s)

: Circuits actually in use.

Ag(α) : Per circuit marginal cost.

Ag(β) : Per circuit average cost.

Actual incurred cost for 0'A operational circuits: AE

Marginal cost for 0'A operational circuits: BD

Average cost for 0'A operational circuits: AC

Fixed cost ± marginal cost: AB + BD = AD

c) Costs associated with the local network

It is important to realize that the local network is the most expensive part of a telephone network and the most complex to model (for example there were more than 1200 local offices only in the Bell Canada's territory when the NPPS model was elaborated). However, since we are concerned with the long haul services, only the toll related cost of the local network should be estimated. A preliminary study showed a cost of around 4% of the cost of the total toll network, as per NPPS model, could be assigned to that problem.

3.2.3 The Sharing Block

The main inputs are the pre-settlement gross operating revenues by stream of traffic, the facility usages and the facility costs. The output is the post-settlement revenues which enter the Accounting Block.

Four this experimental model, three settlement schemes are included:

1) The "Full Division Plan of Settlement" (Trans Canada Telephone System) All common system revenues are pooled, then each member recovers the expenses assigned to the provision of the revenue generating services. The balance is distributed according to the member's share of the assigned plant value. Relative usage measures are mostly used in cost separation and even the excess capacity costs are assigned.

2) The Old Commonwealth Scheme. The pooled revenues are distributed in the same proportion as the incurred expenses (operating expenses, depreciation and cost of capital).

3) The New Commonwealth Scheme. The revenue of each stream are equally shared between the terminal partner as well as the costs of that stream. Nodes and links unit costs are evaluated including capital and operating costs, then route unit costs are derived and stream costs are computed by multiplying those unit costs by the respective stream usages and summing for all routes of a stream.

3.2.4 The Accounting Block

The major results of that block are, for each carrier or for any consolidated grouping:

1) The Balance Sheet at beginning of the year, the changes and the situation at the end of the year.

2) The Income Statement with the operating revenues, operating expenses, other expenses, income taxes, debt service charges, extraordinary items, net income available for dividends and retained earnings.

3) The Sources and Uses of Funds Statement which algebraically equal the changes in the Balance Sheet.

4) Financial ratios like the debt capital ratio, the ratio of return to equity components.

These financial statements are obtained from a simultaneous equations system (74 exogenous variables coming for the other blocks or from outside and 43 endogeneous variables).

4. Cross-subsidy Tests to Be Used in the NPPS Model

4.1 Description of Suitable Tests for Experimental Purpose

For the purpose of empirical calculations, four tests drawn from the game approach appear relevant for testing cross-subsidization. These tests were previously derived and explained and will be expressed here in their mathematical formulation only. A system (economy, carriers,...) producing and distributing a set $N = \{1,\ldots,i,\ldots,n\}$ of n services is supposed. $R(\cdot)$ and $C(\cdot)$ are respectively the revenue and the cost functions defined for a service or a group of services.

The incremental-cost test (CT) is:

$$R(i) \geq C(N) - C(N-i), \quad \text{for any } i \text{ in } N. \tag{1}$$

The stand-alone test (SAT) is:

$$R(i) \leq C(i), \quad \text{for any } i \text{ in } N. \tag{2}$$

The generalized incremental-cost test (GICT) is:

$$R(S) \geq C(N) - C(N-S), \quad \text{for all subsets } S \text{ of } N. \tag{3}$$

The generalized stand-alone test (GSAT) is:

$$R(S) \leq C(S), \quad \text{for all subsets } S \text{ of } N. \tag{4}$$

It is worth remembering that if the carrier (subset) has to meet a zero-profit constraint, and if cross-elasticities are zero (hypothesis necessarily assumed when no "demand block" exists), then GICT is equivalent to GSAT.

Although not a test but a useful "fair" cost-allocation formula, the following will also be needed:

$$F(i) = \sum_{\substack{G \subset N \\ i \in G}} \frac{(n-g)!\,(g-1)!}{n!} \{C(G) - C(G-i)\}, \qquad (5)$$

where the symbol meanings can be found in section 2.2.

4.2 Qualifications of the Present Cross-Subsidy Tests

The theoretical tests proposed in the previous section involve sets of economic agents and sets of costs. The N.P.P.S. model has been designed to show a fine level of disaggregation for traffic as well as for facility costing. Therefore, it is possible to regroup the demands of the economic agents in a meaningful way and to compute some of the several types of incremental costs used in the cross-subsidization tests. However several points must be stressed in order to show the kinds of interpretations and simplifications which are necessary to implement the theoretical tests.

i) Defining Meaningful Demand Subsets (or Services)

The theories postulate that any individual has perfect knowledge of the alternative subsets he can join and that he has communication and cooperating capacity. Also, any "subset" knows the cost of supplying its own demand. For the problem at hand, it is more realistic to postulate that intermediates (enterprises) regroup individual demands through their offering of services. The meaningful demand subsets are thus characterized by communication-streams involving: origin-destination, types of service, time of day, time of week. For instance a subset could be: "all public message traffic between 100 and 500 miles, from 8 to 18 hours in the business day". The computing cost of tests which present a combinatorial nature will force us to limit the number of demand subsets. Moreover, the regulating agencies already in place impose the regrouping in a limited number of services.

ii) Hypothesis on Demand Reactions to Prices and Quality

The demand subsets having been defined, it should be noted that empirical demand functions are not known to us. For the time being, only requirements (in C.C.S. and number of circuits for public message or in number of circuits only for all other services) for a base period or projected requirements for future periods are available. Tests involving demand elasticities (direct and crossed) are thus beyond our scope. However, since the cross-subsidy theory of the game theoretic type is not sufficiently developed to include demand elasticities, these will not be considered in this study.

iii) Data Availability for Costing the Services Required by the Demand Subsets

The cost associated with a given demand subset is theoretically the minimum total cost to supply that subset alone. In real situations, the "initial state" must take care of the physical network and institutional organizations already in place. The "moves" of any subset of economic agents are not built from scratch. The actual network design will impose its structure of nodes and links and most of the incremental cost configurations. For stand-alone tests of relatively small demand subsets, the cost functions available in N.P.P.S. and network configurations will not be satisfactory, since the available network has not been designed for such demands.

iv) Dynamic Aspect, Hidden and Explicit

Up to now, the N.P.P.S. model, except the Accounting Block, is a one period model. It computes results for one current year. The cross-subsidy theories above do not have time explicitly as a variable. One can always think of a typical year or of a planning horizon during which decisions are made but the computing of costs is quite different in each case. The latter case requires facility expansion features linked to forecast demands. Some conceptual development has been done along that line (see [1]) but no software is available yet.

Even if the one period method is retained, the hidden dynamic characteristics are represented by the existence of excess capacities which can be justified by the indivisibility of installed facilities and other economies of scale combined with growing demands. Therefore, the cost of excess capacities should be either excluded or imputed to the cost of the tested services. Several solutions will be proposed in the second serie of simulations.

4.3 Empirical Test Proposals

In the previous sub-section, it was seen that the networks initial states and the data availability were of paramount importance. Although embedded cost scenarios can be run with the use of the Aging, Indexing and Depreciation programs, the costing concept retained for present computations is the incurred cost based on the reproduction cost. Since our asset valuation functions are of the "fixed cost" type, among many others, two obvious possibilities are available for each existing network element. Average cost or marginal cost (link or node) from which the incremental cost of a "service" (a requirement subset for the entire switching and transmission networks) is computed. The tests will be executed with both concepts (the definition of which is recalled in Figure 3.1).

Four cross-subsidy tests will be presented:

i) Public Messages Versus Private Lines

This is a recurrent question. Private lines should at least pay for their incremental cost.

ii) Origin-destination pairs less than or equal to 1 000 miles apart versus pairs more than 1 000 miles apart

It is possible that very long lines were favored. Time did not permit the regrouping of mileage bands used in tariff tables using a clustering device as: a new (larger) mileage band is created if the tariffs that form it do not deviate from the average tariff by more than a fixed amount. By such reasoning, long distance calls can be approximately clustered in equi-tariff bands: 0 to 180 miles, 181 to 540 miles, 541 to 1 200 miles, over 1 200 miles.

iii) Regional-adjacent-non-adjacent (including U.S.) traffics

Negociations between carriers distinguish these three types of traffic.

iv) Peak-hour Traffic Versus Non Peak Traffic

A thorny question in economies is whether peak users are subsidized or not by off-peak users. A possible formulation of such a question may be the following: we know that the traffic matrices are dimensioned with respect to the peak demand (rather a kind of average peak demand). If we are given the information that the average demand is about 70% of the peak demand, what is the incremental cost from that average to the full 100%? And does the incremental revenue cover this cost? Alternatively, any percentage down from peak demand could be costed.

v) Full Allocation Versus "Fair" Formula

This is not a cross-subsidy test, but a comparison between two cost allocations: full allocation based on usage and "fair" formula based on the "fair" postulates enumerated in section 2.2, a formula which is a weighted average of all possible incremental costs that a service can add when it joins all possible combinations of other services.

4.4 Treatment of Excess Capacities

The existence of excess capacities can be explained in several ways: simple planning error, redundancy for survivability, decreasing demand along a cycle or trend, indivisibility of optimal facilities associated with relatively small demands, growth reserve accumulated to protect against any large positive demand variation, growth reserve built to take advantage of economies of scale when the enterprise faces a sustained growing demand. In telecommunication networks, "protection" facilities and indivisibilities leading to economies of scale are frequently mentioned explanations that we can associate with rapidly growing demand. In other words, in such a dynamic setting, growth reserves will benefit future as well as present customers. It is therefore important to impute at least part of the excess capacities to actual services.

In devising several methods to take account of excess capacity when computing incremental costs, we will initially allocate all excess capacity between services, first according to the "fair" formula approach and second, proportionally to utilization. Secondly, keeping in mind that allocation may be made according to game theory or to usage, we will distinguish pure excess capacity and growth reserve by introducing growth rates for services. A last method of treating excess capacity will propose some trade-off between present and future.

The five methods depicted below all obey the same pattern. The incremental cost to be used in the incremental-cost test will be modified. It will be the sum of the previously calculated incremental cost and a term representing a certain part of the excess capacity. Thus, the incremental cost IC(i) for service i will be:

$$IC(i) = C(N) - C(N-i) + EC(i)$$

where EC(i) is the value of excess capacity imputed to service i. Of course, the expression obtained is not a "true" incremental cost, but an "exhaustive" incremental cost in the sense that excess capacity is taken into consideration in the procedure. The methods described obviously may be applied to any service.

In the first two methods, the principle is the same. We admit that the cost of excess capacity must be supported by present customers, whether excess capacity is a growth reserve for the future or an incorrect forecast of future demand. We thus run the model with a specified demand and. obtain the magnitude of excess capacity. This excess capacity may be priced on a marginal basis or with average coefficients. This procedure permits the cost of excess capacity to be obtained by link or node.

METHOD A

With the first method, we want to allocate the cost of excess capacity proportionally to the usage of the element. We then multiply the cost of excess capacity on each link by the relative usage of this link. To obtain the term EC(i) for service i, we proceed in the same way for all links.

This method puts the weight of the cost of excess capacity only on the shoulders of the present generation. Moreover, it is based on the actual relative utilization of the elements and this may be completely out of line with the future usages.

METHOD B

This method adopts the same approach as that employed in method A but allocates excess capacity according to a fairness and game theoretic view. We remember that the cost-allocation formula:

$$F(i) = \sum_{\substack{G \subset N \\ i \in G}} \frac{(n-g)!\,(g-1)!}{n!} \{C(G) - C(G-i)\}$$

allows a fair separation of common costs. We thus can allocate the cost of excess capacity in proportion to these game theoretic coefficients. The term EC(i) would then equal:

$$EC(i) = \frac{F(i)}{\sum_{i \in N} F(i)} EC(N)$$

where EC(N) is the cost of the total excess capacity for all services over the whole network.

This method, as well, puts the burden for excess capacity, on the present generation only.

The next four methods try to make a distinction between growth reserve which tends to meet an expanding demand as accurately as possible, and what is called pure excess capacity which is the surplus of capacity over the growth reserve. It is probable that the notion of pure excess capacity will require a new interpretation when multiplexing costs are more thoroughly understood. For the moment, we shall accept this concept.

In this perspective, we shall choose a moving horizon of three years since it is admitted that facility installations are anticipated for a period of at least two to six years. Hence, we run the model successively for three years, increasing the demand for each service according to a growth rate particular to each service and determined exogenously. This rate might be of the multiplicative form with $d_i(t) = a_i(1 + r_i)^t$, where $d_i(t)$ represents the demand of service i after a lapse of t years, a_i is the demand of service i presently, and r_i is the growth rate.

The philosophy of these two methods lies in the hypothesis that only growth reserve must be imputed to customers and then allocated between the services. Pure excess capacity must be supported only by the carrier.

Two cases are possible after three years. First, all the links are saturated. In that case, all excess capacity is growth reserve, and all excess capacity has to be separated between services. This possibility reduces to the first two methods previously discussed. In the second case, there is excess capacity on some or all links after having run the model with demand $d_i(3)$. Pure excess capacity is therefore present in the network and must be borne by the carrier. We need only allocate growth reserve in order to execute the cross-subsidy test. This method, however, necessitates some expansion features in the model since after each year some links could be saturated and block future growth even if ample excess capacity still existed on most of the links.

METHOD C

This method allocates the growth reserve only and does it on the basis of present utilization. It represents an improvement on method B since the burden imposed on present consumers corresponds only to their probable growing demand.

METHOD C'

This method is very similar to method C but allocates growth reserve on the basis of future utilization.

METHOD D

This method espouses the same spirit as method C since it attempts to allocate only growth reserve. However, here the principle on which separation is grounded is the fair allocation formula. The new incremental cost would be:

$$IC(i) = C(N) - C(N-i) + \frac{F(i)}{\sum_{i \in N} F(i)} GR(N)$$

where the F(i)'s constitute a fair separation of costs incurred by the present level of utilization and GR(N) is the value of the total growth reserve to be allocated.

METHOD D'

This method is similar to method D but allocates growth reserve on the basis of future utilization.

5. Empirical Evaluation of Selected Cross-Subsidy Tests

5.1 General

We present in this section the results obtained for various tests performed using the NPPS model. This first series of tests simply aimed at comparing generated revenues of a service to its stand-alone and incremental costs based on current use of the telephone plant. Obviously, these testswere performed on groups of services where cross-subsidy was suspected, i.e.:

- Public messages / private lines
- Short distance / long distance toll traffic
- Peak traffic / off-peak traffic
- Regional / adjacent / non-adjacent and U.S. traffic

These tests led to the preliminary conclusion that, based on current usage, incremental cost tests were always satisfied given the paramount importance of the plant commonly used by all services and consequently two avenues of research were explored.

- the first one consisted of increasing incremental costs by incorporating the required growth reserve associated with the service. This led to the elaboration of tests based on the prospective use of equipment;
- the second consisted of imposing a definition of cross subsidy much more demanding than the one based on stand-alone and incremental costs alone. This led to the definition of various full cost allocations formula.

This second series of tests was performed on a single separation of services, namely, public messages vs private lines.

5.2 Tests Based on Current Use of Equipment

5.2.1 Public Messages and Private Lines

Table 5-1 shows the total costs incurred in the switching and the transmission networks required to accommodate first public messages alone and secondly both services. The difference is the incremental cost of private lines. Since this service is not a switched service, there is obviously no incremental cost in the switching network Also appearing in Table 5-1 are the revenues derived from the services considered as estimated in N.P.P.S. All figures are shown separately for Bell Canada and for the whole network, the relationship between incremental costs and revenues not being always the same at the carrier level.

These comparisons must however be handled very carefully since estimated revenues and costs are not strictly comparable. As a matter of fact, revenues correspond to the part of the service generated in the carrier's territory while costs are those associated with satisfying the whole service over the said territory. For instance, the incremental cost of non-adjacent traffic for the Bell is constituted by the cost of originating, terminating and going through non-adjacent traffic, while calculated revenues are those generated by originating traffic only.

TABLE 5-1

Incremental Cost of Private Lines
(incurred costs and revenues in $ millions)

Service	Carrier	Switching	Transmission Cost(1)	Total	Estimated Revenues(2)
Public Messages	Bell	64.4	16.0	80.4	316.9
	Network	95.8	23.6	119.4	395.2
Both Services	Bell	64.4	21.9	86.3	341.5
	Network	95.8	33.3	129.1	436.7
Incremental Costs & Revenues of Private Lines	Bell	0	5.9	5.9	24.6
	Network	0	9.7	9.7	41.6

(1) Using average cost.
(2) US revenues excluded.

5.2.2 O-D pairs < 1 000 miles apart / O-D pairs > 1 000 miles apart

In order to test whether one group of customers cross-subsidized another, three simultations were performed:

- one with all traffic between cities more than 1 000 miles apart;
- one with all traffic between cities less than 1 000 miles apart;
- one with both types of traffic.

Since destination/origin points in the U.S. are not precisely known, U.S. traffic was deliberately omitted from all three simulations.

Table 5-2 is very similar to Table 5-1 and yields the incremental costs of both types of traffic. It can be seen that for pairs > 1 000 miles apart revenues exceed incremental costs by a factor of about 17. For pairs < 1 000 miles apart, the ratio is somewhat lower at about 7.

5.2.3 Regional/Adjacent/Non-adjacent and U.S. traffic

A three-service experiment gives us the possibility of performing six incremental cost tests and requires seven simulations.

Total incurred costs for each subset of services are shown in Table 5-3. Resulting incremental costs for each service or combination of two services appear on Table 5-4 where they are compared to corresponding revenues. In all cases, revenues are larger than incremental costs; in order words, all tests are passed. The ratio of revenues over incremental costs varies however quite substantially between simulations and between carriers as shown in the last column of Table 5-4.

5.2.4 Peak Hour Traffic / Non-peak Traffic

Traffic profiles during an average business day have the general form shown in Figure 5-1.

TABLE 5-2

Incremental Cost of O-D Pairs More or Less Than 1 000 Miles Apart ($ millions)

Simulation and Carrier	Switching Cost	Transmission Cost(1)	Total Cost	Estimated Revenues
Pairs < 1 000				
Bell	60.7	10.2	70.9	298.4
Network	90.7	13.8	104.5	352.9
Pairs > 1 000				
Bell	53.6	2.0	55.6	18.5
Network	83.6	4.7	88.3	42.2
Both Services				
Bell	61.7	11.6	73.3	316.9
Network	94.7	17.2	111.9	395.1
Incremental Costs & Revenues for pairs < 1 000				
Bell	8.1	9.6	17.7	298.4
Network	11.1	12.5	23.6	352.9
Incremental Costs & Revenues for pairs > 1 000				
Bell	1.0	1.4	2.4	18.5
Network	4.0	3.5	7.5	42.2

(1) Using average cost.

TABLE 5-3

Three-service Experiment
Total Incurred Costs and Revenues
($million)

Simulation & Carriers	Switching Cost	Transmission Cost (1)	Total Cost	Estimated Revenues (2)
Reg + Adj + N-Adj + US				
Bell	64.4	16.1	80.5	316.9
All carriers	95.8	23.5	119.3	395.2
Adj + N-Adj + US				
Bell	56.1	7.7	63.8	33.6
All carriers	87.5	14.6	102.1	96.1
Reg + N-Adj + US				
Bell	63.7	15.3	79.0	305.2
All carriers	94.7	21.4	116.1	352.9
Reg Adj				
Bell	60.7	9.9	70.6	292.2
All carriers	90.7	12.5	103.2	341.4
N-Adj + US				
Bell	56.1	6.9	63.0	21.8
All carriers	86.6	12.4	99.0	53.8
Adj				
Bell	53.6	1.3	54.9	11.8
All carriers	83.6	3.3	86.9	42.3
Reg				
Bell	59.8	9.0	68.8	283.4
All carriers	89.9	10.0	99.9	299.1

(1) Using average cost.
(2) Excluding US

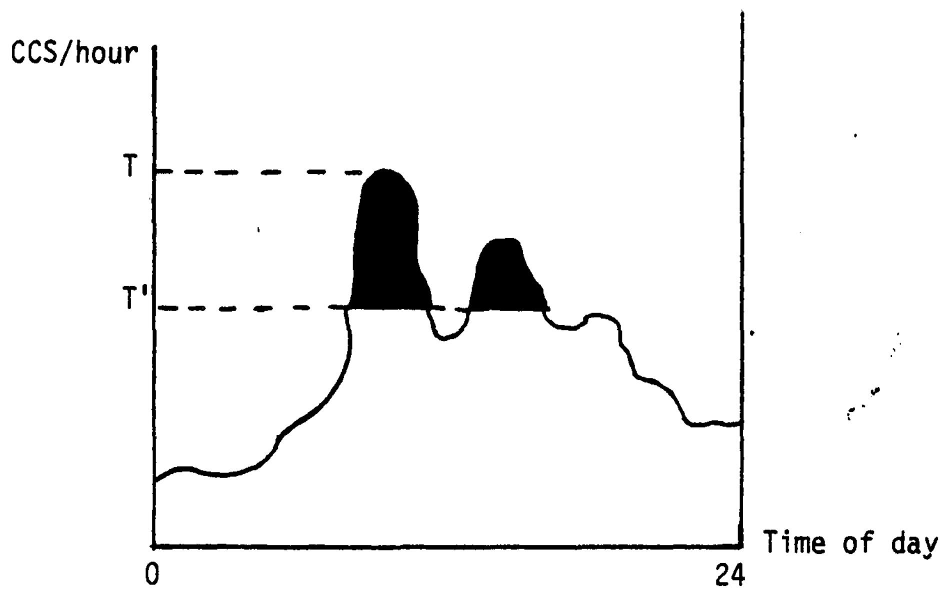
TABLE 5-4

Three-service Experiment
Incremental Costs
($ million)

Subset test	Carriers	Incremental cost in switching	Incremental cost in transmission	Total incremental cost(2)	Revenues N.P.P.S. Estimates (1)	Ratio (1)/(2)
Adj +	Bell	4.6	7.1	11.7	33.6	2.9
Non-Adj + US	All	5.9	13.6	19.5	96.1	4.9
Reg +	Bell	10.8	14.8	25.6	305.2	11.9
Non-Adj + US	All	12.2	20.3	32.1	352.9	11.0
Reg +	Bell	8.3	9.2	17.5	292.2	16.7
Adj	All	9.2	13.2	22.4	341.4	15.2
Non-Adj + US	Bell	3.7	6.2	9.9	21.8	2.2
	All	5.1	11.1	16.2	53.8	3.3
Adj	Bell	.7	.8	1.5	11.8	7.9
	All	1.1	2.2	3.3	42.3	12.8
Reg	Bell	8.3	8.4	16.7	283.4	17.0
	All	8.3	9.1	17.4	299.1	17.2

FIGURE 5-1

Typical Traffic Profile



The network is dimensioned for the peak hour traffic T and costs C(T). Should the peak-hour traffic be smaller, say T', a smaller cost would result C(T'). The test hence consists in comparing the incremental cost of peak-hour traffic (i.e. C(T) - C(T')) to the revenues it generates. These revenues are calculated by multiplying the shaded area of Figure 5-1 by the appropriate tariff. For this experiment, T' was arbitrarily set at 70% of T.

Total incurred costs for peak and reduced peak simulations are presented in Table 5-5. The incremental cost of peak traffic is derived in Table 5-6 and compared to its revenues.

It can be observed that once more incremental revenues largely exceed incremental costs.

TABLE 5-5

Total Incurred Costs
Peak/Off-peak Traffic
($ millions)

Simulations & Carrier	Switching Costs	Transmission Costs[1]	Total Costs
Peak			
Bell	64.4	16.1	80.5
All carriers	95.8	23.5	119.3
Reduced Peak			
Bell	61.5	13.0	74.5
All carriers	92.8	19.3	112.1

(1) Using average cost.

TABLE 5-6

Incremental Costs of Peak Traffic
($ millions)

Carrier	Switching	Transmission	Total[2]	Revenues[1] N.P.P.S. Estimates (*)	(1)/(2)
Bell	2.9	3.1	6.0	41.5	6.9
All carriers	3.0	4.2	4.2	51.8	7.2

(*) Excluding US

5.2.5 Preliminary Comments on First Series of Tests

The tests presented so far seem to indicate that the incremental cost is always satisfied. In addition, the ratio of revenues over incremental cost is so large that it could hardly be reduced to values inferior to 1 simply by improving certain approximations of the model.

It appears in certain instances that revenues exceed the stand-alone cost of a service (e.g. public messages). Strictly speaking, the stand-alone cost of a service should be representative of all facilities required to support this service. Consequently, the stand-alone cost of any long distance service would include the cost of the local network. Given its relative importance in the total plant (see section 3), it becomes clear that stand-alone cost tests are also satisfied.

The most important point to notice however is the large discrepancy which exists between the cost of the existing transmission network and the part which is allocated to the various services or groups of services tested in this section. The total cost incurred in the toll network (as estimated in N.P.P.S.) is compared to the cost allocable to public messages and private lines in Table 5-7.

TABLE 5-7

Comparison of Total Cost of Plant to Cost Allocable to Public Message and Private Lines ($ millions)

Switching	Total incurred cost of plant as estimated in N.P.P.S. (2)	Cost allocated to PM and PL (3)	(3)/(2) %
Switching Network	106.7	94.7	89
Transmission Network (1)	184.8	33.3 (4)	18
Total	291.5	128.0	44

(1) Excluding channels used for video.
(4) It will be seen in Table 5-8 that when using the (fixed cost + marginal cost) formula this value becomes 86.1.

It becomes clear from this table that this difference has to be explained before any further tests are performed and we give below a list of possible contributing factors.

i) Circuit requirements as estimated by dimensioning the switching network are far below those contained in the data base (14 1000 vs 23 600). It must be remembered that the dimensioning algorithm is applied to traffic which

- is estimated based on limited data (traffic between 17 cities during two weeks of July 1971);
- does not include WATS, TWX and data transmission.

ii) It was mentioned earlier that costing the transmission network with the average cost formula is a poor approximation when the link loading is low.

It will be seen, see Table 5-8, that costing transmission facilities with the (fixed + marginal) cost approach would result in a total cost of $86.1 millions to be compared to $33.3 millions obtained with the average cost formulation (Table 5-7).

iii) It is known that trade-offs between multiplexing and ratio costs result in a channel loading which generally does not exceed 75%.

iv) A certain amount of unused equipment is includedin the plant as a growth reserve.

v) Finally, it must be remembered that the N.P.P.S. allocation procedure does not take survivability constraints into account and therefore yields and allocation which is cheaper than it would be in reality.

5.3 Tests Based on Prospective Use of Equipment

5.3.1 General

In view of the results presented in the previous section, a new seriesof tests was performed. It was decided to concentrate on the appropriate calculation of costs rather than on various splits of the services taken into account. All tests were consequently based on a public message/private lines separation. In order to improve estimation of costs and in line with the observations of section 5.2.5, the following rules were applied:

i) Transmission facilities were costed using the fixed cost + marginal cost approach.

ii) The multiplexing plan was approximated by the formulation suggested that circuit requirements constitute integer number of groups, the loading of which does not exceed 75%.

iii) Since no precise definition of the growth reserve is available, various policies were tested by which growth reserve was defined as the incremental cost associated with the growth of a service over 1, 2 and 3-year periods.

5.3.2 Description of Simulation Runs and Incremental Cost Tests

Five simulation runs were performed. The first one is based on present demand. The next three consider prospective demand 1, 2 and 3 years from now using:

- a 12% annual growth rate for public messages(1)
- an 18% annual growth rate for private lines(1).

To test the sensitivity of the results to growth rates, a fifth simulation was performed considering prospective demand in year 3 but with a 10% annual growth rate for private lines.

---

(1) These rates were applied uniformly to all existing demands and no new demands were considered.

The results of the five simulations are presented in Table 5-8. It must be remembered that, private lines being a non-switched service, only transmission costs have been analyzed.

One will also notice that the total cost of the transmission network increases with the length of the planning horizon since capacity had to be increased on a certain number of links in order to render the allocation feasible. The corresponding incremental cost of private lines can easily be derived from these results and is shown below.

Basis of calculation for growth reserve	Incremental cost of private lines including growth reserve ($ millions)
No growth reserve	10.1
One year planning horizon	12.9
Two year planning horizon	15.2
Three year planning horizon	16.8
Three year planning horizon (lower growth rate for P.L.	13.4

Although consideration of a 3-year growth reserve increases the incremental cost of private lines by 66%, the revised incremental cost figure still remains much smaller than revenues estimates of $41.6 millions.

5.3.3 Tests Based on Full Allocation of Costs

All tests performed so far have shown that the incremental costs of private lines is always covered by generated revenues. If one examines closely the total transmission cost of supporting both private lines and public messages, it can be broken down as follows:

TABLE 5-8

Simulation Results
Incurred costs in $ millions

Simulation	Services considered	Incurred fixed cost	Incurred variable cost[2]	Total incurred cost	Cost of excess capacity[1]	Total cost of transmission NW (excluding channels used for video)
#1						
Present demand	P.M. only	49.4	26.6	76.0	108.8	184.8
	P.L. only	39.1	10.7	49.8	135.0	
	Both services	49.4	36.7	86.1	98.7	
#2						
Demand after one year of growth	P.M. only	49.4	28.6	78.0	107.8	185.8
	P.L. only	39.1	12.2	51.3	134.5	
	Both services	49.4	41.5	90.9	94.9	
#3						
Demand after two years of growth	P.M. only	49.4	31.1	80.5	109.3	189.8
	P.L. only	39.1	13.7	52.8	137.0	
	Both services	49.4	46.2	95.7	94.1	
#4						
Demand after three years of growth	P.M. only	49.4	34.1	83.5	108.2	191.7
	P.L. only	39.1	15.6	54.7	137.0	
	Both services	49.9	50.4	100.3	91.4	
#5						
Demand after three years of growth (lower rate for P.L.)	P.M. only	49.4	34.1	83.5	107.8	191.3
	P.L. only	39.1	13.4	52.5	138.8	
	Both services	49.4	47.5	96.9	94.4	

(1) Including $9.2 millions for links not used at all.
(2) It can be seen that the incurred variable cost associated with both services is generally lower than the sum of individual variable costs. This results from the rounding procedure which, when applied to both services, results in requirements smaller than the sum of individual rounded requirements.

	$ millions
Incremental cost of private lines	10.1
Incremental cost of public messages	36.3
Cost of equipment used jointly	39.7
Total transmission cost (excluding growth reserve)	86.1

If one further considers a 3-year growth reserve and compares all these costs to total costs of the existing transmission network one obtains a graph of the form shown in Figure 5-2, where surfaces are proportional to costs.

It becomes clear then that a definition of cross-subsidy based on incremental costs alone is not sufficient given the importance of the common costs and other non directly allocable costs and given that total costs must eventually be recovered.

Two questions then arise:

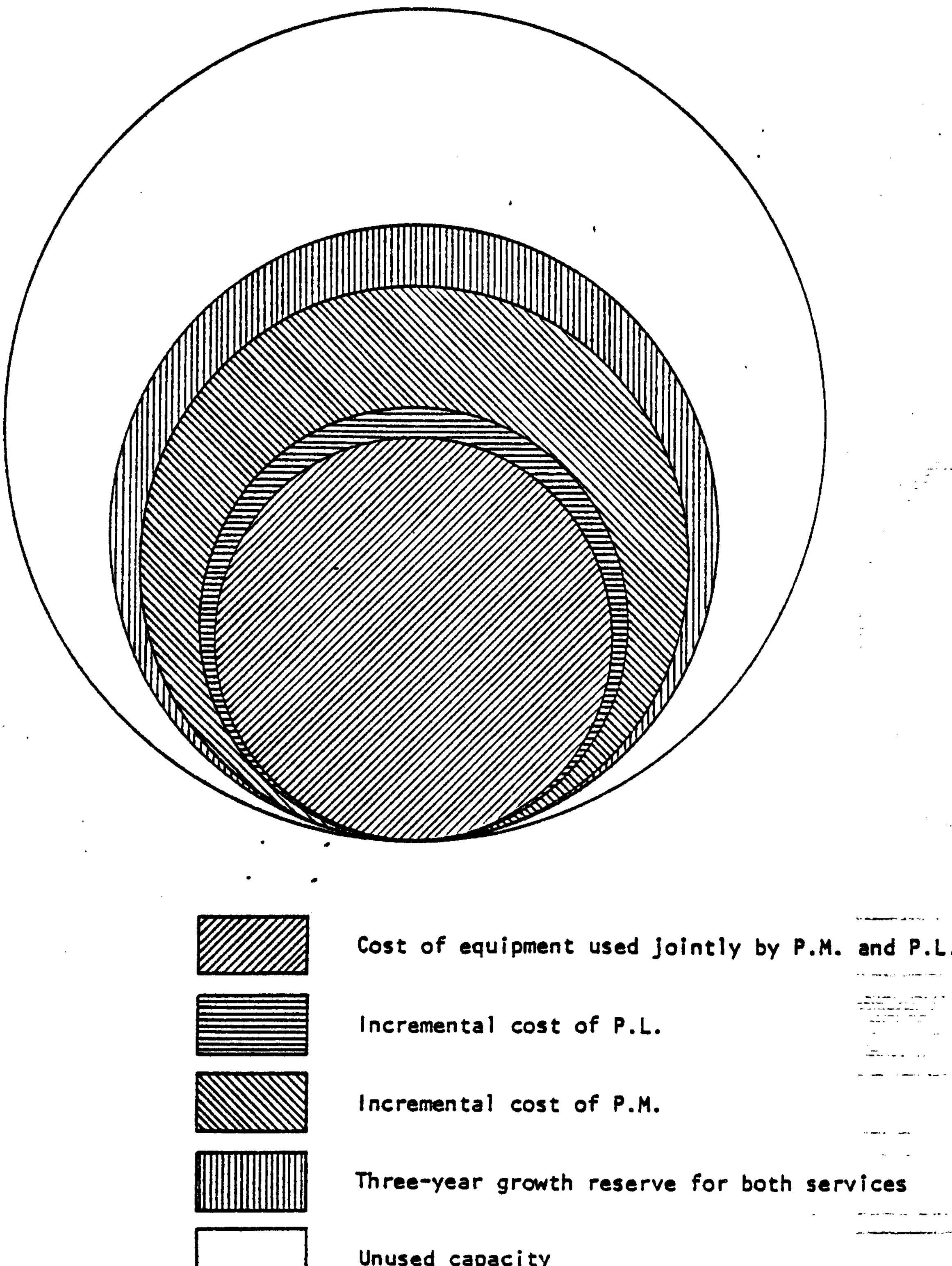
i) How should common cost be allocated?

ii) Which common costs should be allocated, namely, should the cost associated with the so called "pure excess" capacity be paid by the consumer or by the carrier. This depends obviously on the origin of this excess which could result from:

- deficiencies of the model (i.e. not enough traffic, no survivability constraints...);
- a larger planning horizon than used in our calculations (i.e. (i.e. more than three years);
- a very safe and/or suboptimal planning of the network by the carriers;
- a mixture of the three above-mentioned factors.

FIGURE 5-2

Allocation of Incurred Costs in Transmission Network as Estimated by N.P.P.S.
(Surfaces Are Proportional to Estimated Costs)



This leads us to the application of cost separation formula presented in section 4.2 of this paper.

Table 5-9 presents all data necessary to calculate cost allocations using the methods described above. The first three columns (Standalone cost, incremental cost and "fair" allocation of used capacity) are directly derived from Table 5-8. The allocation based on usage was obtained by the N.P.P.S. model.

Table 5-10 presents cost separations based on methods A and B. It can be noticed that for both methods, full allocated costs of private lines exceed the estimated revenues of $41.6 millions. It can also be seen that the "exhaustive" incremental cost (defined the true incremental cost plus a "fair share" of excess) yielded by method B for private lines also exceeds revenues (i.e. $44.6 vs $41.6 millions).

Cost allocations based on methods C and C' for various planning horizons are exhibited in Table 5-11. Both these methods do not allow for the estimation of an "exhaustive" incremental cost and only full allocations are computed. It can be seen however that revenues of private lines always exceed their fully allocated cost independently of the planning horizon and the method chosen.

Cost allocations based on methods D and D' for various planning horizons appear in Table 5-12. It can be seen that the "exhaustive" incremental cost of private lines never exceeds $15.2 millions while the fully allocated cost varies between $31.7 and $35.1 millions according to the planning horizon and the method selected. One can also notice that fully allocated costs based on the game theoretic approach always disfavour private lines when compared to allocations based on usage. As a matter of fact, a game theoretic allocation splits evenly the costs of the cost among participating services while the split is proportional to usage in the other method.

TABLE 5-9

Cost Allocation of Used Capacity
($ millions)

Simulations	Services	Stand alone cost	Incremental cost	"Fair" allocation of used capacity	Allocation of used capacity based on usage (derived from the N.P.P.S. model)
#1	P.M.	76.0	36.3	56.1	61.9
Present	P.L.	49.8	10.1	30.0	24.2
#2	P.M.	78.0	39.6	58.8	65.3
One year	P.L.	51.3	12.9	32.1	25.6
#3	P.M.	80.5	42.9	61.7	66.8
Two years	P.L.	52.8	15.2	34.0	28.9
#4	P.M.	83.5	45.6	64.5	68.3
Three years	P.L.	54.7	16.8	35.8	32.0
#5	P.M.	83.5	44.4	64.0	70.9
Three years slower growth for P.L.	P.L.	52.5	13.4	32.9	26.0

TABLE 5-10

Cost Allocations Based on Methods A and B
($ millions)

		P.M.	P.L.	P.M.	P.L.
Used Capacity	Common Costs	61.9	24.2	19.8	19.8
	Incremental Cost			36.3	10.1
Unused Capacity		71.0	27.7	64.3	34.5
Total		132.9	51.9	120.4	64.4

TABLE 5-11

Cost Allocations Based on Methods C and C'
($ millions)

Planning Horizon		Method C		Method C'	
		P.M.	P.L.	P.M.	P.L.
One year	Used capacity	61.9	24.2	61.9	24.2
	Growth reserve	3.5	1.3	3.4	1.4
	Total	65.4	25.5	65.3	25.6
Two years	Used capacity	61.9	24.2	61.9	24.2
	Growth reserve	6.9	2.7	6.7	2.9
	Total	68.8	26.9	68.6	27.1
Three years	Used capacity	61.9	24.2	61.9	24.2
	Growth reserve	10.3	4.0	9.7	4.5
	Total	72.2	28.2	71.6	28.7
Three years (lower growth on P.L.)	Used capacity	61.9	24.2	61.9	24.2
	Growth reserve	7.8	3.0	7.9	2.9
	Total	69.7	27.2	69.8	27.1

Estimated Revenues of Private Lines: 41.6

TABLE 5-12

Cost Allocation Based on Methods D and D'
($ millions)

Cost Component	Planning Horizon: One year P.M.	One year P.L.	Two years P.M.	Two years P.L.	Three years P.M.	Three years P.L.	Three years (lower growth P.L.) P.M.	Three years (lower growth P.L.) P.L.
Used capacity[(1)]								
Common costs	19.8	19.8	19.8	19.8	19.8	19.8	19.8	19.8
Incremental cost	36.3	10.1	36.3	10.1	36.3	10.1	36.3	10.1
Total	56.1	30.0	56.1	30.0	56.1	30.0	56.1	30.0
Growth reserve according to D	3.1	1.7	6.3	3.3	9.3	4.9	7.0	3.8
Exhaustive incremental cost according to D	39.4	11.8	42.6	14.4	45.6	15.0	46.3	13.9
Growth reserve according to D'	3.1	1.7	6.2	3.4	9.1	5.1	7.1	3.7
Exhaustive incremental cost according to D'	39.4	11.8	42.5	14.5	45.4	15.2	43.4	13.8
Total allocable cost according to D	59.2	31.7	62.4	33.3	65.4	34.9	63.1	33.8
Total allocable cost accodding to D'	59.2	31.7	62.3	33.4	65.2	35.1	63.2	33.7

(1) Identical for both methods and independent of planning horizon.

6. Assessment of Tests Performed

6.1 Sumary of Results

Going from theoretical test statements to empirical implementation with the logic of the NPPS model and the data at hand, the computations have shown that the generalized incremental tests (GICT) are passed for all partitions of services chosen in the tested examples. Moreover, in each example, if the sub-additivity hypothesis as well as the hypothesis asserting that the revenue of the "grand coalition" equals its cost, are true, the generalized stand alone tests (GSAT) are also passed without having to be computed. This somewhat reduces the problems since the actual network configuration and its associated costs are often not appropriate for a small service to stand alone. Therefore, if one is willing to accept the notion of cross-subsidy as described earlier, it follows that no such subsidy has been detected in our examples.

Another clear and interesting finding is the fact that incremental costs are often relatively small with respect to common cost. This could explain the large difference observed between the revenue generated by a subset of services and its incremental cost. As a further result, it should be noted that throughout the test series it has been recognized that a relatively large installed excess capacity was present in the network model over a normal three years growth reserve. However, some transmission links had been found to be saturated in the prospective use base tests.

In the course of this paper, we have outlined a number of model and formulation qualifications which could affect the outcome of tests performed. However some sensitivity analyses have been done with the model. Among those sensitivity studies, let us mention the introduction of the demand by WATS and TWX, the take into account of some indivisibilities in the transmission network by costing it in integer numbers of channels, by increasing the annual growth rate of demand, and by modifying the treatment of the multiplexing cost and finally the introduction of survivability constraints. It can be shown (see [7]) that any of the factors considered could not, independently, invalidates the results of our private lines incremental cost test.

6.2 Guidelines for Further Work

In the above assessment several unsolved difficulties were mentioned: the role of the cross-elasticities, the treatment of excess capacities in the tariff determination, the inter-carrier cross-subsidies.

The recent literature on multiservice regulated companies frequently discusses the cross-elasticity concept. In a more concrete vein, the empirical effort of the F.C.C. explicitely requires, in Method 7, a role for empirically evaluated elasticities for future tariffs. The NPPS team should as far as possible introduce these price reaction coefficients in some Demand Block for simulation purposes. The introduction can be more or less sophisticated and at different levels of service aggregation but the difficulty is heightened by the lack of data and the need to forecast the elasticities.

Excess capacities can always be apportioned to existing services on a more or less arbitrary basis. The challenge is to redefine services over several periods and to prove that economies of scale are benefiting present and future consumers. Phenomena of technology diffusion must also be taken into account since new equipment is progressively introduced along with retirement plan implementation. Finally, fast growing new service is a problem. What is the best way to finance the required capacity if large indivisible facilities are economically the best choice? It appears that finding the best capital deployment in a multiservice, multiperiod scheme, is certainly a difficult thing, and finding a financing scheme, which will be a burden only for the responsible service(s) without intertemporal cross-subsidy is another challenging problem.

All tests evaluated in this paper were done at the level of the national system. The problem which the regulator faces however, involves additional complexity since he is testing for cross-subsidies at the carrier level. It is evident that the non-existence of cross-subsidies among services at the national level does not necessarily imply the same at the carrier level. Furthermore, tests at the national level do not allow the probing of eventual cross-subsidies among carriers. In the course of section 5, we have presented partial results at the carrier level but indicated that the costs thereby derived can only be compared to generated revenues if the service considered is regional. Otherwise it should be compared to post-settlement revenues. In the latter case, if any of those incremental costs did exceed corresponding post-settlement revenues, it could indicate either that the sharing principle is disadvantageous to the considered carrier or that the prevailling rate structure is not cross-subsidy free. We present below a methodology which could determine which case applies.

i) First Series of Tests

The first series of tests would compare the incremental cost on the whole network of interregional (i.e. adjacent, non-adjacent, and U.S.) traffic originating in one carrier's territory with (pre-settlement) revenues collected by the considered carrier. If one such test was not passed, it would imply that one set of customers does not pay for its incremental cost and consequently that the interregional rate structure is not subsidy-free.

ii) Second Series of Tests

Assuming all tests to be passed in the first series, the second series of tests would compare the incremental cost incurred by one carrier for all interregional traffic with post-settlement revenues collected by the said carrier. A positive test in this case would imply that the sharing scheme utilized discriminates against the considered carrier.

Given the results obtained in the course of this project, namely the relatively small magnitude of incremental costs when compared to total costs, it is probable that incremental cost tests as previously defined would not lead to any positive conclusion. Therefore, more stringent tests may have to be developed in the same line as those presented in section 4.2 (e.g. taking account of growth reserve or excess capacity).

Finally, once the prospective costs become an important part in the tariff determination, the whole question of reliability of forecasts has to be introduced in the concept of accountability of the carriers. What are the acceptable errors? Or rather, which are the best methodologies available for forecasting and planning purposes in telecommunications? Who must pay for the errors? These are questions which need a theoretical basis.

Reference

[1] Autin, C.; St-Cyr, G. Expansion sur plusieurs périodes. Note technique, projet N.P.P.S., troisième phase, avril 1975.

[2] Autin, C.; LeBlanc, G. "A National Telecommunications Planning and Policy Simulation Model", in Models and Decisions in National Economies, J.M.L. Janssen, L.F. Pau, A. Straszak (eds), North-Holland Publishing Company, 1979.

[3] Baumol, W.J.; Bradford, D.F. "Optimal Departures from Marginal Cost Pricing". Ann. Econ. Rev., vol. 60 (1970), p. 265-283.

[4] Faulhaber, G.R. On Subsidization: Some Observations and Tentative Conclusions. Conf. on Communication Policy Res., Office of Telecommunication Policy, Washington, 1972.

[5] Faulhaber, G.R. "Cross-subsidization: Pricing in Public Enterprises". Bell Lab. Econ., Discussion Paper, May 1975.

[6] Hazelwood, A. "Optimum Pricing as Applied to Telephone Service". Rev. Econ. Studies, vol. 18, no 2 (1950-51), p. 67-78.

[7] Laboratoire d'économétrie; Sorès Inc. Development and Empirical Evaluation of Cross-subsidy Tests. Unpublished DOC, March 31, 1977.

[8] Littlechild, S.C. "Marginal-cost Pricing with Joint Costs". Econ. Journal, vol. 80 (1970), p. 323-335.

[9] Littlechild, S.C. "A Game-theoretic Approach to Public Utility Pricing". Western Econ. Journal, vol. 8, no 2 (1970), p. 162-166.

[10] Loehman, E.; Whinston, A. "A New Theory of Pricing and Decision-making for Public Investment". The Bell Journal of Economics, vol. 1, no 2 (1971), p. 606-625.

[11] Loehman, E.; Whinston, A. "An Axiomatic Approach to Cost Allocation for Public Investment". Public Finance Quarterly, vol. 2, no 2 (1974), p. 236-251.

[12] Panzar, J.C.; Willig, R.D. "Free Entry and the Sustainability of Natural Monopoly." The Bell Journal of Economics, Spring 77, vol. 8, no 1.

[13] Shapley, L.S. "Cores of Convex Games". International Review of Game Theory (1971), p. 11-26.

[14] Zajac, E.E. Some Preliminary Thoughts on Subsidization. Conf. on Communication Policy Res., Office of Telecommunication Policy,, Washington, 1972.

ECONOMIC EVALUATION OF U.S. TELECOMMUNICATIONS POLICY PROPOSALS

J. H. Rohlfs
A. R. Goldstein
E. P. Marfisi

1.0 INTRODUCTION

The Telecommunication Policy Model (TELPOL) is a computer model designed to aid analysts as well as policy makers in formulating and evaluating changes in telecommunications policies. In designing TELOPL we have attempted to address the question of "Whose ox will be gored" under different telecmmunications policies. We have used available data to shed some insight into the economic impacts of U.S. telecommunications policy changes on elderly, poor and rural families served by the Bell System in the 48 contiguous states. The economic assumptions and model structure underlying TELPOL are documented elsewhere.[1] This paper illustrates the use of TELPOL by applying it to several current policy proposals including the reform of jurisdictional separations in the context of price regulation for Bell. Section 1.1 gives a general discussion of the issues. Section 1.2 discusses full-cost pricing. Section 1.3 estimates how reform of jurisdictional separations would affect telecommunication prices and quantities and estimates the aggregate impact on consumers. Section 1.3 also estimates how different socio-economic groups in the economy would be affected by the policy.

---

[1] Rohlfs, J.H., Goldstein, A.R. and Marfisi, E.P. "Evaluating Telecommunications Policies: Whose Ox Will Be Gored?" unpublished memo available on request from the authors.

1.1. USING TELPOL TO EVALUATE POLICY PROPOSALS:

We use TELPOL to analyze several widely discussed policies that are summarized in Table 1.1. At the national level, the policies we examine involve full-cost pricing (FCP) for Bell's interstate services and jurisdictional separations reform. FCP entails allocating non-attributable costs to services on the basis of regulated formulae.[1] Prices are then set so that revenues from each service equal the fully allocated costs of that service. We are emphatically not advancing FCP as the best, or even a satisfactory, method of regulating Bell's prices. Rather, we are attempting to understand the ramifications of FCP because it may have a central role in the future telecommunications industry.

At present, jurisdictional separations reform is receiving considerable attention by regulators and Congress. The debate focuses on whether it is desirable or even feasible to maintain the present jurisdictional allocation of non-traffic sensitive interstate costs (NTS)[2] as the industry becomes more competitive.

---

[1]The specific rules employed for allocating non-attributable costs in TELPOL are discussed below.

[2]As set forth in Separations Manual: Standard Procedures for Separating Telephone Property Costs, Revenues, Expenses, Taxes and Reserves (NARUC, 1971). It should be noted that under current groundrules, the NTS interstate cost assignment is scheduled to increase quite substantially over the next decade.

TABLE 1.1 THE TELECOMMUNICATIONS POLICIES EXAMINED

<u>FEDERAL POLICIES</u>:

. BOTH FULL COST PRICING FOR BELL'S INTERSTATE SERVICES

. AND FULL COST PRICING COMBINED WITH JURISDICTIONAL SEPARATIONS REFORM FOR BELL AND OCC'S

<u>STATE REGULATORS' RESPONSE TO THE ADDITIONAL REVENUE REQUIREMENT RESULTING FROM SEPARATIONS REFORM</u>:

. EITHER PROPORTIONATE INCREASES IN ALL LOCAL EXCHANGE CHARGES

. OR PROPORTIONATE INCREASES IN LOCAL CALLING CHARGES

We analyze the effects of substantially reducing the assignment of NTS interstate costs for Bell and OCC's. The policy for separations reform that we analyze entails a 30 percent reduction in the assignment of NTS interstate costs. This particular type of jurisdictional separations reform allows us to illustrate how TELPOL can be used to evaluate proposals for separations reform, and has no other significance. As Table 1.1 makes plain, we analyze the implications of separations reform in the context of FCP for Bell's interstate services.

Reform of the jurisdictional separations process will have important financial effects at the local and state levels. Many proposals for separations reform involve shifting revenue requirements from the interstate jurisdiction to state jurisdictions. The method state regulators choose to raise the additional funds will have great impact on the overall success of the reform.

As shown in Table 1.1, we analyze two of the possible responses which state regulators might make to the reform of separations. The first approach increases the charges for all local exchange services, while the second increases local calling rates. As our analysis demonstrates, the two methods have vastly different socioeconomic effects. Indeed, one informative lesson to be learned from this exercise of TELPOL is that the overall success of the policy for separations reform may be largely dependent upon which approach state regulators adopt.

We next consider the restructure of interstate rates which corresponds to FCP for Bell's interstate services. There are numerous FCP methods available for allocating joint and common costs to specific services. The allocation formula we use is much in the spirit of Fully Distributed Cost Method 7. We allocate non-attributable interstate costs to interstate services in proportion to each service's directly-attributable costs.

In Section 1.3, we turn our attention to separations reform. We focus on measuring the impact that pricing decisions made at the state level may have on the overall effectiveness of separations reform. In each case the measure of economic impacts that we employ is "consumer surplus".[1] This is the tool most economists use to arrive at a scalar measure of the economic benefit or damage a customer incurs as the result of a complex, vector-valued price change. If the policies we are evaluating result in price increases for all telecommunication services, then we can be confident that consumers are made worse off by those policies without resorting to consumer surplus. Similarly, if the policies result in price decreases for all telecommunication services we can be sure that customers are better off. However, we are actually considering policies that result in complex price changes, some increases and some decreases. It is in evaluating complex price changes that the

---

[1]See, for example, Mansfield, Edwin, Microeconomics: Theory and Applications, third edition (New York: W.W. Norton, 1979), page 94, for an elementary discussion of consumer surplus.

widely-accepted tool of consumer surplus is especially useful. The change in consumer surplus stemming from a price change is a scalar measure of the dollar value customers attach to that price change.

TELPOL calculates changes in consumer surplus at two levels of aggregation: First, TELPOL calculates changes in aggregate consumer surplus for Bell and OCC customers. This is the simple arithmetic sum of consumer surplus changes for individuals. On the basis of this aggregate estimate, the analyst can determine whether or not those who gain consumer surplus from the policy change gain more in aggregate than the losses incurred on the part of those who suffer economic harm as a result of the policy change. Frequently, however, policy makers are concerned more with the impact of their decisions on particular groups than with aggregate consequences. For this reason we also provide rough estimates of the impact of policy changes in the consumer surplus for selected, policy-relevant customer groups including the elderly, poor and rural households.

1.2 RESTRUCTURING INTERSTATE RATES IN ACCORDANCE WITH FULL COST PRICING FOR BELL

Table 1.2 shows the rates for Bell's major interstate services in 1978 (row 1). It also shows TELPOL's estimates of equilibrium rates under FCP. In interpreting these results, we should recognize that Bell's market for interstate day messages is primarily a business market, whereas the market for non-day is predominantly residential.[1] Furthermore, Bell's smaller business customers purchase more MTS than WATS or PL services.

As Table 1.2 makes plain, FCP would mean a 12 percent decrease in Bell's interstate day rates combined with a 22 percent increase in Bell's interstate non-day rates. By and large, this restructure involves a rate increase for Bell's residential customers combined with a rate decrease for Bell's smaller business customers - i.e., business customers who use MTS exclusively. The price increase for Full Business Day (FBD) WATS and the price reduction for Measured Time (MT) WATS result in little price change for Bell's largest business customers and price reductions for Bell's medium business customers - i.e., business customers who use MT WATS predominantly. Roughly speaking, the move from current rates to rates based on FCP would increase prices for Bell's residential customers, and decrease prices for Bell's business customers.

---

[1] Specifically, in 1978, 67 percent of Bell's daytime interstate MTS service was billed to business customers, while 92 percent of Bell's non-day MTS was billed to residential customers. See MTS Statistics: 1978 Supplement (Bedminster, N.J.: AT&T Long Lines, 1979).

Table 1.2 PRELIMINARY ESTIMATES OF BELL'S EQUILIBRIUM INTERSTATE PRICES UNDER FULL COST PRICING REGULATION

	Day MTS (¢/min.)*	Non-Day MTS (¢/min.)	FBD WATS ($/line/mo.)	MT WATS ($/line/mo.)	PL ($/line/mo.)
Current	33	18	1500	600	170
Full-Cost Pricing	29	22	1600	500	170

*1978 dollars

Upon consideration, these results are not surprising. Bell's current rate structure evolved while regulators endeavored to keep residential rates low. Consequently, realigning these rates to reflect fully-distributed costs results in price increases for residential customers.

1.3 EVALUATING THE ECONOMIC EFFECTS OF SEPARATIONS REFORM

1.3.0 Aggregate Welfare Analysis

We next turn our attention to measuring the changes in consumer surplus stemming from the kind of separations reform described previously. The baseline policy for this analysis involves FCP for Bell's interstate services and actual 1978 levels of the interstate assignment of NTS costs per minute of Bell's services. In addition, our scenario presumes that OCC's pay the same assignment of NTS costs that Bell pays per minute and have costs that are identical to Bell's. TELPOL users may have different forecasts of OCC access charges and other costs and can employ whatever data they deem appropriate. In any event, our assumptions enable us to show how to use the model.

As we turn to our discussion of results, we caution the reader to beware of using TELPOL to search for "the numerical answer." A most valuable aspect of the computer tool is to facilitate sensitivity analyses. With the uncertainty surrounding our estimates of demands and costs in the future, it is hopelessly naive to look to computer tools for "the numerical answer." TELPOL's strength is that it enables the user to evaluate policies under a wide range of assumptions regarding model parameters. And, as we turn to our results, the real question is whether our policy conclusions are robust with respect to the model parameters. We examine this issue further below.

Relative to our baseline scenario, we estimate that aggregate consumer surplus would increase $400 m/yr if the assignment of NTS interstate costs were reduced by 30 percent, for both Bell and OCC's. This estimate relies on two important assumptions. One, we assume that there would be no "network externality".[1] Two, we assume that the additional intrastate revenue requirement would be met through a proportionate increase in all local exchange charges--whether usage-sensitive or not. We discuss each of these assumptions in turn.

In part, the value to an individual of being connected to the public network depends on the number of other subscribers. In general, we would expect that the fewer the subscribers, the lower

---

[1]The network externality is discussed extensively in Rohlfs, J., Economically-Efficient Bell-System Pricing (Bell Laboratories Economic Discussion Paper #138, January 1979) and in "A Theory of Interdependent Demand for a Communications Service," Bell Journal of Economics and Management Science, Spring 1974. Also see Squire, L., "Some Aspects of Optimal Pricing for Telecommunications," Bell Journal of Economics and Management Science, Autumn 1973, Artle, R. and Averous, C., "The Telephone System As a Public Good-Static or Dynamic Aspects," Bell Journal of Economics and Management Science, Spring, 1973. Also see Littlechild, S. C., "Two Part Tariffs and Consumption Externalities," Bell Journal of Economics and Management Science, volume 6, number 2, 1975.

the value of a subscription to an individual. However, individuals confronted with a price increase for local exchange service -- as may be the case with separations reform -- may decide to drop their subscription without considering the effect of their decision on other subscribers. An individual's decision can therefore affect other individuals in a way that is not reflected in prices. This phenomenon is called the network externality. The $400 M per year estimated increase in aggregate consumer surplus is based on the assumption that the network externality does not exist. The implications of making different assumptions regarding the magnitude of the network externality will be examined later when we discuss our sensitivity analysis.

As was mentioned earlier, many types of jurisdictional separations reform would increase revenue requirements in state jurisdictions. These monetary requirements could amount to billions of dollars. To supply these additional funds, state regulators must raise intrastate prices. The issue of which prices are raised is a matter of considerable importance. The results for aggregate consumer surplus stated above assume that state regulators respond with proportionate increases in all local exchange rates. As will be discussed, should state regulators respond with increases in local calling rates rather than increases in basic monthly local exchange rates, these results would be substantially improved.

Can we quantify the probable effect of the network externality on our results for aggregate consumer surplus? Is it possible to establish the magnitude of network externality? Unfortunately, it is extraordinarily difficult to measure such externalities. To date, these values have not been established. However, we can estimate plausible bounds for the network externality by making certain limiting assumptions.

At the lower extreme, we assume a complete absence - or complete internalization - of any network externality. This assumption leads to the results we have cited. At the opposite extreme, we assume that two parties place the same value on a communications link between them. Therefore, the value that any particular individual attaches to being connected to the network is exactly one-half of the value the entire body of subscribers, including the new subscribers, attaches to that connection. This assumption represents a plausible upper limit on the effect of the access externality because it ignores all possibilities that exist for internalizing the externality. In fact, ample opportunities exist for partially internalizing the network externality. For example, let us consider businesses that subscribe to the telecommunications network in order to accommodate their customers. These firms are probably able to pass at least part of their telecommunications costs onto their customers by charging higher prices. Hence, the beneficial effect of the business' use of telecommunications is reflected in the prices customers pay for non-telecommunication goods and services.

By raising the local calling rates instead of the monthly local exchange rates, state regulators could satisfy the demand for added revenue stemming from separations reform. Such a policy would also reduce and perhaps eliminate any adverse effects stemming from the network externality, and enhance aggregate surplus well above levels that could be achieved through changes in local exchange rates. Our baseline results are that an added $300 M/year in aggregate consumer surplus will result if the intrastate deficit were recovered through increases in local usage charges rather than increases in local exchange rates.

To this point our policy analysis suggests that a reduction in the NTS interstate cost assignment improves aggregate surplus. Further, our analysis indicates that if state regulators meet the additional revenue requirements stemming from jurisdictional separation reform by increasing local calling rates instead of basic service rates, surplus is improved by our additional amount equal to hundreds of millions of dollars per year. The question we now address is, "Are these conclusions robust with respect to our estimates of the model parameters?"

To analyze this question we conduct sensitivity analyses with respect to our estimates of the industry demand elasticities, the OCC demand parameters, the relative economic costs of Bell and OCC's, and the access externality. In particular we consider high and low estimates of the model parameters in each of the dimensions. The range for the high and low estimates generally involve 50 percent changes in the variables of interest. In an effort to put concrete bounds on our estimates of the surplus change stemming from jurisdictional separations reform we consider contributions of high and low parameter estimates in each of these dimensions simultaneously. Thus, for example, we consider how our estimates of the increase in surplus would change if the industry demand elasticities were much higher than our baseline estimates and simultaneously OCC demand parameters were much lower. With a single exception we find that the basic conclusions outlined in the preceeding paragraphs are extremely robust with respect to the model parameters. The one exception concerns the access externality.

By way of illustration we find that if the magnitude of the access externality is negligible, then under the local exchange pricing option:

- A 50 percent change in the baseline estimates of the industry demand elasticities results in a change of ± $200 M/yr in our baseline estimate of the surplus change stemming from jurisdictional separations reform.

- Substantial changes in the OCC demand parameters (e.g. 50 percent changes in the discount OCC's must offer to induce customers to switch from Bell to OCC's) result in ± $200 M/yr in our baseline estimates of the surplus change stemming from jurisdictional separations reform.

In all cases considered jurisdictional separations reform increases aggregate consumers surplus even if state regulators raise basic local exchange rates in response to the reform, provided the network externality is negligible. Additionally, we find that the consumer surplus gains are comparable under the local exchange and local use pricing options, again, under the assumption that the access externality is negligible.

If the access externality is important in the sense that its magnitude approaches the upper limit discussed before, these results change remarkably. Our sensitivity analysis indicates that under the local exchange pricing option for state regulators, the access externality could eliminate the gains stemming from jurisdictional separations reform. In fact under some scenarios separations reform would reduce aggregate surplus if state regulators increase local exchange rates.

On the other hand if state regulators respond to separations reform with increases in local calling rates, welfare losses which might be exacerbated by the access externality under the local exchange pricing option are largely avoided. In all cases, our sensitivity analysis indicates that in the presence of the access externality aggregate surplus is improved if state regulators increase charges for local calling rather than raise basic local exchange rates. Further, we find that if state regulators raise local calling rates, separations reform always improves aggregate consumer surplus regardless of the magnitude of the access externality. This stands in sharp contrast to our finding that jurisdictional separations reform could reduce aggregate surplus under the local exchange pricing option.

This section has used TELPOL to evaluate some illustrative policy proposals, to conduct sensitivity analyses on aggregate results, and to suggest policies that may be socially desirable from the standpoint of aggregate consumer surplus. As was mentioned earlier, changes in policy can improve aggregate consumer surplus but still affect some individuals adversely. There will be groups of consumers that gain as a result of the new policies and others that lose. Frequently, the policy decision hinges more on the identity of the groups that are adversely affected than on changes in aggregate economic well-being. The next section addresses the distributional effects of telecommunications policy changes.

1.3.1 IMPACT ON DIFFERENT SOCIO-ECONOMIC GROUPS

1.3.1A Introduction

This section considers how policies affect different socio-economic groups. That is, we address the question of "whose ox gets gored".

We consider the following three groups in this analysis: The poor, the rural, and senior citizens. For example, the change in consumer surplus of an average poor household is compared to the change in consumer surplus of an average non-poor household. This is known as a binary comparison. Similarly, we compare rural to non-rural, and senior to non-senior.

It is important to note that these classifications may overlap. For example a poor household can also be rural. Thus, we cannot draw meaningful conclusions by comparing the change in surplus of poor households against the change in surplus of non-rural households. We can only compare groups in the same category; so that the poor are compared only to the non-poor, the rich only to the non-rich, and the rural only to the non-rural.

We have used the following definitions of consumer groups. A household is considered poor if it has an income of less than $7,500 in 1978 dollars. In a senior household the head of the household is over 65 years old. Our definition of rural is essentially the same as the one used by the Bureau of Labor Statistics in their Consumer Expenditure Survey 1972-3.

The change in consumer surplus of a particular group consists of two components: a <u>direct impact</u> and an <u>indirect impact</u>.

The direct impact is the change in welfare brought about by changes in the prices of telecommunications services used directly by households, such as local calls, MTS-day etc. As a result of such price changes, a household's telephone bill will rise or fall, making the household better or worse off.

This is not, however, the only result of changes in telecommunications prices. Businesses will also face new prices for the services they use, such as private line, WATS, etc. This affects the prices of inputs that firms use to produce their outputs. It is critical to observe that changes in these input prices are not "absorbed" by firms. A business confronted with an increase in the price of an input must either increase the price it charges customers for its output, or use some of its profits to pay for the added cost of the input. In either case, the changes in input prices are flowed through to people.

In general, we would expect that changes in telecommunication prices would cause output prices, as well as the net earnings of businesses, to change. Empirically, we would expect the proportionate change in output prices to be relatively small because telecommunications expenditures account for so small a part of the total costs of businesses - often only one or two percent. In addition, the extent to which businesses are affected depends on the ease with which they can substitute other services for telecommunications.

Thus, the indirect or the <u>flowthrough</u> impact of the change in telecommunications prices consists of two subcomponents. Firstly, there is the <u>price-effect</u>; e.g., the price of an airline ticket may rise as the price of a Full Business Day WATS line rises. Secondly, there is the <u>net-earning-flowthrough</u> effect; e.g., profits of the airline industry may decline.

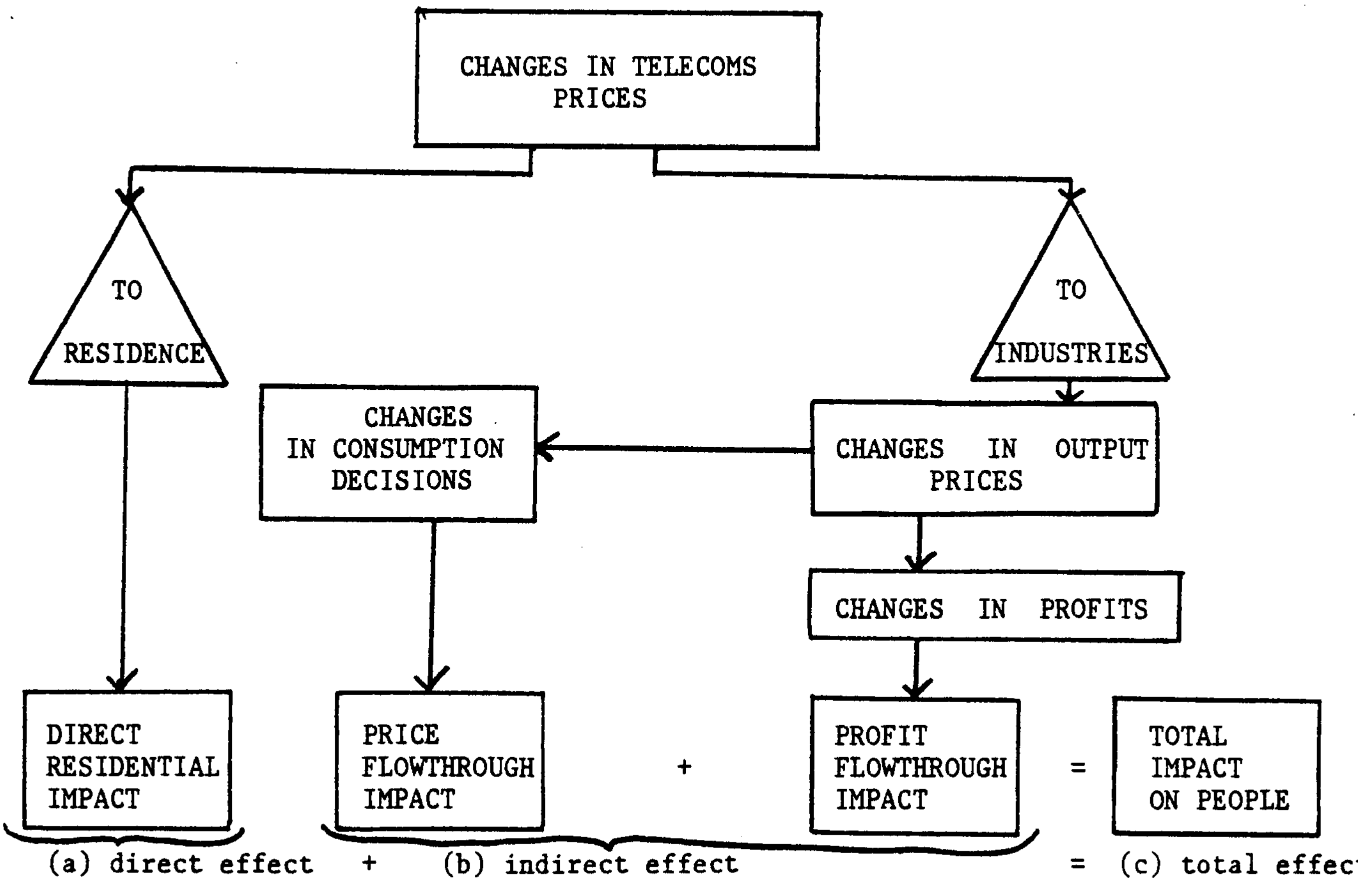
Additionally, the change in a business' profits affects the people who own the business. Increases in profits may result in increased dividends to shareholders, or may be retained by the firm. In the latter case, stockholders are likely to experience a gain in their equity.

These concepts and the interrelationships among them are shown schematically in figure 1.1.

The next two sections discuss this analysis in more detail. Section 1.3.2 considers the analysis of indirect impacts, with the BFM, or Business Flowthrough Model. Section 1.3.3 analyzes the direct impact on households, in a relatively simple application of microeconomic principles.

Section 1.3.4 presents some preliminary results of one of the policies examines earlier; namely jurisdictional separations reform (from a full cost pricing baseline).

FIGURE 1.1: A SCHEMMATIC VIEW OF THE TOTAL EFFECT OF CHANGES IN THE PRICES OF TELECOMMUNICATIONS SERVICES.



1.3.2. THE BUSINESS FLOWTHROUGH MODEL: THE ANALYSIS OF INDIRECT IMPACTS

1.3.2.A. Preliminaries

In gauging the total distributive effect of policies, we must account for the flowthrough impact. Previous analyses of distributive effects have not considered the question of indirect effects at all, or have done so in only a rudimentary way.[1] The indirect effects are, however, significant, and the Business Flowthrough Model (or BFM) estimates them.

1.3.2.B. A general description of the Business Flowthrough Model

The BFM is a 27 industry input-output model of the US economy.[2] Our input-output model has a number of features which are not usually associated with standard input-output models. Specifically, we allow each industry to respond flexibly to changes in telecommunication prices by substituting telecommunications services for labor and capital. The ease with which business sectors can substitute capital and labor for telecommunications, is known as the elasticity of substitution. In our model, this elasticity is controlled by the user.

---

[1] For example, see Appendix A of the Charles River Associates Report The Economics of Competition in the Telecommunications Industry, August 1979.

[2] For an introduction to input-output models see Chenery, H., and P. Clark, Interindustry Economic, Wiley, 1967.

Our analysis also attempts to construct a model of the market structure of the various industries. This point is important and requires some elaboration. Standard economic theory assumes that all markets are perfectly competitive.[1] Since there are no economic profits in such a model, all changes in the prices of inputs will be fully reflected by changes in the prices of outputs and economic rents.

In reality, the economy is not perfectly competitive, and this has important implications for our analysis. First, the way industries adjust their prices in response to a change in the price of an input depends on how the market is structured. Second, if we are interested in the impact on profits of changes in the prices of telecommunications, we must clearly have a model that allows profits to exist.

The model also allows a firm's expenditures on telecommunications to be treated as fixed expenditures: costs that are not affected by changes in the levels of outputs. Alternatively, the amount spent on telecommunications may be dependent upon the level of output. With this model, we can vary the proportion of telecommunication costs that are fixed from zero to one hundred percent.

---

[1]Perfect competition is an often employed assumption in economic analysis. It means that all market participants are 'small' relative to the market so that no single unit can noticably affect prices. Additional, it also requires that no barriers to entry and exit exist. This implies that in the long run equilibrium economic profits will be zero in all industries.

This aspect of our analysis of telecommunications expenditures has the following implications. For instance, assume that all telecommunication expenditures are fixed. As we use what is essentially variable-cost pricing,[1] increasing the price of telecommunications will not affect output prices under this assumption. In this scenario changes in profits are thus the only response to changes in telecommunications prices. To the extent that the distribution of consumption of outputs among households is different from the distribution of business ownership among households, distinction between fixed and variable telecommunications costs will affect our results.

1.3.2.C. How the Model works

In this section we describe in general terms how the BFM operates. The BFM maps telecommunications price changes into equilibrium price changes and output changes for all the industries in our model.

The model starts from an initial equilibrium of baseline telecommunications prices. This initial or benchmark equilibrium is then upset by a change in an index of business telecommunications prices. This change in the price of telecommunications leads industries to revise their prices and their methods of producing goods by combining inputs in a least cost fashion. New output prices cause con-

---

[1]This is essentially marginal cost pricing, but allows the existence of profits.

sumers to buy differently. Furthermore, any industry which uses the outputs of other industries as inputs, will now be faced with a general rise in the costs of these inputs. This price increase will result in further revisions of production and pricing decisions. This process, if left undisturbed by additional changes, will end with a new set of output prices for all industries, referred to as long run equilibrium.

Once new equilibrium prices are computed, other variables such as industry sales, costs, and profits in the new equilibrium are easily determined. At that point we have the information necessary to compute the changes in consumers' surplus caused by indirect effects on the various demographic groups.

1.3.2.D. The data used by the model

The BFM is a mini-model of the US economy. This model requires a great deal of data. We have used five primary data sources. The input-output data was constructed from the INFORUM database at the University of Maryland. That database is based on the input-output accounts published by the Bureau of Economic Analysis at the US Department of Commerce, for 1972. The data on consumption expenditures by demographic groups is based on the Consumers' Expenditure Survey of 1972-3, conducted by the US Department of Labor. Data on factor payments (capital and labor expenditures) are taken from the Gross Product Originating Accounts developed by the Bureau of

Economic Analysis at the US Department of Commerce. We also make use of data on capital-output ratios by industry developed by John Kendrick at the Conference Board - Division of Economic Research. Lastly, we use some consumption elasticities developed by Clopper Almon at the University of Maryland.

In addition to the data described above, the model requires the user to supply information on the cost of capital, the elasticity of substitution between telecommunications and capital and labor and the proportion of telecommunications expenditures that does not depend on the level of output.

One set of data that we would like to have has been unobtainable: the distribution of ownership of industry by our consumers' groups. Once we compute the total change in profit in the economy, we need information on how the ownership of industry is distributed among our defined consumer groups in order to assign changes in profit over people. Unfortunately no reliable data appear to exist for age and location, and only incomplete data exists on the distribution of ownership by income. Faced with this lack of information, we have resorted to certain bounding assumptions in place of these data. We have chosen assumptions which we consider conservative. These assumptions bias our result against the targeted groups.

1.3.3 ANALYSIS OF THE DIRECT RESIDENTIAL IMPACT

1.3.3.A. Methodology

The changes in surplus, caused by the direct effect on households, are much simpler to compute than the indirect effect. We begin our analysis of the direct effect by assuming that households use three telecommunications services: denoted in TELPOL as services 1, 2, and 3. (Urban household also consume services 7 and 8-OCC Day and Non-Day message services). Thus, given price changes for the three services, consumer surpluses for each of our specified groups can be computed by using data on local and long distance usage by demographic characteristics.

As with BFM, some critical data were not immediately available, in particular local usage by demographic groups. Consequently, we again resorted to the conservative technique of limiting assumptions. Specifically, some of the policies considered in Section 1.2 involve an increase in the price of local calls and a decrease in the price of long distance calling. It is clear that the distribution of local calling by the various demographic groups will be critical in assessing the direct impact. The issue then becomes the following: what is a reasonable limiting assumption regarding the distribution of local calls by demographic groups. We believe that the assumption of equal local usage per household is a conservative one. The fragmentary empirical evidence on this issue suggests that local calling increases with income. To the extent that this is true, assuming that local calling is the same for all groups will overstate the amount of local calling done by poor households. When the price of locals calls in our analysis is increased, our assumption will cause our model to over-estimate the extent to which poor households will be adversely affected by this price increase.

1.3.3.B. THE DATA USED IN THE ANALYSIS OF DIRECT EFFECTS

The data on long distance usage (by demographics) used in our analysis has been derived from the Bell System MARC data base (Marketing Analysis of Revenues and Customers). In addition, estimates of the elasticities of the two long distance services were used. These elasticities are the ones used in the Telecommunications Industry Pricing Model. Further demographic information on rural customers was obtained from the Consumers' Expenditure Survey 1972-3.

1.3.4. MODEL APPLICATION

This section applies our technique to a policy discussed in Section 1.2: jurisdictional separations reform combined with increases in local calling charges in the context of full-cost pricing for Bell's interstate services.

The follow limiting assumptions were used in our analysis:

1) our consumer groups made the same number of local calls per household as the national average.

2) Stock ownership was distributed among our consumers in the following fashion.

- In the poor vs. non-poor comparison we assume that the poor do not own any equity.

- In the rural vs. non-rural comparison we assume that rural households do not own any equity.

- In the senior vs. non-senior, we assume that senior households have same amount of equity as the rest of the population.

We believe these assumptions to be conservative. To the extent that targeted groups make local calls at levels below the national average, they will not be as adversly affected as our analysis suggests. Similarly, to the extent that the poor and rural groups own equity, they will do better than our model indicates. This is because profits generally increase when telephone rates to businesses are reduced.

By assuming that the poor and rural do not own any equity, we rule out the potential benefits they may receive from increased profits. With the senior group, assuming that they own the same amount of equity as the rest of the population also understates the effects of changes in business profits. This is because the elderly tend to have higher than average share in business profits.

Table 1.4 shows the changes in consumer surplus, in 1978 dollars, for each of the binary comparisons (rounded to the nearest multiple of $5.)

The numbers shown are total surplus changes, including direct and indirect effects in consumers' surplus for the <u>average household in the group</u>. As was previously discussed, only one on one comparisons are meaningful in this context. (We cannot, for example, conclude that the non-rural are less favorably affected than the poor.)

From Table 1.4 we observe that <u>no average household</u> experiences a negative change in surplus. The average poor household is better off by approximately five 1978 dollars per year, while the non-poor household is ahead by over $10 per year. This is due to the fact that the non-poor tend to have a higher level of toll usage, and therefore have a more surplus originating in the reduction of toll prices. The same applies to the nonsenior households, who are $10 per year better off. Nonsenior households tend to have a higher toll usage than senior households. In the rural non-rural comparison, we observe that both groups appear to be similarly affected, in average. All targeted groups, the poor, rural, and the senior group, experience a net benefit. The counter groups tend to do better than the targeted groups, but no group is adversely affected. This result does not imply that every household benefits, only that the average household in every group benefits. Clearly, households that make only local calls will be adversely affected. From these results, we estimate that in terms of average household 'nobody's ox gets gored'.

Two additional runs, one yielding a high aggregate surplus and one yielding a low aggregate surplus, have been conducted to gauge the sensitivity of these results. Based on these two sensitivity runs it would appear that qualitative nature of our results is robust.

What would our intuition lead us to expect in the way of quantitative results? The change in telephone rates has a simple interpretation. Since local rates increase and long distance rates fall, this effect will be negative for a household that uses only local service and positive for a household that makes only long distance calls. In most cases, households make both kinds of calls. Thus, the extent to which the direct impact on such consumers is negative or positive will depend upon the relative proportion of local vs. long distance.

The indirect effect has a different interpretation. The first order price-flowthrough effects are clear. People who use more services will get more price-flowthrough benefit than people who use less. If rich households simple consume more of the same goods as poor people, they will derive a greater dollar benefit from a general fall in prices.

A second order effect occurs when consumers buy different goods in different proportions. Some groups may buy more goods whose production requires more use of telecommunication services than others. For example, poor households may spend most of their income on food and other basic necessities, whose production does not use much telecommunications. They therefore benefit less than households buying goods which more telecommunications, such as brokerage services or airline tickets. As it takes more use of telecommunications to produce these goods, their prices will tend to be more sensitive to changes in telephone prices than the prices of basic necessities.

In summary, our analysis shows that jurisdictional separations reform when combined with increases in local calling rates is a policy proposal which on average has no adverse distributive consequences. The overall gain in economic efficiency allows all groups to benefit in spite of the redistributional factors that favor the counter groups.

TABLE 1.4

PRELIMINARY ESTIMATES OF THE TOTAL DISTRIBUTIVE EFFECTS UNDER LIMITING ASSUMPTIONS. (See SECTIONS 1.3.2.D and 1.3.4)

UNITS: CHANGE CONSUMERS' SURPLUS
(1978 $, PER HOUSEHOLD, PER YEAR)

POOR	5	NON-POOR	10
RURAL	10	NON-RURAL	10
SENIOR	5	NON-SENIOR	10

NEW SERVICES

RENTABILITE ECONOMIQUE DU TRAFIC DES NOUVEAUX SERVICES SUR LES RESEAUX DE TELECOMMUNICATION

JEAN DE LA BRUNETIERE

Direction Générale des Télécommunications

P.T.T. France

1. INTRODUCTION

Le développement prochain des nouveaux services de télécommunication engendrera un trafic additionnel dont l'écoulement nécessitera d'importants investissements de réseau. Etudier la rentabilité de ces investissements est nécessaire pour pouvoir les justifier à la fois vis à vis des organismes qui contribuent à leur financement, et vis à vis des institutions de tutelle de l'Administration des Télécommunications.

Or une part importante de ces services nouveaux s'adressera à un public uniquement raccordé au réseau téléphonique, ou n'accèdant à d'autres réseaux (réseau de commutation par paquet TRANSPAC, par exemple) que par l'intermédiaire du réseau téléphonique. De ce fait le réseau téléphonique supportera une proportion importante du trafic des nouveaux services et sa capacité devra être accrue en conséquence.

Les investissements nécessités par l'écoulement d'un trafic additionnel sur le réseau téléphonique dépendent de la part de ce trafic qui s'écoule à l'heure chargée du réseau. A volume de trafic additionnel donné, et donc à recettes données, le montant des investissements de réseau nécessaires sera d'autant plus faible, et donc leur rentabilité d'autant plus forte, que la part de ce trafic s'écoulant à l'heure chargée du réseau sera plus faible.

On peut donc penser à priori que le trafic des nouveaux services s'écoulant sur le réseau téléphonique, aura une rentabilité importante puisqu'il n'aura probablement pas la même répartition horaire que le trafic téléphonique qui, restant prépondérant, définira l'heure chargée. Néanmoins, on conçoit qu'à mesure que le volume du trafic des nouveaux services augmentera, celui-ci jouera un rôle plus important dans la définition de l'heure chargée et que sa rentabilité risque d'en être diminuée.

Par ailleurs, le trafic des nouveaux services aura en général deux origines. Il proviendra pour une part d'une substitution à certaines communications téléphoniques actuelles, et pour une autre part, de l'apparition de communications supplémentaires induites, qui n'avaient pas d'équivalent auparavant sur le réseau téléphonique. Or cette seconde catégorie de communications est vraisemblablement assez sensible aux actions de promotion commerciale qui pourront être entreprises par l'Administration des Télécommunications. Il est donc utile de savoir dans quelle mesure des actions de stimulation du trafic induit par les nouveaux services sur le réseau téléphonique peuvent être opportunes pour accroître la rentabilité des investissements réalisés.

Nous cherchons dans ce papier à répondre aux questions précédentes en étudiant quelle est la rentabilité d'un trafic de nouveaux services s'écoulant sur le réseau téléphonique et comment elle varie en fonction du volume de ce trafic.

Pour y parvenir, nous adopterons d'abord un point de vue statique, en examinant le cas d'un trafic de nouveaux services constant dans le temps, et sans faire d'hypothèse restrictive sur l'évolution de la courbe de charge du réseau en fonction du volume de trafic des nouveaux services. Ainsi, on supposera dans un premier temps que l'heure chargée du trafic total du réseau téléphonique peut varier de façon continue en fonction du volume de trafic des nouveaux services ; puis on envisagera le cas où elle ne se modifie que de façon discrète, par basculements successifs.

Nous examinerons ensuite la rentabilité d'un échéancier de trafic de nouveaux services, à deux points de vue :

-tout d'abord, on étudiera comment varie, dans le cas général, en fonction du volume de trafic des nouveaux services, la rentabilité des investissements marginaux successifs associés à l'accroissement de capacité du réseau téléphonique, et on établira une relation reliant cette rentabilité marginale à l'échéancier de trafic de nouveaux services retenu.

-dans un deuxième temps, on examinera comment varie la rentabilité globale des investissements associés à un échéancier de trafic de nouveaux services, en fonction de celui-ci. Pour pouvoir mener les calculs à leur terme, nous seron conduits à particulariser la forme des échéanciers de trafic téléphonique et de nouveaux services (en leur supposant une croissance annuelle à taux constant) et à admettre que l'heure chargée du réseau ne varie que de façon discrète. On pourra alors déterminer l'échéancier maximum de trafic de nouveaux services compatible avec un taux de rentabilité donnée, et définir, dans certains cas, un échéancier de trafic nouveau optimum du point de vue de la rentabilité des investissements associés.

Enfin, deux exemples d'application des résultats obtenus seront fournis :

- le premier exemple concerne la détermination d'un échéancier de trafic maximum compatible avec un taux de rentabilité marginale donné. Ce problème sera résolu dans le cas théorique où on peut spécifier les courbes de charge de trafic téléphonique et de trafic de nouveaux services sous forme de fonctions continues.

- le second exemple concerne la rentabilité globale des investissements de trafic de nouveaux services dans le cas du vidéotex. On adoptera ici des répartitions horaires empiriques pour le trafic téléphonique et le trafic de nouveaux services et on déterminera les échéanciers maximum de trafic vidéotex compatibles avec une rentabilité donnée des investissements de réseau.

2 - Rentabilité économique d'un trafic de nouveaux services constant.

On considère ici un réseau écoulant un trafic téléphonique de base auquel on superpose un trafic additionnel permanent dû aux services nouveaux, et on étudie la rentabilité des investissements de réseau nécessaires à l'écoulement de ce trafic additionnel. On établira d'abord la relation qui permet de calculer le taux de rentabilité du trafic des nouveaux services. Puis on étudiera comment varie ce taux en fonction du volume de ce trafic.

2.1. Calcul du taux de rentabilité d'un trafic de nouveaux services constant.

Par définition le taux de rentabilité d'un investissement est la valeur du taux d'actualisation qui en annule le bilan actualisé. Celui-ci s'obtient par différence entre les recettes et les coûts totaux actualisés.

2.1.1. Calcul des coûts actualisés

On admet ici que seul a un coût non nul le trafic s'écoulant sur le réseau à l'heure de pointe. On estimera donc d'abord la valeur du trafic de pointe engendré par le trafic additionnel. Puis on en déduira successivement celles des coûts actualisés d'investissement et de fonctionnement qui lui sont associés.

a) Estimation du trafic de pointe imputable au trafic additionnel

Soit :

n_0 = nombre d'erlangs-heure quotidiens de trafic téléphonique de base

n = nombre d'erlangs-heure quotidiens de trafic additionnel de nouveaux services

s_0 nombre d'erlangs-heure quotidiens du trafic téléphonique de base remplacé par du trafic des nouveaux services

$\rho_0(h)$ = répartition horaire du trafic de base ($\sum_h \rho_0(h) = 1$)

$\rho(h)$ = répartition horaire du trafic additionnel des nouveaux services ($\sum_h \rho(h) = 1$)

$T_1(h)$ = trafic horaire résultant (en erlangs)

On suppose que le trafic substitué a la même structure horaire que le trafic de base.

h_0 = heure de pointe du trafic de base

h_1 = heure de pointe du trafic résultant

T_ρ = trafic de pointe imputable au trafic des nouveaux services (en erlangs)

On a : $T_1(h) = (n_0 - s_0)\,\rho_0(h) + n\rho(h)$

D'où : $T_\rho = (n_0 - s_0)\,\rho_0(h_1) + n\rho(h_1) - n_0\,\rho_0(h_0)$

b) Coûts d'investissement de réseau :

Soit :

i = coût unitaire d'investissement par erlang

D = durée de vie du réseau

a = taux d'actualisation annuel

Le coût d'investissement total du trafic supplémentaire net actualisé à la date de mise en service est :

$$I(a) = T_\rho \cdot i \; \frac{(1+a)^D}{(1+a)^D - 1}$$

c) Coût de fonctionnement du réseau

Soit :

f = coût annuel de fonctionnement par erlang

Le coût de fonctionnement total actualisé du trafic supplémentaire net est :

$$F(a) = T_\rho . f \left[\frac{1}{1+a} + \frac{1}{(1+a)^2} + \dots \right] = T_\rho . \frac{f}{a}$$

2.1.2. Calcul des recettes actualisées

Soit :

J = nombre de jours moyens de trafic par an

Tx = valeur de la taxe de base

p(h) = période d'impulsion moyenne du trafic de nouveaux services (en sec.)

$p_o(h)$ = période d'impulsion moyenne du trafic téléphonique substitué (en sec.)

La période d'impulsion moyenne d'un trafic dépend de sa structure spatiale et de sa répartition horaire. On supposera ici que les structures spatiales du trafic de nouveaux services et du trafic téléphonique substitué sont les mêmes, ce qui est cohérent avec l'hypothèse faite implicitement ci-dessus que le coût de l'erlang est le même dans les deux cas.

La recette annuelle nette dûe au trafic de nouveaux services est donc :

$$A = 3600.J.Tx \left[n \left(\sum_h \frac{\rho(h)}{p(h)} \right) - n_o \left(\sum_h \frac{\rho_o(h)}{p_o(h)} \right) \right]$$

La recette totale actualisée correspondante est donc =

$$R(a) = \frac{A}{a}$$

2.1.3. Calcul du taux de rentabilité

a) Le taux de rentabilité est r tel que : B(r) = o

avec $$B(r) = \frac{A - T_p.f.}{r} - T_p.i \; \frac{(1+r)^D}{(1+r)^D - 1}$$

Posons $\varepsilon = \frac{T_p.i}{A - T_p.f}$, et admettons que $A > T_p.f$

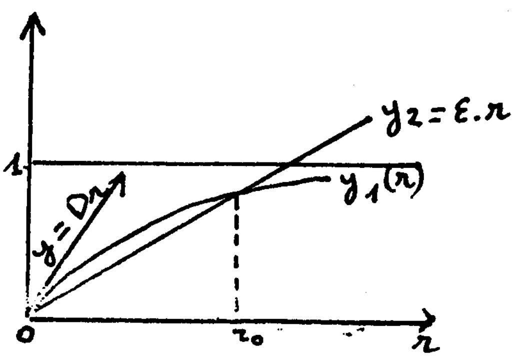
ε représente l'inverse du rapport :

$\frac{\text{recette annuelle nette}}{\text{investissement}}$ associé à un erlang

de trafic de nouveau service. Ce rapport est
parfois appelé "rentabilité immédiate" de l'investissement

Si l'on trace les courbes :

$y_1 = 1 - (1+r)^{-D}$
$y_2 = \varepsilon . r$



On voit que l'équation B(r)=o
n'a de solution ro # o que si :

$o \leq \varepsilon < D$

b) On peut écrire : $\varepsilon = \frac{1}{\frac{A}{T_p.i} - \frac{f}{i}}$. Or $\frac{f}{i}$ représente le

rapport entre les dépenses de fonctionnement annuel et les dépenses d'investissement correspondant à une augmentation de capacité du réseau. Ce rapport caractérise le réseau, indépendamment du trafic qu'il écoule. A réseau donné, ε ne dépend donc que de $\frac{A}{T_p.i}$ dont il est une fonction décroissante

Par conséquent, à réseau donné (caractérisé par D et $\frac{f}{i}$) le taux de rentabilité d'un trafic supplémentaire est une fonction croissante de $\frac{A}{T_p \cdot i}$.

Pour savoir comment cette rentabilité varie en fonction du volume de trafic, il suffit d'examiner comment varie $\frac{A}{T_p \cdot i}$ en fonction de celui-ci.

2.2. Variation de la rentabilité d'un trafic de nouveaux services constant en fonction de son volume

En général, la courbe de charge du trafic téléphonique de base et celle du trafic nouveau service ne sont pas les mêmes. Il en résulte que l'heure chargée finale du réseau dépend des volumes respectifs de trafic téléphonique et de trafic de nouveau service qu'il écoule. Lorsque le rapport entre ces volumes se modifie, l'heure chargée varie. Néanmoins, il peut arriver que, dans certains intervalles de variation du volume de trafic nouveau, l'heure chargée reste approximativement la même. Parfois même, si la courbe de charge du trafic de nouveau service est voisine de celle du trafic téléphonique, ou si le volume de trafic de nouveau service est faible par rapport à celui du trafic téléphonique, on pourra faire l'hypothèse que l'heure chargée reste celle du trafic téléphonique de base lorsque le volume de trafic de nouveau service varie.

On examinera donc ici d'abord le cas général où l'heure chargée varie en fonction du volume de trafic de nouveau service, puis celui où elle reste fixe, au moins par intervalles.

2.2.1. Cas général : l'heure chargée dépend du volume de trafic des nouveaux services

Posons : $\alpha = 3600 \ J.Tx \left(\sum_h \frac{\rho^{(h)}}{\mu^{(h)}} \right)$

$\beta = 3600 \ J.Tx \qquad \left(\sum_h \frac{\rho_0(h)}{\mu_0(h)} \right) = \beta$

$\gamma = i \, \rho(h_1)$

$\delta = i \quad \left[n_n \left(\rho_0(h_0) - \rho_0(h_1) \right) + s_n \, \rho_n(h_1) \right]$

α est la recette annuelle d'un erlang de trafic de nouveaux services et γ l'investissement correspondant
β représente la recette annuelle unitaire associée au trafic substitué , et δ le coût unitaire d'investissement de la capacité du réseau rendue disponible par erlang-heure de trafic substitué , compte tenu de la variation de l'heure de pointe.

On suppose $T_\rho > o$, c'est-à-dire $n > \frac{\delta s_o}{\gamma}$

On a = $\frac{A}{T_\rho . i} = \frac{n\alpha - \beta s_o}{n\gamma - \delta s_o}$

où = α et β ne dépendent pas de n
et où = γ et δ en dépendent par l'intermédiaire de l'heure de pointe h_1.

Le sens de variation de $\frac{A}{T_\rho . i}$ en fonction de n dépend du signe de :

$$\Delta = s_o(\beta\gamma - \alpha\delta) - (n\alpha - \beta s_o)\left[n\frac{d\gamma}{dn} - s_o\frac{d\delta}{dn}\right]$$

lorsque Δ a un sens, c'est-à-dire lorsque γ et δ sont dérivables par rapport à n.

Deux cas sont donc à distinguer :

a) γ et δ sont dérivables par rapport à n, ce qui suppose que h_1 est dérivable par rapport à n, (on admet que $\rho_o(h)$ et $\rho(h)$ sont dérivables par rapport à h)

L'hypothèse précédente revient à admettre qu'il n'y a pas de basculement discontinu de l'heure de pointe h_1 lorsque n augmente.
Dans ce cas :

$$n\frac{d\gamma}{dn} - s_o\frac{d\delta}{dn} = o \text{ car :}$$

$$\frac{d\gamma}{dn} = i\left(\frac{d\rho}{dh}\right)_{h_1}\frac{dh_1}{dn}$$

$$\frac{d\delta}{dn} = \frac{i}{s_o}(s_o - n_o)\left(\frac{d\rho_o}{dh}\right)_{h_1}\frac{dh_1}{dn}$$

et, par définitiion de l'heure chargée :

$$(n_0 - s_0)\left(\frac{d\rho_0}{dh}\right)_{h_1} + n\left(\frac{d\rho}{dh}\right)_{h_1} = o$$

Donc $\Delta = s_0\,(\beta\gamma - \alpha\delta)$

Il en résulte que le sens de variation de $\frac{A}{T_{p.i}}$ en fonction de n dépendra du signe de $(\beta\gamma - \alpha\delta)$, qui dépend lui-même de n par l'intermédiaire de γ et δ .

Donc :

- $\frac{A}{T_{p.i}}$ croîtra en fonction de n, si $\frac{\alpha}{\gamma} < \frac{\beta}{\delta}$, c'est-à-dire si la rentabilité immédiate du trafic de nouveau service est inférieure à celle du trafic substitué. Il en sera de même du taux de rentabilité (d'après 2.1.)

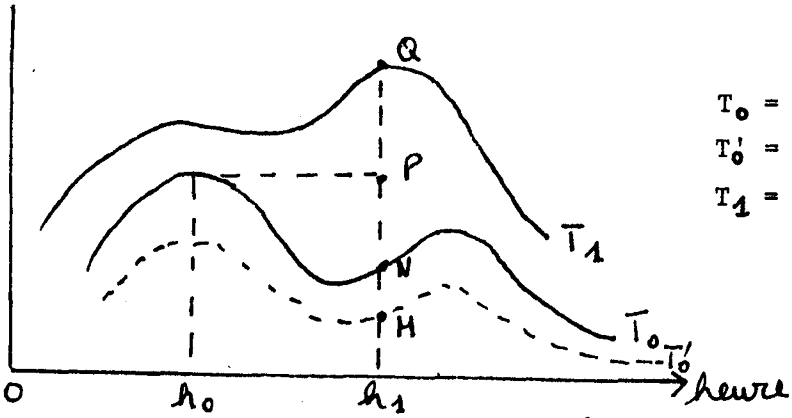
- $\frac{A}{T_{p.i}}$ décroîtra en fonction de n si $\frac{\alpha}{\gamma} > \frac{\beta}{\delta}$, c'est-à-dire si la rentabilité immédiate du trafic de nouveau service est supérieure à celle du trafic substitué. Il en sera de même pour le taux de rentabilité.

On peut fournir pour ces résultats la justification intuitive suivante :

- si le trafic perdu (=substitué) est plus rentable que le trafic additionnel, la perte relative de rentabilité, dûe à la substitution ,très sensible pour un trafic additionnel faible, le sera de moins en moins à mesure que l'importance du trafic additionnel augmentera. Donc le taux de rentabilité de ce trafic, compte tenu de la substitution associée, croîtra.
- si le trafic perdu est moins rentable que le trafic additionnel le gain de rentabilité sera d'autant moins sensible que le trafic additionnel sera plus important. Le taux de rentabilité décroîtra donc.

Remarquons que dans ce cas, le sens de variation de $\frac{A}{T_{p.i}}$ peut s'inverser si $(\beta\gamma - \alpha\delta)$ change de signe pour une valeur donnée de n.

Graphiquement, les divers trafics peuvent se représenter comme suit :



T_0 = trafic initial de base
T'_0 = trafic après substitution
T_1 = trafic résultant (= initial − substitué + additionnel)

Les rapports $\frac{\alpha}{\gamma}$ et $\frac{\beta}{\delta}$ s'écrivent alors respectivement :

$$\frac{n.\alpha}{i.HQ} \quad \text{et} \quad \frac{s_0\beta}{i.HP}$$

La condition ci-dessus s'exprime donc comme suit :
le taux de rentabilité du trafic additionnel, compte tenu des substitutions qu'il entraîne, sera une fonction croissante de son volume si $\frac{n\alpha}{s_0\beta} < \frac{HQ}{HP}$, décroissante dans le cas contraire.

b) γ et δ ne sont pas dérivables par rapport à n en n_c car h_1 ne l'est pas.

Il y a donc pour $n=n_c$ basculement discontinu de l'heure de pointe. Ce basculement correspond nécessairement à une augmentation de $\rho(h_1)$, et par contre coup à une diminution de $\rho_0(h_1)$. En effet, on peut admettre que T_p varie continuement en fonction de n. En appelant h_1^- et h_1^+ les heures de pointe correspondant respectivement à $n=n_c^-$ et $n = n_c^+$, on a :

$$T_p(n_c^-) = n_c\,\rho(h_1^-) + (n_0 - s_0)\,\rho_0(h_1^-) - n_0\,\rho_0(h_0)$$

$$T_p(n_c^+) = n_c\,\rho(h_1^+) + (n_0 - s_0)\,\rho_0(h_1^+) - n_0\,\rho_0(h_0)$$

Puisque $T_p(n_c^-) = T_p(h_c^+)$, on a :

$$n_c\left[\rho(h_1^+) - \rho(h_1^-)\right] = (n_0 - s_0)\left[\rho_0(h_1^-) - \rho_0(h_1^+)\right]$$

Si $\rho(h_1^+) > \rho(h_1^-)$, alors $\rho_0(h_1^-) > \rho_0(h_1^+)$

Cependant, il est clair que la rentabilité ne dépend que de $\frac{A}{T_{p}.i}$ qui est une fonction continue de n. Par conséquent, <u>cette rentabilité ne subit pas de discontinuité en n_c</u>. De part et d'autre de n_c, son sens de variation en fonction de n dépendra de la valeur de $(\beta\gamma - \alpha\delta)$.

2.2.2. Cas particulier : l'heure chargée ne dépend pas du volume du trafic des nouveaux services (dans le domaine de variation de n)

Dans ce cas, γ et δ ne dépendent pas de n et $(\beta\gamma - \alpha\delta)$ garde un signe constant quand n varie. Donc la rentabilité varie de façon monotone en fonction de n, entre 2 basculements de l'heure de pointe.

Les situations particulières suivantes peuvent se rencontrer :

- absence de substitution : $s_0 = 0$. Deux cas se présentent

. si l'heure de pointe est celle du trafic de base = $h_1 = h_0$. Alors $\delta = 0$ et $\frac{A}{T_{p.i}} = \frac{\alpha}{\gamma} = c^{te}$: le taux de rentabilité ne dépend pas de n.

. s'il y a modification de l'heure de pointe $(h_1 \neq h_0)$: $\frac{\alpha}{\gamma} > 0 = \frac{\beta}{\delta}$ donc le taux de rentabilité du trafic additionnel décroît en fonction de son volume.

- absence de modification de l'heure de pointe : $h_1 = h_0$. Dans ce cas, le taux de rentabilité du trafic additionnel est une fonction croissante de son volume si sa rentabilité immédiate est inférieure à celle du trafic substitué, et en est une fonction décroissante dans le cas contraire.

3. Rentabilité économique d'un échéancier de trafic de nouveaux services

On examine maintenant le cas où le trafic des nouveaux services varie dans le temps selon un échéancier donné. On étudiera successivement la rentabilité des investissements élémentaires successifs qu'il nécessite et la rentabilité globale de l'ensemble de ces investissements.

3.1. Rentabilité marginale des investissements

On s'intéresse ici à l'évolution de la rentabilité de l'investissement marginal occasionné par le trafic des nouveaux services au cours du temps. Dans ce but on considère : les échéanciers suivants :

$n_0(t)$ = trafic téléphonique de base

$n\ (t)$ = trafic de nouveaux services

$s_0(t)$ = trafic téléphonique remplacé par du trafic de services nouveaux

$n_i(t)$ = trafic de nouveaux services induit sur le réseau téléphonique

$T_\rho(t)$ = trafic de pointe imputable aux nouveaux services.

On admet de plus que l'échéancier de trafic des nouveaux services peut s'écrire :

$$n(t) = \theta\, s_0(t) + n_i(t) = \theta\, \tau\, n_0(t) + n_i(t)$$

avec $s_0(t) = \tau\, n_0(t)$

et θ = rapport entre la durée d'une communication de nouveau service et la durée de la communication téléphonique correspondante qu'elle remplace.

On s'[illegible] la rentabilité de l'investissement marginal nécessaire entre t et (t+dt) pour écouler le supplément de trafic nouveau dn(t).

Pour cela on calcule :

a) le trafic supplémentaire de pointe $dT_\rho(t)$ imputable au supplément de trafic nouveau dn(t)

On a : $T_\rho(t) = [n_0(t) - s_0(t)]\, \rho_0[h_1(t)] + n(t)\, \rho[h_1(t)] - n_0(t)\, \rho_0[h_0]$

avec, d'après la définition de l'heure de pointe $h_1(t)$:

$$[n_0(t) - s_0(t)]\, d\rho_0[h_1(t)] + n(t)\, d\rho[h_1(t)] = 0$$

On en déduit :

$$dT_\rho(t) = [\theta\, \tau\, \rho(h_1(t)) + (1-\tau)\, \rho_0(h_1(t)) - \rho_0(h_0)]\, dn_0(t) + \rho(h_1(t))\, dn_i(t)$$

Notons que si $h_1(t)$ n'est pas différentiable en $t=t_c$, on peut définir par continuité des limites h_1^- et h_1^+ correspondant aux cas où t tend vers t_c respectivement par défaut et par excès. Il leur correspond deux variations élémentaires $dT_p(t_c^-)$ et $dT_p(t_c^+)$ définies par la formule ci-dessus.

b) le coût total actualisé de ce supplément de trafic de pointe

$$dT_p(t) \left[i \frac{(1+a)^D}{(1+a)^D - 1} + \frac{f}{a} \right]$$

c) la recette nette totale actualisée associée à dn(t) :

$$\frac{1}{a}\left[\alpha\, dn(t) - \beta\, dn_o(t)\right] = \frac{1}{a}\left[(\alpha\theta\tau - \beta\tau)\, dn_o(t) + \alpha\, dn_i(t)\right]$$

d) la valeur du coefficient $\mathcal{E}$, calculé en 2-1 :

$$\mathcal{E} = \frac{i \cdot dT_p(t)}{(\alpha\theta - \beta)\tau\, dn_o(t) + \alpha\, dn_i(t) - f\, dT_p(t)}$$

soit :

$$\mathcal{E} = \frac{i \cdot \frac{dT_p(t)}{dn_o(t)}}{(\alpha\theta - \beta)\tau + \alpha \frac{dn_i(t)}{dn_o(t)} - f \frac{dT_p(t)}{dn_o(t)}}$$

On en déduit, d'après 2-1 que la condition à satisfaire pour que la rentabilité de l'investissement marginal soit supérieure à une valeur r_o est que :

$$\boxed{\frac{dT_p(t)}{dn_o(t)} \leq \frac{\mathcal{E}_o \left[(\alpha\theta - \beta)\tau + \alpha \frac{dn_i(t)}{dn_o(t)}\right]}{i + \mathcal{E}_o f} = \eta_o}$$

avec

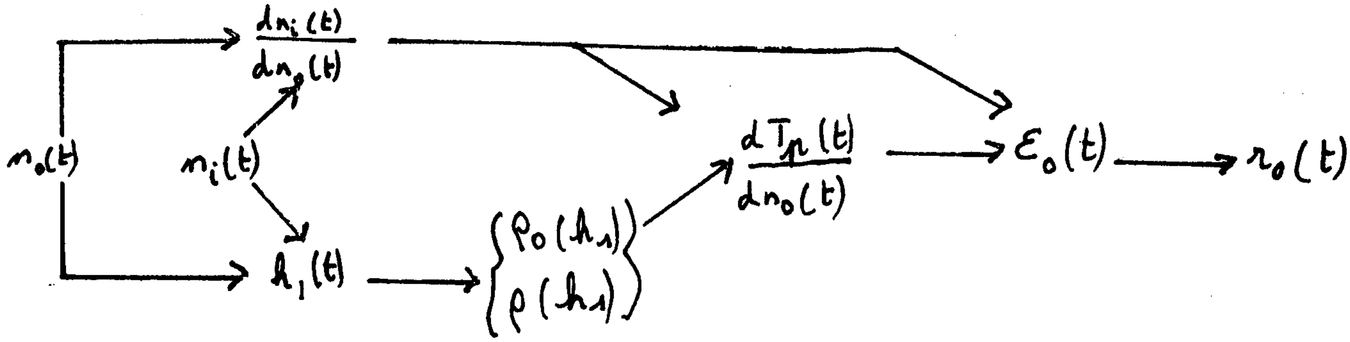
$$r_o \mathcal{E}_o = 1 - (1 + r_o)^{-D}$$

et

$$\frac{dT_p(t)}{dn_o(t)} = \theta\tau p(h_1(t)) + (1-\tau) p_o(h_1(t)) - p_o(h_o) + p(h_1(t)) \frac{dn_i(t)}{dn_o(t)}$$

La relation ci-dessus fournit donc la contrainte que doit respecter l'échéancier de trafic de nouveaux services, pour que l'investissement marginal de réseau ait une rentabilité au moins égale à une valeur donnée.

Plus précisément, on peut déduire des valeurs de $n_0(t)$ et de $n_i(t)$, celles de $r_0(t)$, par l'enchaînement suivant :



Par contre, dans le cas général, il n'est pas possible d'effectuer le passage inverse et de faire correspondre à $r_0(t)$ la fonction $n_i(t)$ associée. On ne peut donc opérer que par <u>simulation</u>, en adoptant différents échéanciers $n_i(t)$ et en ne retenant que ceux qui fournissent $r_0(t) \geqslant r_0$

Dans le cas particulier où on adopte : $n_i(t)=k\ n_0(t)$, hypothèse qui revient à supposer que la répartition du trafic des nouveaux services entre trafic substitué et trafic induit reste constante, l'heure de pointe résultante $h_1(t)$ ne varie plus dans le temps et on a :

$$\frac{dT_p(t)}{dn_0(t)} = (\theta\tau + k)\,\rho[h_1] + (1-\tau)\,\rho_0[h_1] - \rho_0[h_0]$$

$$\varepsilon = \frac{i\,\dfrac{dT_p(t)}{dn_0(t)}}{\alpha(\theta\tau + k) - \beta\tau - f\,\dfrac{dT_p(t)}{dn_0(t)}}$$

et la condition :

$$\frac{dT_p(t)}{dn_0(t)} \leq \frac{\varepsilon_0[\alpha(\theta\tau + k) - \beta\tau]}{i + \varepsilon_0 f}$$

Ainsi, la relation reliant ε à k devient homographique. On constatera plus bas que cette même relation intervient dans le calcul de la rentabilité globale des investissements qui est fait ci-dessous sous les mêmes hypothèses. Elle sera discutée à cette occasion. D'ailleurs dans ce cas, la rentabilité globale des investissements est égale à celle des investissements marginaux.

Rappelons pour terminer qu'un exemple de détermination d'un échéancier maximum de trafic de nouveau service compatible avec un taux de rentabilité donné de l'investissement marginal sera traité au paragraphe 4.

3.2. Rentabilité globale des investissements

On étudie maintenant le rentabilité globale de l'ensemble des investissements de réseau nécessités par un échéancier de trafic de nouveaux services. Pour pouvoir mener les calculs à leur terme, on fera les hypothèses suivantes :

- le trafic téléphonique de base croît à taux annuel constant : $n_0(t) = n_0 e^{\lambda t}$
- le trafic des nouveaux services est proportionnel au trafic téléphonique de base et la part du trafic induit y est constante: $n(t) = (\theta \tau + k)\, n_0(t)$

Dans ces conditions, l'heure chargée sur le réseau reste fixe dans le temps, à importance du trafic des nouveaux services donnée, c'est-à-dire à coefficient $(\theta \tau + k)$ fixé.

Par contre elle varie en fonction de l'importance du trafic des nouveaux services (si la répartition horaire de celui-ci diffère de celle du trafic téléphonique). On supposera que l'heure chargée ne varie que de façon discrète et qu'elle reste donc constante par intervalle de variation du coefficient $(\theta \tau + k)$.

3.2.1. Calcul du taux de rentabilité global des investissements dûs aux nouveaux services

On reprend ici les notations utilisées antérieurement, a désigne le taux d'actualisation continu. On pose $\mu = \tau\theta + k$

a) calcul des recettes nettes actualisées.

La recette nette dûe au trafic des nouveaux services perçue entre t et (t+dt) est :

$$dR(a) = [\alpha n(t) - \beta s_0(t)]\,dt = [\alpha\mu - \beta\tau]\, n_0(t)\, dt$$

La recette nette totale actualisée est donc :

$$R(a) = [\alpha\mu - \beta\tau] \int_0^{\infty} n_0\, e^{-(a-\lambda)t}\, dt = \frac{n_0[\alpha\mu - \beta\tau]}{a-\lambda}$$

b) calcul des coûts d'investissement actualisés.

- Le trafic de pointe élémentaire supplémentaire imputable aux nouveaux services entre t et t+dt est :

$$dT_p(t) = [\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)]\, dn_0(t)$$

- L'investissement élémentaire correspondant a pour coût total actualisé (y compris renouvellement toutes les D années) :

$$dC_I^1(a) = i \cdot \frac{e^{-at}}{1-e^{-aD}}\, dT_p(t)$$

- L'investissement total réalisé à partir de t=o a donc pour coût actualisé :

$$C_I^1(a) = [\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)]\, \frac{i\, n_0 \lambda}{1-e^{-aD}} \left(\frac{1}{a-\lambda}\right)$$

- L'investissement initial en t=o a pour valeur :

$$C_I^0(a) = \frac{i}{1-e^{-aD}}\, T_p(0) = [\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)]\, \frac{i\, n_0}{1-e^{-aD}}$$

- L'investissement total a donc pour coût actualisé :

$$C_I(a) = C_I^0(a) + C_I^1(a) = [\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)]\, \frac{i\, n_0}{1-e^{-aD}} \left(\frac{a}{a-\lambda}\right)$$

c) calcul des coûts de fonctionnement actualisés

. Le trafic total de pointe imputable aux nouveaux services est:

$$T_p(t) = n_0\left[\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)\right] e^{\lambda t}$$

. Le coût de fonctionnement total actualisé est donc :

$$C_F(a) = \left[\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)\right] n_0 f \int_0^{\infty} e^{-(a-\lambda)t}\,dt$$

$$= \left[\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)\right] \frac{n_0 f}{a-\lambda}$$

d) calcul du taux de rentabilité

Le taux de rentabilité r satisfait l'équation $R(r) = C_I(r) + C_F(r)$

soit : $1 - e^{-rD} = \varepsilon . r$

avec : $\varepsilon = \dfrac{\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)}{\mu\alpha - \tau\beta - f\left[\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)\right]}$

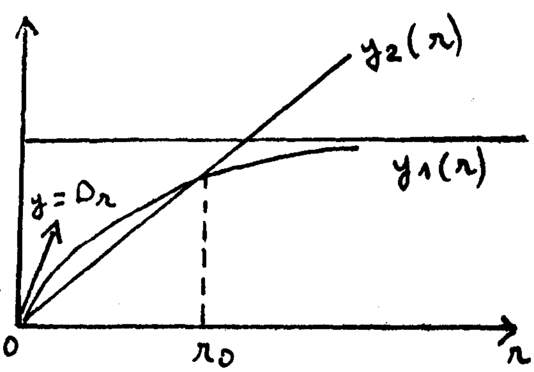
Le calcul du taux de rentabilité admet une solution graphique simple, par intersection des courbes d'équation

$y_1(r) = 1 - e^{-rD}$

et $y_2(r) = \varepsilon r$

la problème n'admet de solution $r_0 > 0$ que si $0 \leq \varepsilon < D$

et r_0 est une fonction décroissante de ε



3.2.2. Variation du taux de rentabilité en fonction du volume de trafic de nouveaux services

a) Variation du trafic de pointe imputable aux nouveaux services

Le calcul du taux de rentabilité des investissements induits par les nouveaux services, n'a de sens que si le trafic de pointe qui leur est imputable est positif.

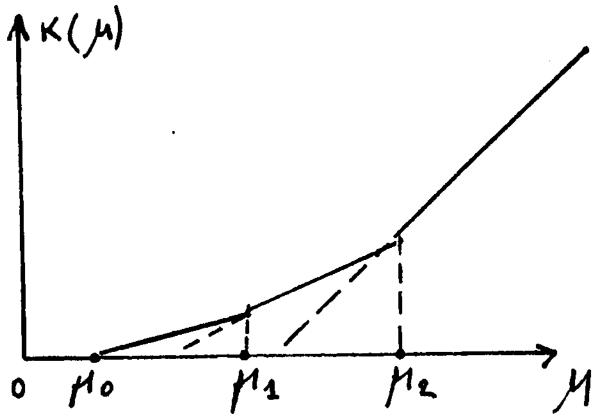
A la date t, ce trafic de pointe supplémentaire s'écrit :

$$T_p(t) = K(\mu)\, n_0\, e^{\alpha t}$$

avec

$$K(\mu) = \mu\, \rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)$$

Puisque, par hypothèse, l'heure chargée h_1 ne varie que de façon discrète en fonction de μ , les variations de $K(\mu)$ ont l'allure suivante :



En effet : le trafic des services nouveaux ne nécessite d'investissement supplémentaire que si

$$\mu > \mu_0 = \frac{1}{\rho(h_1)} \left[\rho_0(h_0) - (1-\tau)\rho_0(h_1)\right]$$

Les valeurs de μ correspondant aux basculements successifs de l'heure chargée sont, $\mu_1, \mu_2, \ldots\ldots, \ldots.$ Ainsi, pour, $\mu_1 \leq \mu < \mu_2$ l'heure chargée sera notée h_1^1 .

Puisqu'un basculement de l'heure chargée de h_1^{i-1} à h_1^i implique que, $\rho(h_1^{i-1}) < \rho(h_1^i)$, la courbe représentant les variations de $K(\mu)$ a l'allure d'une ligne brisée dont les segments ont une pente croissante. On voit de plus que dès que $\mu > \mu_0$, $K(\mu)$ reste positif quels que soient les basculements successifs.

Enfin l'abscisse du point de basculement μ_i est donné par $K(\mu_i^-) = K(\mu_i^+)$, soit

$$\boxed{\mu_i = (1-\tau)\,\frac{\rho_0(h_1^{i-1}) - \rho_0(h_1^{i})}{\rho(h_1^{i}) - \rho(h_1^{i-1})}}$$

De plus, puisque $\varepsilon(\mu) = i\,\dfrac{K(\mu)}{\mu\alpha - \tau\beta - f K(\mu)}$ les basculements de l'heure chargée <u>n'entrainent pas de discontinuïté</u> dans la valeur de $\varepsilon(\mu)$ ni donc dans celle du taux de rentabilité r_0.

b) variations de $\varepsilon(\mu)$, à heure chargée fixée.

Nous supposerons ici que l'heure chargée reste fixée et nous examinerons comment varie $\varepsilon(\mu)$ lorsque μ varie entre deux points de basculement de l'heure chargée, par exemple μ_1 et μ_2.

Soit donc h_1 l'heure chargée et $\mu_1 < \mu < \mu_2$ l'intervalle de variation de μ .

On a : $$\varepsilon(\mu) = i\,\frac{\mu\rho(h_1) + (1-\tau)\rho_0(h_1) - \rho_0(h_0)}{\mu[\alpha - f\rho(h_1)] - \tau\beta - f(1-\tau)\rho_0(h_1) + f\rho_0(h_0)}$$

. La dérivée $\dfrac{d\varepsilon}{d\mu}$ a le même signe que :

$$\boxed{Q = \alpha[\rho_0(h_0) - \rho_0(h_1)] + \tau[\alpha\rho_0(h_1) - \beta\rho(h_1)]}$$

. Les asymptotes de $\varepsilon(\mu)$ sont :

* verticalement :

$$\mu_v = \frac{1}{\alpha - f\rho(h_1)}\left[\tau\beta + f(1-\tau)\rho_0(h_1) - f\rho_0(h_0)\right]$$

on vérifie que : $\begin{cases} \text{si } Q > 0, \ \mu_v < \tau\dfrac{\beta}{\alpha} \\ \text{si } Q < 0, \ \mu_v > \tau\dfrac{\beta}{\alpha} \geqslant 0 \end{cases}$

* horizontalement :

$$\varepsilon(\infty) = \frac{i\rho(h_1)}{\alpha - f\rho(h_1)} \geq 0$$

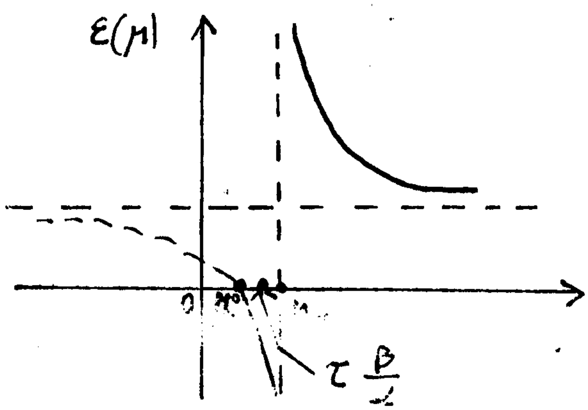
. Les intersections avec les axes sont :

$$\varepsilon(\mu) = 0 \qquad \mu^{\circ} = \frac{1}{\rho(h_1)}\left[\rho_0(h_0) - (1-\tau)\rho_0(h_1)\right] \geq 0$$

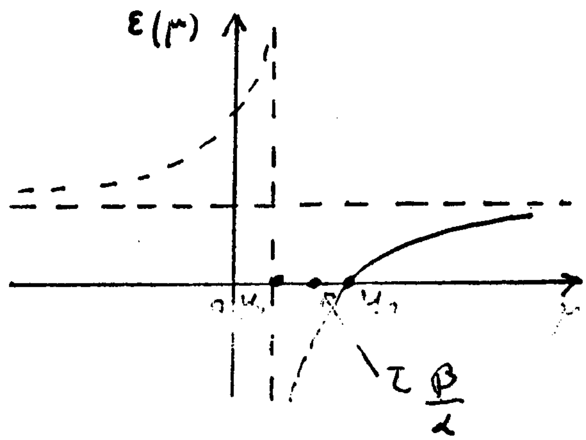
$$\mu = 0 \qquad \varepsilon(0) = i\,\frac{\rho_0(h_0) - (1-\tau)\rho_0(h_1)}{[\beta - f\rho_0(h_0)] - (1-\tau)[\beta - f\rho_0(h_1)]}$$

. Les variations de $\varepsilon(\mu)$ en fonction de μ correspondent donc à un des 2 cas suivants :

$Q < 0$



$Q > 0$



Sur les graphiques précédents, on n'a tracé en traits pleins que la partie des courbes qui correspondait à un trafic de pointe imputable aux nouveaux services positif, donc à K (μ) $>$ o.

Il est clair cependant que, dans le cas où $Q < 0$, seules les valeurs de $\mu > \mu_1$ permettent d'obtenir un $\varepsilon(\mu) > 0$, donc éventuellement une valeur de $n_0 > 0$.

c) variations de $\varepsilon(\mu)$ en cas de basculement de l'heure chargée

Lorsque μ franchit un point de basculement de l'heure chargée, les coefficients de $K(\mu)$ se modifient et ceux de $\varepsilon(\mu)$ aussi. Néanmoins, comme on l'a vu plus haut, il n'y a pas de discontinuité dans les valeurs de $K(\mu)$, ni donc dans celles de $\varepsilon(\mu)$. Les deux courbes représentant les variations de $\varepsilon(\mu)$ avant et après le basculement se raccordent donc en un point ayant pour abscisse la valeur de μ qui provoque le basculement. Nous nous proposons de situer ces deux courbes l'une par rapport à l'autre, en distinguant 2 cas :

⟶ La courbe $\varepsilon(\mu)$ initiale est croissante ($Q>0$)

. L'abscisse μ_2 du point de basculement est nécessairement supérieure à μ^0, donc à $\tau - \frac{\beta}{\alpha}$. On en déduit facilement que $(Q_2 - Q_1) > 0$, donc que $Q_2 > 0$ (si Q_1 et Q_2 sont les quantités Q associées respectivement aux courbes avant et après le basculement). Par conséquent, <u>si l'on part d'une courbe $\varepsilon(\mu)$ croissante, les basculements successifs ne pourront modifier le sens de variation de $\varepsilon(\mu)$</u>.

. On vérifie de plus que, pour $\mu > \mu_2$, on a $\varepsilon_1(\mu) < \varepsilon_2(\mu)$ puisque $K_1(\mu) < K_2(\mu)$, les indices 1 et 2 concernant respectivement les courbes avant et après basculement. Donc la courbe $\varepsilon_2(\mu)$ est située au-dessus de $\varepsilon_1(\mu)$, à droite du point de basculement.

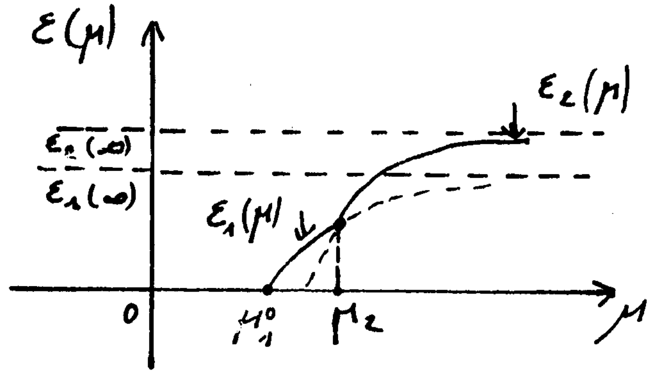
. Enfin on vérifie avec des notations analogues, que :

$$\varepsilon_1(\infty) < \varepsilon_2(\infty), \text{ car } \rho(h_1) < \rho(h_2)$$

et que

$$\mu_1^0 > \mu_2^0$$

. La position respective des deux courbes $\varepsilon_1(\mu)$ et $\varepsilon_2(\mu)$ est donc la suivante :



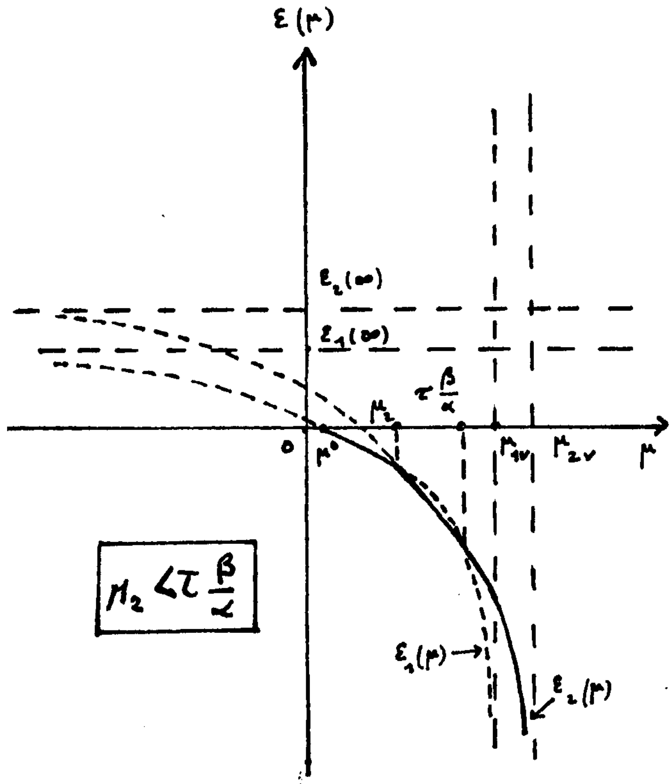
⟶ La courbe $\varepsilon(\mu)$ initiale est décroissante. $(Q < 0)$

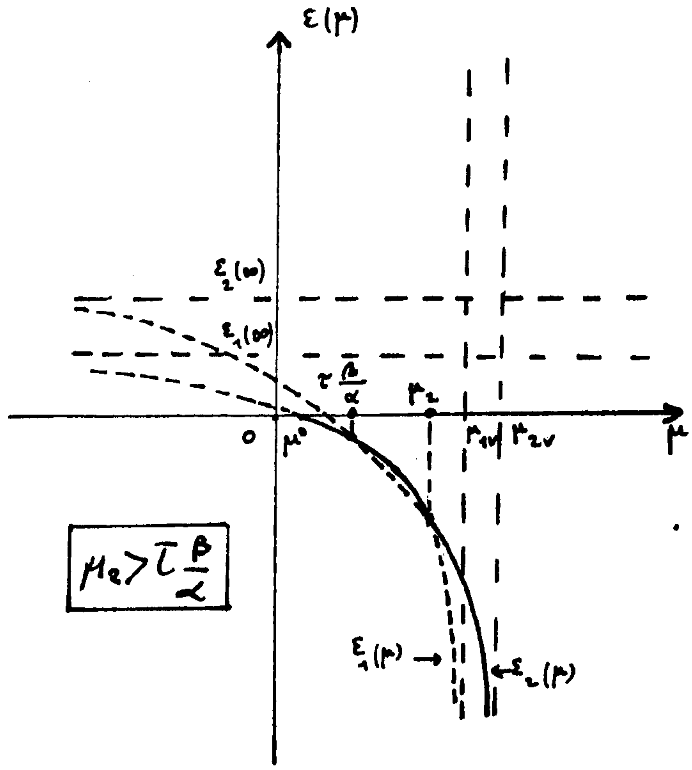
. si le basculement a lieu en un point d'abscisse μ_2 tel que $\mu^0 < \mu_2 < \mu_v$, alors $\varepsilon(\mu_2) < 0$. Puisqu'il n'y a pas de discontinuïté entre $\varepsilon_1(\mu)$ et $\varepsilon_2(\mu)$ en μ_2, et que le seul cas où l'on rencontre des valeurs de $\varepsilon(\mu) < 0$ pour $\mu > \mu^0$ est celui d'une fonction $\varepsilon(\mu)$ décroissante, la fonction $\varepsilon_2(\mu)$ sera nécessairement décroissante, elle aussi.

On a de plus : $\varepsilon_1\left(\tau \frac{\beta}{\alpha}\right) = \varepsilon_2\left(\tau \frac{\beta}{\alpha}\right)$. Par ailleurs :

* si $\mu_2 < \mu < \tau \frac{\beta}{\alpha}$: $\varepsilon_1(\mu) > \varepsilon_2(\mu)$.
* si $\mu > \sup\{\mu_2, \tau \frac{\beta}{\alpha}\}$: $\varepsilon_1(\mu) < \varepsilon_2(\mu)$.

On en déduit que $\mu_{2v} > \mu_{1v}$: le basculement de l'heure chargée agrandit l'intervalle de variation de μ à l'intérieur duquel on ne peut obtenir de taux de rentabilité $r_o > 0$. Les positions des courbes $\varepsilon_1(\mu)$ et $\varepsilon_2(\mu)$ sont respectivement les suivantes :



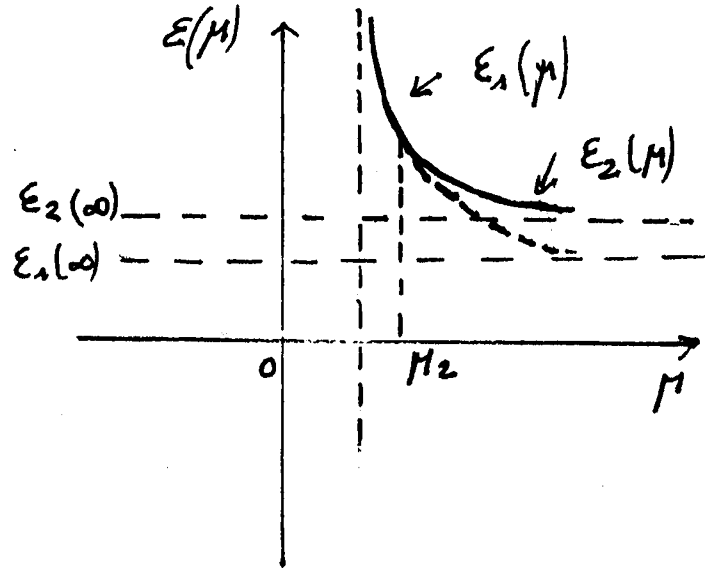


. Si l'on admet que le nombre de basculements possibles de l'heure chargée est fini, arrivera nécessairement un moment où μ , deviendra supérieur à l'abscisse μ_v de l'asymptote verticale de la courbe $\varepsilon(\mu)$. Dans ce cas, $\varepsilon(\mu)$ redevient positif. S'il n'y a pas de basculement pour $\mu > \mu_v$, $\varepsilon(\mu)$ tend vers une limite finie quand $\mu \to \infty$. Sinon, soit μ_2 un point de basculement.

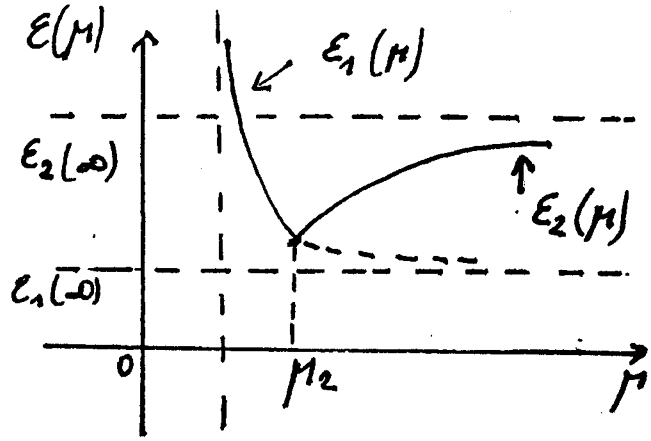
On voit facilement que :

* pour $\mu > \mu_2$, $\varepsilon_2(\mu) > \varepsilon_1(\mu)$: la seconde courbe est située au-dessus de la première.
* $\varepsilon_2(\infty) > \varepsilon_1(\infty)$: l'asymptote horizontale de la deuxième courbe a une ordonnée supérieure à celle de la première.
* si $Q_2 < 0$: $\varepsilon_2(\mu)$ décroît, comme $\varepsilon_1(\mu)$
  si $Q_2 > 0$: $\varepsilon_2(\mu)$ croît

On peut donc rencontrer l'une des deux situations suivantes :



$Q_2 < 0$



$Q_2 > 0$

On met donc en évidence dans le cas où $Q_2 > 0$ un point de rebroussement qui constitue un minimum pour $\varepsilon(\mu)$, donc un éventuel maximum pour le taux de rentabilité r. (si $\varepsilon(\mu_2) < D$)

On trouvera en annexe 1 deux exemples numériques attestant que l'on peut effectivement rencontrer les deux situations précédentes.

d) Variation du taux de rentabilité en fonction du volume de trafic des nouveaux services.

A chaque valeur de $\varepsilon(\mu)$ correspond une valeur $r_o > 0$ du taux de rentabilité, à condition que $\varepsilon(\mu) < D$ .Les variations de r_o en fonction de μ se déduisent donc de celles de $\varepsilon(\mu)$, en tenant compte de cette contrainte.

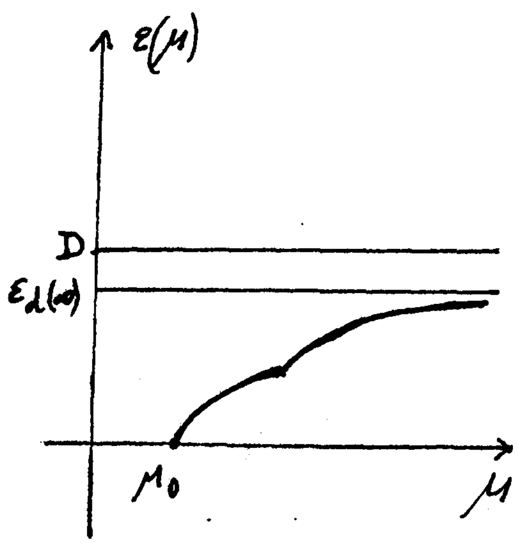
- si $\alpha[P_0(\lambda_0) - P_0(\lambda_1^0)] + \tau[\alpha P_0(\lambda_1^0) - \beta P(\lambda_1^0)] > 0$,

la courbe initiale $\varepsilon(\mu)$ est croissante et le reste quand μ augmente à partir de μ_0 .

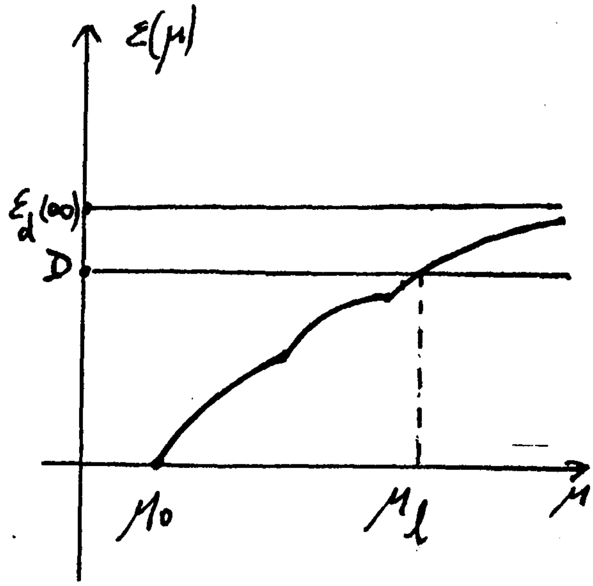
Le taux de rentabilité, initialement infini pour $\mu = \mu_0$, décroît jusqu'à ce que l'on rencontre l'une des deux situations suivantes :

. ou bien, après le dernier basculement de l'heure chargée, $\varepsilon(\mu) \to \varepsilon_d(\infty) \leq D$: la limite inférieure du taux de rentabilité est celle associée à $\varepsilon_d(\infty)$ et elle est atteinte pour $\mu \to \infty$

. ou bien, il existe une valeur μ_l telle que $\varepsilon_l(\mu_l) = D$ sur la courbe $\varepsilon_l(\mu)$ associée à μ_l .Dans ce cas le taux de rentabilité s'annule pour $\mu = \mu_l$.



$\varepsilon_d(\infty) < D$



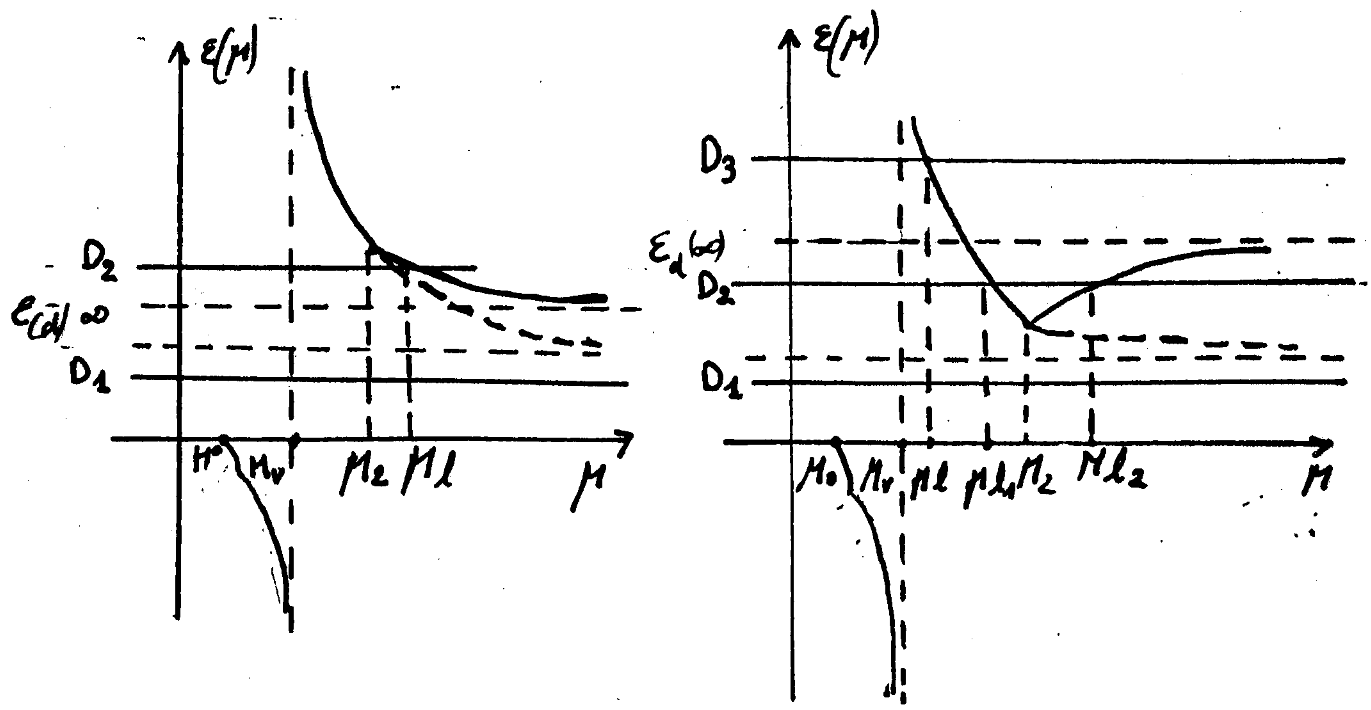
$\varepsilon_d(\infty) > D$.

- si, au contraire,

$$\alpha\left[f_0(h_0) - f_0(h_1^\circ)\right] + \tau\left[\alpha f_0(h_1^\circ) - \beta f(h_1^\circ)\right] < 0$$

la courbe initiale $\varepsilon(\mu)$ est décroissante quand μ croît à partir de μ_0.
De plus $\varepsilon(\mu)$ reste négatif jusqu'à ce que μ dépasse l'abscisse μ_v de son asymptote verticale. Pour $\mu > \mu_v$, $\varepsilon(\mu)$ décroît de l'infini jusqu'à ce qu'on rencontre éventuellement un point de basculement μ_2 de l'heure chargée. Alors deux situations peuvent se présenter :

• ou bien la nouvelle courbe $\varepsilon(\mu)$ est décroissante et on est ramené au cas précédent jusqu'au point de basculement suivant s'il y en a un. Sinon $\varepsilon(\mu)$ décroît jusqu'à son asymptote $\varepsilon_d(\infty)$
Si $D=D_1 < \varepsilon_d(\infty)$, le taux de rentabilité n'est jamais positif.
Si $D=D_2 > \varepsilon_d(\infty)$, le taux de rentabilité est positif pour $\mu > \mu_\ell$ et il croît jusqu'à la valeur associée à $\varepsilon_d(\infty)$ atteinte pour μ infini.



• ou bien la nouvelle courbe $\varepsilon(\mu)$ est croissante, et elle le reste ensuite. Selon les valeurs de D on peut alors :

. soit n'avoir jamais de taux de rentabilité positif ($D = D_1$)

. soit avoir un taux de rentabilité positif dans un intervalle $[\mu_{\ell_1}, \mu_{\ell_2}]$ ($D=D_2$)

. soit avoir un taux de rentabilité positif pour tout $\mu > \mu_\ell$. ($D = D_3$)

Dans les deux derniers cas, le taux de rentabilité est maximum pour la valeur μ_2 de μ qui correspond au point de rebroussement.

L'existence et la valeur d'un échéancier maximum de trafic de nouveaux services compatible avec un taux de rentabilité donné, se déduisent de la discussion précédente. Cet échéancier correspond, lorsqu'il existe, à la valeur maximum de μ compatible avec le taux de rentabilité fixé.

4- EXEMPLES D'APPLICATION

On traitera ici deux exemples, l'un théorique concernant la rentabilité marginale des investissements, l'autre, empirique concernant la rentabilité globale du trafic des nouveaux services.

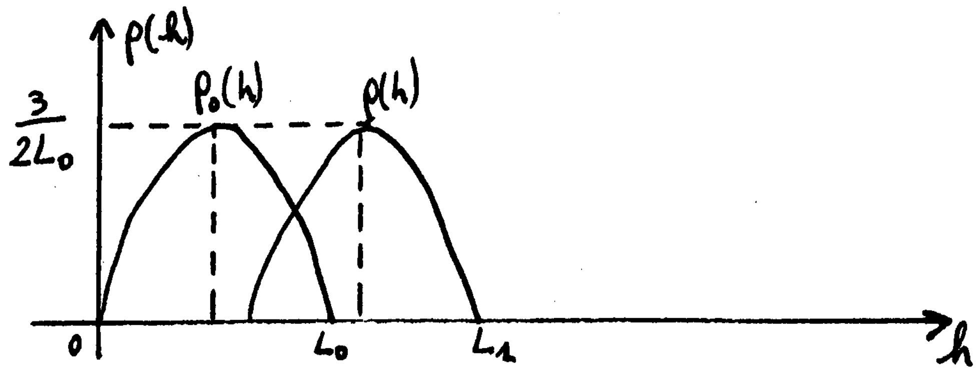
4-1- Détermination d'un échéancier maximum de trafic de nouveaux services compatible avec une rentabilité marginale donnée des investissements.

a) on retient dans cet exemple les hypothèses suivantes :

- les courbes de charge du trafic téléphonique de base et du trafic des nouveaux services sont respectivement, en fonction de l'heure h :

$$\rho_0(h) = \frac{6}{L_0^3}\, h(L_0 - h) \text{ et } \rho(h) = \frac{6}{L_0^3}(h - L_1 + L_0)(L_1 - h)$$

vérifiant $\displaystyle\int_0^{L_0} \rho_0(h)\,dh = \int_{L_1 - L_0}^{L_1} \rho(h)\,dh = 1$



- on suppose que le trafic des nouveaux services s'écrit : $n(t) = k n_0(t)$, et que le taux de substitution est nul.

b) dans ces conditions :

- l'heure chargée h_1 ne varie pas dans le temps, et ne dépend que de k :

$$h_1 = \frac{L_0 + (2L_1 - L_0)k}{2(1+k)}$$

- la relation qui assure une rentabilité marginale des investissements au moins égale à r_o s'écrit :

$$\rho_0(h_1) - \rho_0(h_0) + k\,\rho(h_1) \leq \eta_0 k$$

$$\text{avec } \eta_0 = \frac{\varepsilon_0 \alpha}{i + \varepsilon_0 f}$$

et : $$1 - (1 + r_0)^{-D} = \varepsilon_0 . r_0 .$$

- on peut montrer que :

 . si $\eta_0 \geq \frac{3}{2L_o}$, c'est à dire si r_o est suffisamment faible

il n'y a pas de contrainte sur k : le volume de trafic des nouveaux services peut donc être aussi important qu'on le veut

 . si $0 < \eta_o < \frac{3}{2L_o}$:

 + ou bien $\eta_1 = \frac{3}{2L_0^3}(3L_0 - 2L_1)(2L_1 - L_0) < \eta_0 < \frac{3}{2L_0}$

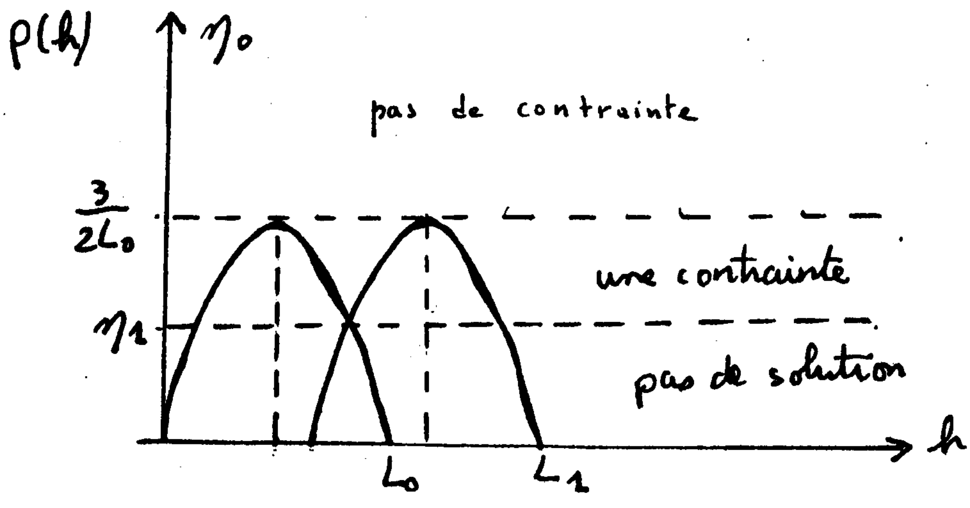
alors l'échéancier maximum de trafic des nouveaux services compatible avec r_o est :

$$n(t) \leq \left[\frac{12(L_1 - L_0)^2}{3 - 2\eta_0 L_0} - L_0^2 \right] n_0(t)$$

 + ou bien $0 < \eta_o < \frac{3}{2L_o^3}(3L_o - 2L_1)(2L_1 - L_o)$

alors il n'existe pas d'échéancier compatible avec la contrainte de rentabilité imposée.

Ces résultats sont résumés sur la figure ci-dessous



4-2- Détermination de la rentabilité globale du trafic de vidéotex en fonction de son volume.

a) On considère le trafic engendré par le vidéotex, en supposant que ce trafic ne concerne que le réseau local : ce sera le cas, soit si les banques de données sont décentralisées au niveau local, soit si l'accès aux banques de données distantes s'effectue par Transpac (dans ce cas on ne prend en compte ici que la rentabilité des investissements effectués sur le réseau local qui donne accès à Transpac).

On étudie alors les variations du taux de rentabilité des investissements occasionnés par le vidéotex sur le réseau téléphonique et on en déduit l'échéancier maximum de trafic vidéotex compatible avec une rentabilité donnée.

b) Pour y parvenir, on doit se donner :

- la courbe de charge du réseau téléphonique local : on a retenu ici une courbe obtenue pour 1978, par pondération des trafics des différentes catégories d'abonnés : ménages, indépendants, établissements et postes publics.

- la courbe de charge du trafic vidéotex : on a retenu une hypothèse de répartition horaire de ce trafic qui suppose que la part relative des ménages est plus forte dans le trafic vidéotex que dans le trafic téléphonique et que la répartition horaire du trafic vidéotex des ménages privilégie la période d'interruption du milieu de journée et la fin de la journée. Ces deux courbes de charge figurent en annexe 2.

- le pourcentage de trafic téléphonique auquel se substitue du trafic vidéotex : on a retenu trois hypothèses pour la valeur de τ = 5%,10% et 15%. Ce dernier chiffre correspond à la proportion du trafic téléphonique consacré à des recherches d'informations locales. On le considérera ici comme un maximum pour τ .

- la tarification des communications sur le réseau téléphonique : on a admis qu'elle serait d'une taxe de base par 5 mn. Pour le calcul des recettes annuelles de trafic, on supposera de plus que l'année équivaut à 235 jours.

- les caractéristiques du réseau :
  = coût d'investissement par erlang local : 121 680 F 1980
  =coût de fonctionnement annuel par erlang local : 7 764 F 1980
  =durée de vie du réseau : 20 ans.

Les estimations de coûts fournies ici proviennent d'une analyse de la comptabilité analytique des Télécommunications.

c) On désigne comme précédemment par μ la valeur du trafic vidéotex en proportion du trafic téléphonique de base.
La valeur minimum μ_o que peut prendre μ pour que le calcul du taux de rentabilité ait un sens est la valeur minimum pour laquelle

$$K(\mu) = \mu \rho(h_1) + (1 - \tau) \rho_0(h_1) - \rho_0(h_0)$$

est positif ou nul, h_1 étant l'heure chargée associée à μ.

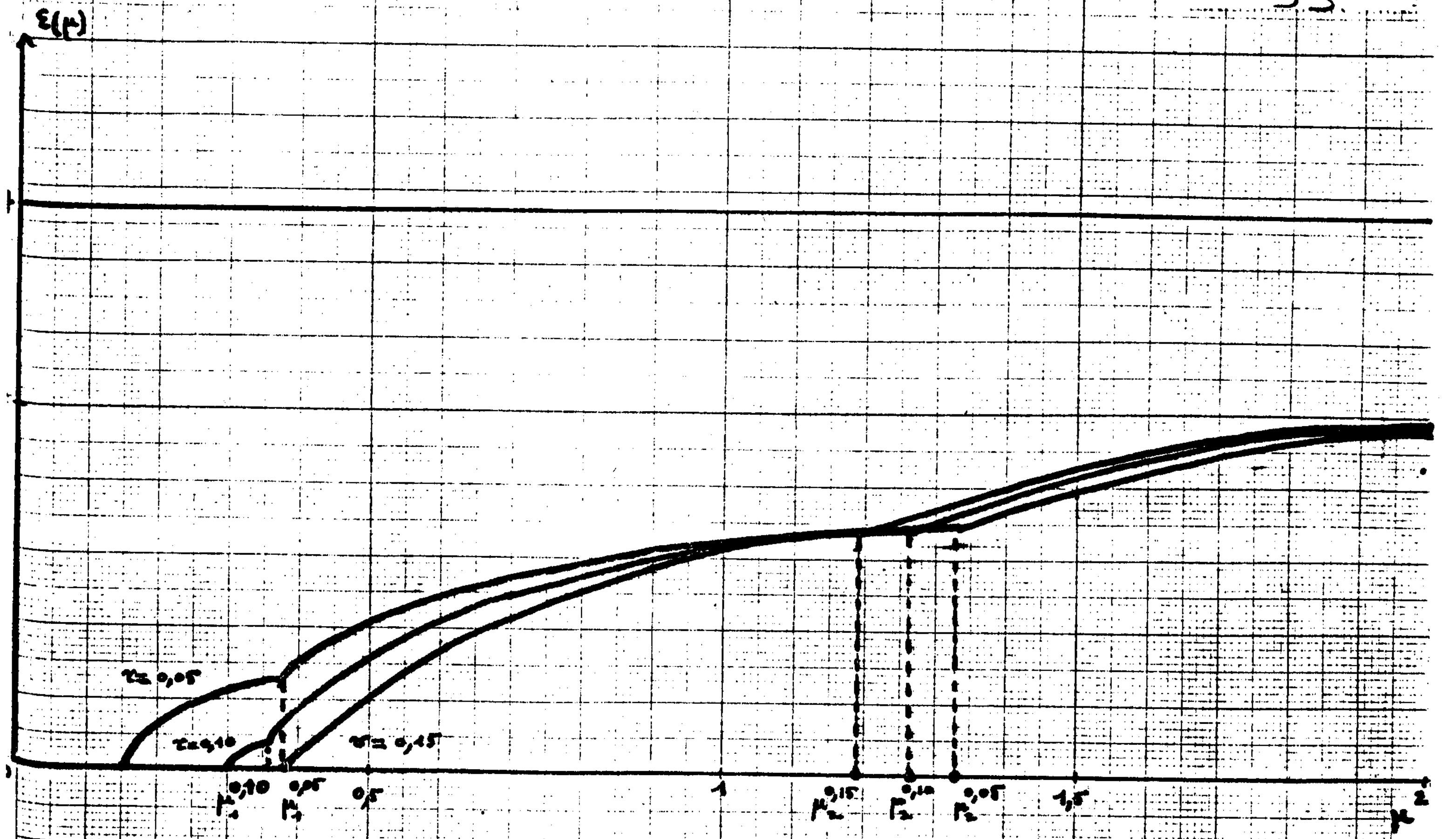
Compte tenu des courbes de charge retenues ci-dessus, on trouve pour μ_o et h_1^o associée les valeurs suivantes :

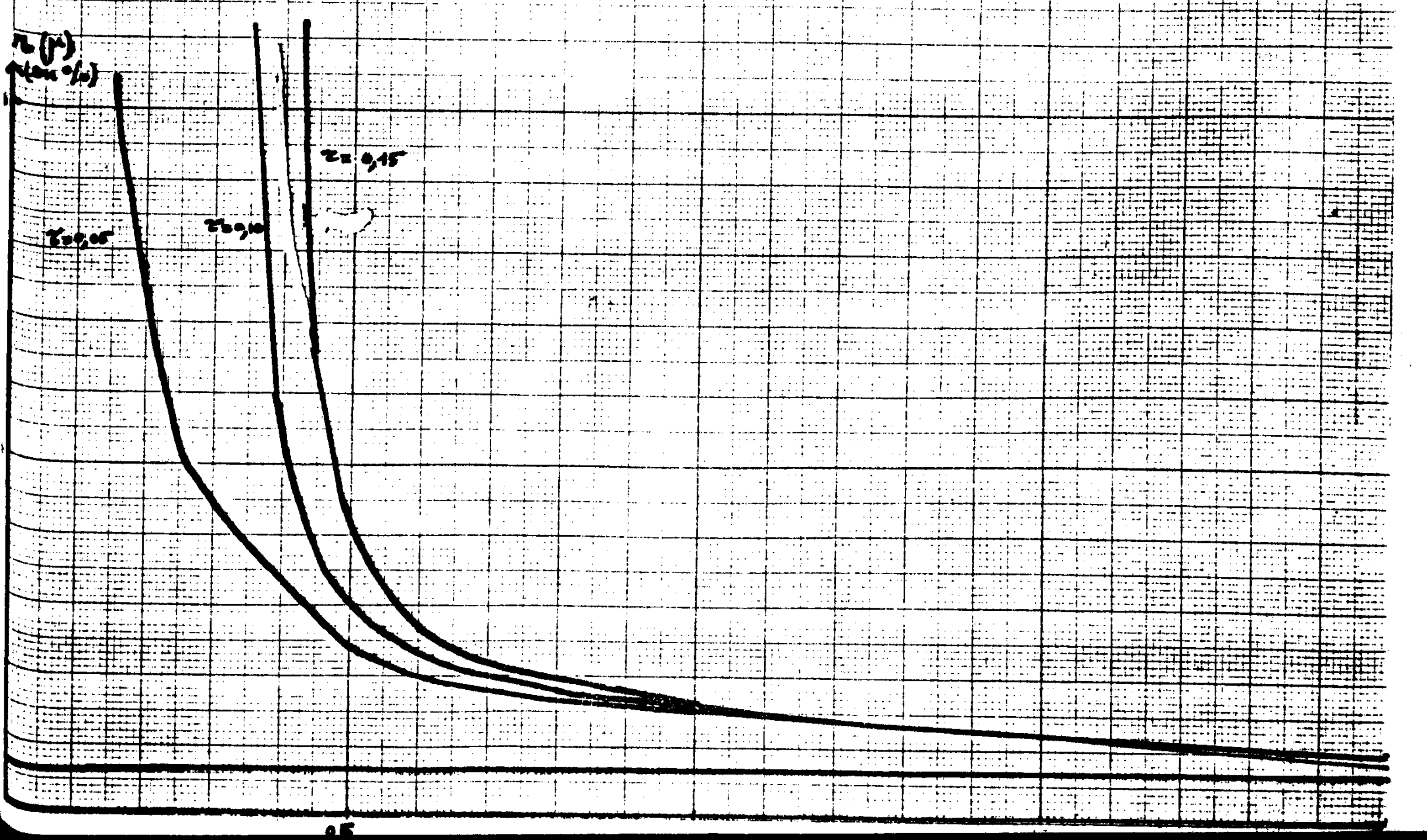
τ	0,05	0,10	0,15
h_1^0	9h30-10h	9h30-10h	14h-14h30
μ_0	0,145	0,29	0,387

d) les valeurs de $\varepsilon(\mu)$ et du taux de rentabilité $r(\mu)$ sont, en fonction de τ et de μ les suivantes :

μ \ τ	0,05		0,10		0,15		Heure chargée
	$\varepsilon(\mu)$	$r(\mu)$(%)	$\varepsilon(\mu)$	$r(\mu)$	$\varepsilon(\mu)$	$r(\mu)$	
0,15	0,08	576,2					9h30-10h
0,20	0,7	75,8					
0,25	1,0	52,0					
0,30	1,1	43,5	0,08	576,2			
0,35	1,3	39,2	0,4	117,5			
0,40	1,5	33,7	0,9	54,7	0,2	286,0	14h-14h30
0,45	1,8	27,7	1,3	37,4	0,7	66,3	
0,50	2,1	24,2	1,7	29,9	1,2	41,9	
0,55	2,3	22,0	1,9	25,8	1,5	32,5	
0,60	2,4	20,4	2,2	23,1	1,8	27,5	
0,65	2,6	19,2	2,3	21,3	2,0	24,4	
0,70	2,7	18,3	2,5	20,0	2,2	22,3	
0,75	2,8	17,6	2,6	19,0	2,4	20,8	
0,80	2,9	17,0	2,7	18,1	2,5	19,6	
0,85	3,0	16,5	2,8	17,5	2,7	18,7	
0,90	3,1	16,1	2,9	17,0	2,8	18,0	
0,95	3,2	15,8	3,0	16,5	2,9	17,4	
1,00	3,25	15,5	3,1	16,1	2,95	16,9	
1,50	3,9	12,7	4,0	12,5	4,1	12,2	19h30-20h
2,00	4,8	10,4	4,9	10,2	4,9	10,1	
∞	7,7	6,5	7,7	6,5	7,7	6,5	

Les points de basculement de l'heure chargée ont pour abscisses :
$\mu_1 = 0,4\ (1-\tau)$ et $\mu_2 = 1,4\ (1-\tau)$.
Ces résultats sont reproduits sur les graphiques qui suivent :


E(μ)
1
τ = 0,05
τ = 0,10
τ = 0,15
μ1 0,10
μ1 0,05
0,5
1
μ2 0,15
μ2 0,10
μ2 0,05
1,5
2
μ


τ = 0,15
τ = 0,10
τ = 0,05
0,5

e) Ces résultats correspondent à ce que l'étude théorique laissait prévoir. En effet :

- initialement les courbes $\mathcal{E}(\mu)$ sont croissantes pour $\mu \geq \mu_0$

car

$$Q = \alpha[\rho_0(\lambda_0) - \rho_0(\lambda_1^0)] + \tau[\alpha \rho_0(\lambda_1^0) - \beta \rho(\lambda_1^0)] > 0$$

pour chacune des trois valeurs de τ. Donc elles restent croissantes quand μ augmente, quels que soient les basculements de l'heure chargée.

De plus si $\mu = \mu_0$, $r(\mu) \to \infty$

- la valeur asymptotique $\mathcal{E}_d(\infty)$ vaut $\dfrac{i\,\rho(\lambda_1^d)}{\alpha - f\rho(\lambda_1^d)}$

où h_1^d est l'heure chargée après le dernier basculement c'est à dire l'heure chargée du trafic vidéotex seul. Cette valeur ne dépend pas de τ et vaut bien :

$$\mathcal{E}_d(\infty) = \frac{121.680 \times 0,06}{12 \times 0,50 \times 235 - 7764 \times 0,06} \doteq 7,73$$

On constate que $\mathcal{E}_d(\infty) < D = 20$. Donc $r(\mu)$ atteint son minimum pour $\mu \to \infty$

On peut déduire de ces résultats la valeur maximum $\mu(r_0)$ de μ permettant d'atteindre un taux de rentabilité global r_0 donné.

De plus on en tire la valeur maximum du trafic vidéotex induit (en proportion du trafic téléphonique) compatible avec r_0 :

$k(r_0) = \mu(r_0) - \tau\theta$

Ainsi, pour $\tau = 0,20$ et $\theta = 2$(cas où la communication vidéotex a une durée double de celle de la communication téléphonique à laquelle elle se substitue), si l'on veut $r > r_0 \doteq 30\%$

il faudra que :

$k \leq 0,50 - 0,10 \times 2 = 0,30$

ANNEXE 1

EXEMPLES DE FONCTIONS $\mathcal{E}(\mu)$ INITIALEMENT DECROISSANTES

Nous avons vu que lorsque la quantité

$$Q = \alpha\left[\varphi_0(h_0) - \varphi_0(h_1^0)\right] + \tau\left[\alpha\varphi_0(h_1^0) - \beta\varphi(h_1^0)\right]$$ était positive, la

fonction initiale $\mathcal{E}(\mu)$ était décroissante pour $\mu \geqslant \mu_0$. De plus, s'il se produit un basculement de l'heure chargée, la nouvelle fonction $\mathcal{E}(\mu)$ peut être soit décroissante, soit croissante. Nous fournissons ici un exemple de chacune de ces deux possibilités.

1. Cas où la nouvelle courbe $\mathcal{E}(\mu)$ reste décroissante

a. Les conditions à remplir pour avoir deux courbes $\mathcal{E}(\mu)$ décroissantes avant et après un point de basculement μ_1 de l'heure chargée sont, en supposant que l'heure chargée initiale h_1^0 est celle h_0 du trafic initial (ce que l'on vérifiera a posteriori):

$$① \; Q_1 < 0 \implies \beta > \alpha \frac{\varphi_0(h_0)}{\varphi(h_0)}$$

$$② \; Q_2 < 0 \implies \beta > \frac{\alpha}{\tau} \frac{\varphi_0(h_0) - (1-\tau)\varphi_0(h_1)}{\varphi(h_1)}$$

$$③ \; \mu_1 > \mu_{1v} \implies (1-\tau)\frac{\varphi_0(h_0)-\varphi_0(h_1)}{\varphi(h_1)-\varphi(h_0)} > \tau \frac{\beta - f\varphi_0(h_0)}{\alpha - f\varphi(h_0)}$$

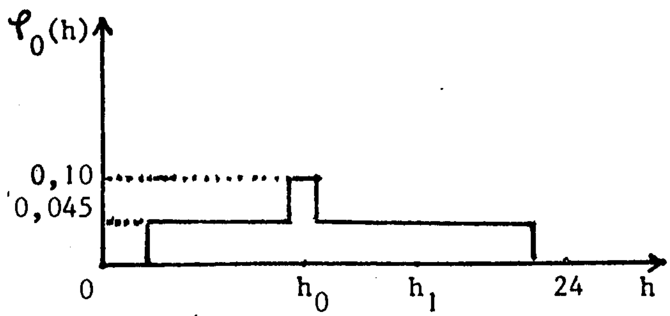
b. On peut donc construire un cas respectant les conditions ci-dessus, en :

- supposant $\alpha = \beta$
- choisissant $\varphi(h_1) > \varphi(h_0) > \varphi_0(h_0) > \varphi_0(h_1)$ (condition ①)
- calculant $\tau_0 = \dfrac{\varphi_0(h_0) - \varphi_0(h_1)}{\varphi(h_1) - \varphi_0(h_1)}$
- choisissant $\tau > \tau_0$ (condition ②)

- vérifiant que h_0 est bien l'heure de pointe initiale associée à μ_0
- vérifiant que $\varphi(h_1) < \varphi(h_0)+(1-\tau)\left(\dfrac{\varphi_0(h_0)-\varphi_0(h_1)}{\tau}\right)\left(\dfrac{\alpha - f\,\varphi(h_0)}{\beta - f\,\varphi_0(h_0)}\right)$

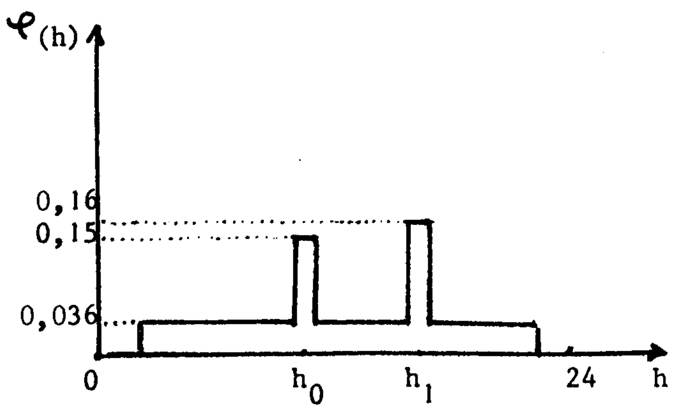
c. Exemple numérique

- Courbe de charge du trafic téléphonique :



$\varphi_0(h_0) = 0,10$
$\varphi_0(h) = 0,045$ pendant 20 heures
$\varphi_0(h) = 0$ pendant 3 heures

- Courbe de charge du trafic nouveau



$\varphi(h_0) = 0,15$
$\varphi(h_1) = 0,16$
$\varphi(h) = 0,036$ pendant 19 heures
$\varphi(h) = 0$ pendant 3 heures

- alors $\tau_0 = 0,48$
- prenons $\tau = 0,60$
- on vérifie alors que pour $\mu = \mu_0 = \tau \dfrac{\varphi_0(h_0)}{\varphi(h_0)} = 0,4$, h_0 est bien l'heure

chargée. En effet la seule heure chargée alternative serait h_1. Or, pour $\mu = \mu_0$, le trafic à l'heure h_1 vaut :

$T(h_1) = \mu_0\,\varphi(h_1) + (1-\tau)\,\varphi_0(h_1) = 0,0$...

alors que, par construction $T(h_0) = \varphi_0(h_0) = 0,10$.

- On vérifie enfin que $\varphi(h_1) = 0,16 < 0,164$ (condition ③)

Alors, le point de basculement de l'heure chargée de h_0 en h_1 correspond à : $\mu_1 = 2,2$.

2. Cas où la nouvelle courbe $\mathcal{E}(\mu)$ devient croissante.

a. Les conditions à remplir pour avoir deux courbes $\mathcal{E}(\mu)$ décroissante puis croissante avant et après un point de basculement μ_1 de l'heure chargée sont, en supposant que l'heure chargée initiale h_1^0 est celle h_0 du trafic initial (ce que l'on vérifiera a posteriori) :

① $Q_1 < 0 \implies \beta > \alpha \dfrac{\varphi_0(h_0)}{\varphi(h_0)}$

② $Q_2 > 0 \implies \beta < \dfrac{\alpha}{\tau} \dfrac{\varphi_0(h_0) - (1-\tau)\varphi_0(h_1)}{\varphi(h_1)}$

③ $\mu_1 > \mu_{1v} \implies (1-\tau) \dfrac{\varphi_0(h_0) - \varphi_0(h_1)}{\varphi(h_1) - \varphi(h_0)} > \tau \dfrac{\beta - f\,\varphi_0(h_0)}{\alpha - f\,\varphi(h_0)}$

b. On peut donc construire un cas respectant les conditions ci-dessus, en :

- supposant $\alpha = \beta$
- choisissant $\varphi(h_0) > \varphi_0(h_0) > \varphi_0(h_1)$ (condition ①)
- choisissant $\varphi(h_1) > \varphi(h_0)$
- calculant $\tau_0 = \dfrac{\varphi_0(h_0) - \varphi_0(h_1)}{\varphi(h_1) - \varphi_0(h_1)}$
- adoptant $\tau < \tau_0$ (condition ②) et tel que la condition ③ soit vérifiée.

c. Exemple numérique

• Courbes de charge du trafic téléphonique et du trafic des nouveaux services : on reprend celles utilisées en 1. , qui satisfont (1).

• On obtient : τ_0 =0,48 (condition (2)).

• On vérifie que pour τ = 0,10, la condition (3) est respectée et que h_0 est bien l'heure de pointe associée à $\mu_0 = 0,10 \times \frac{0,10}{0,15} = 0,067$

Le basculement de l'heure chargée de h_0 en h_1 se produit pour : $\mu_1 = 4,95$.

ANNEXE 2

REPARTITION HORAIRE DES TRAFICS TELEPHONIQUE ET VIDEOTEX

Heure	Trafic téléphonique	Trafic Vidéotex
Avant 8h	1	1
8h-8h30	2,1	3
8h30	4,1	3
9h	5	2
9h30	5,8	2
10h	5,2	2
10h30	5,2	2
11h	5,1	3
11h30	5,1	3
12h	3,5	4
12h30	3,2	5
13h	3,1	5
13h30	3	5
14h	5	4
14h30	4,7	3
15h	2,3	3
15h30	2,0	2
16h	3	3
16h30	3,5	3
17h	4	3
17h30	4,6	4
18h	3	4
18h30	2,6	4
19h	2,3	5
19h30	2,2	6
20h	2,2	5
20h30	2,1	4
21h	1,8	3
21h30	1,1	2
Après 22h	2,2	2
TOTAL	100	100

STRATEGIC ANALYSIS OF EMERGING TELECOMMUNICATIONS MARKETS

GEORGE S. DAY

WILLIAM R. ALLISON

University of Toronto

The decision by a company to enter an emerging telecommunications market is a major strategic commitment made in the face of considerable uncertainty. The purpose of this paper is to clarify these uncertainties and discuss their implications for forecasting the evolution of such markets. Particular attention is given to the need to integrate supply and demand factors in the forecasts and scenarios that are used to identify opportunities for profitable growth.

Since emerging telecommunications markets have many similarities with other high technology markets, we will begin by examining the characteristics of these markets as a basis for defining the significant strategic issues. Next we will review the sequence of development decisions which address these issues on the way to a decision to enter an emerging market. The related forecasting issues are necessarily interactive — progress in technology creates price reduction or feature enhancement opportunities which increase the relative advantage of the new product or service and accelerate the rate of adoption, which in turn expands the experience base and provides further opportunities for cost reduction and/or investment in market development. The last section of the paper deals with progress toward integrated methodologies for the analysis and forecasting of emerging telecommunications markets.

Characteristics of Emerging High Technology Markets

Overall, these markets are characterized by high levels of uncertainty with respect to technology, customer acceptance, competitive behavior, government intervention, and the appropriate introductory strategy. These factors make it extremely difficult to assess whether the forecast profitability will compensate for the level of risk.

(A) Customer acceptance

- low initial levels of awareness of product features and benefits.
- inertia from satisfaction with existing alternatives and unwillingness to change established behavior patterns.
- rate of acceptance depends on trade-off of features versus price (e.g. videodiscs), and the magnitude of marketing efforts to communicate and distribute the product and increase the value (e.g. provide soft-ware support and technical service).
- heterogeneity in needs, and willingness and ability to pay puts a premium on identifying most responsive target segment.

(B) Technological uncertainty

- will the technology function as expected (e.g. Thomson - CSF pursuit of low cost flat screen displays as alternative to usual cathode ray tube)?

- which technology will be dominant (e.g. grooved capacitance systems versus solid-state laser systems for videodiscs)?

(C) Uncertain costs

- high initial costs are usually offset by steep experience curves on critical components (e.g. satellite ground stations).

- potential shortages of essential components (e.g. microprocessors) with high prices reflecting the scarcity.

(D) Competitive uncertainty

- actions taken by services being supplanted to protect their position.

- commitment and resources of prior entrants.

- strategies, resources and numbers of potential direct competitors (e.g. suppliers with forward integration strategies, off-shore competitors, and companies with technological and marketing capabilities).

- other technologies which can provide similar functions (e.g. pay TV and video-tape recorders will be preferred to videodiscs for some applications).

(E) Consumer uncertainty

- heightened by lack of product standardization perceived likelihood of obsolescence, conflicting claims of competitors and erratic quality.

(F) Uncertainty re public sector

- intervention to facilitate via incentives and standards or inhibit via regulation, the development of the market.

The Evaluation of New Ventures

The decision of a company to consider entering an emerging telecommunications will depend on whether it appears to be consistent with the growth strategy or product innovation charter. Such a strategy gives direction in the following areas:

I. Objectives of innovation activities

II. Scope of business (defined in terms of functions, technologies and markets)
   - alternative growth paths

III. Elements of strategies chosen to achieve objectives
   - strengths to exploit
   - weaknesses to avoid
   - timing of entry
   - risk preference
   - method of entry

The growth strategy elements also yield criteria to be used when evaluating the new opportunity as it progresses through a sequence of development stages: screening analyses, economic feasibility analysis, product development, testing and commercialization. At each of these stages new information designed to reduce or control overall project risk is directed at the following criteria:

1. Is the opportunity real?
2. What are the risks?
3. Does it fit?
4. Is it worth doing (compared to alternative ventures?)

The last question places especially severe demands on forecasting methods, since ideally provisional answers should be forthcoming with respect to:

- the source and duration of the competitive advantage
- the size and growth of the potential market and the probable share as a function of the introductory marketing strategy
- the potential for cost reductions
- the financial consequences (ROS, ROI, payback, total investment requirements and impact on cash flow)

Implications for Forecasting

From the perspective of a manager deciding whether to participate in a market for a new telecommunications product or service, a forecast of eventual demand for the total product category is of little value unless it also incorporates information about the rate of market growth (time to saturation), probable market share, price and cost relationships, and the uncertainties of these forecasts. The problem is complicated by the dependence of these forecasts on the strategies of each of the direct and indirect competitors in the market. In short, the position of a competitor in an emerging market with an ill-defined structure will be determined by many interacting factors, only a few of which are controllable. The principal relationships among these factors are summarized in Figure 1. The bulk of this section of the paper will be devoted to reporting progress in analysing each of the factors in Figure 1, and integrating them into a useful set of forecasts which reflect alternative, plausible scenarios. Particular attention will be given to the following concepts and methods.

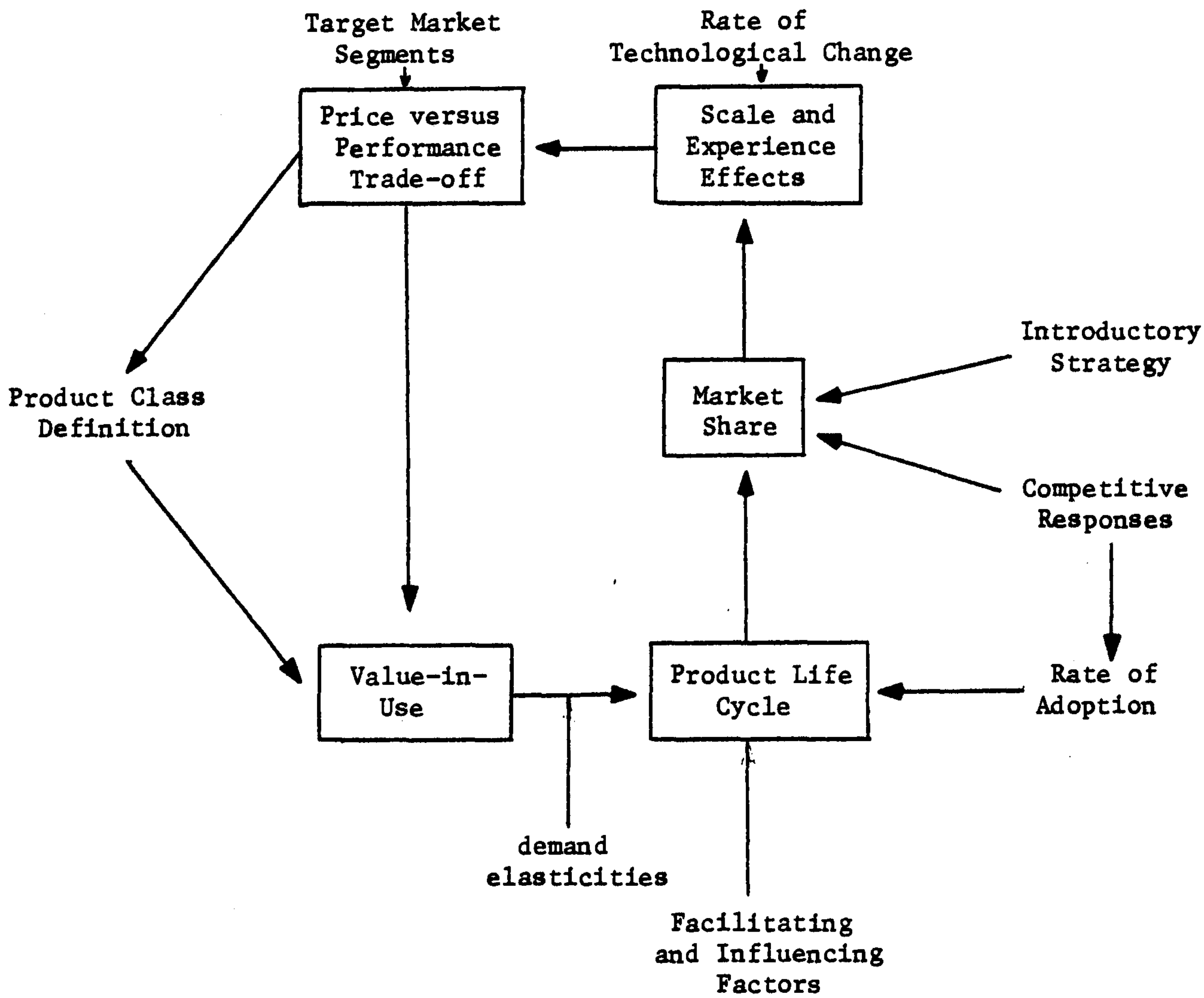
(A) The value-in-use of a new product is a measure of its economic value to the prospective buyer relative to the alternative in a specific usage situation. This in turn will be influenced by the trade-offs made by the buyer between additional features versus lower costs. From this analysis it is possible to estimate demand elasticities.

(B) The product life cycle describes the time path of sales of the total product category. It in turn reflects the timing of adoption which is distributed over the population of buyers but can be facilitated or inhibited by introductory marketing and pricing strategies and competitive responses.

(C) Scale and experience effects result in a predictable inverse relationship between unit costs and cumulative output. To the extent that declining costs are reflected in declining prices, the experience effect will influence the perceived value-in-use of the new product. Whether prices and costs are related depends on the objectives and strategies of present and prospective market entrants.

Figure 1

ELEMENTS OF FORECASTS FOR
NEW TELECOMMUNICATIONS SERVICES



COMPETITION, MONOPOLY, AND VERTICAL INTEGRATION

From:

Proceedings of the Conference on Telecommunications in Canada: Economic Analysis of the Industry

AN OVERVIEW OF THE ANALYSIS OF COMPETITION vs. MONOPOLY IN TELECOMMUNICATIONS SERVICES

JOHN STRICK

Windsor University

INTRODUCTION

The focus of the debate on competition vs. monopoly in telecommunications services has been primarily on the areas of terminal interconnection and transmission. The trends in recent years in Canada and the U.S. have been toward increased competition, supported by the CRTC in Canada, and the FCC and the courts in the U.S.

Volumes of literature in the form of books, journal and magazine articles, submissions to and reports of regulatory agencies, legislative committee hearings and evidence, conference papers, etc. have been generated on this subject. The volumes of material presented against competition have generally been matched, or exceeded, by volumes of material brought forward by pro-competition forces. For each of the arguments that competition would produce adverse effects on services, rates, and general economic harm, a counter-position can be found, together with arguments that a considerable amount of economic good would be generated from increased competition, such as improved services, more rapid innovation, greater efficiency, etc.

This paper seeks to examine the issues that have been at the centre of the question of competition vs. monopoly and to survey the economic analysis to which these issues have been subjected. Part I attempts to place the issue into historical perspective by outlining the developments toward increased competition in telecommunications in the U.S. in the past two decades, with emphasis on the various FCC decisions in this area. Part II reviews the issues involved and examines the analysis of these issues and the conclusions reached by the FCC after reviewing the various arguments and evidence.

PART I

THE TREND TO COMPETITION: HISTORICAL OUTLINE

1. Early Developments

The Bell patents of the late 19th Century provided the Bell System with a legal monopoly in the telephone industry. Bell operated unhindered by competition and provided services primarily to the business-industrial community. A period of competition (1893-1920) followed the termination of the Bell patents. New entrants began to extend services to the residential community and it is contended that it was the vigorous competition in pursuit of new markets that produced substantial rate reductions [Danielian, 1939].

Bell's reaction to this early competition included expansion of its own facilities, refusal to connect with independent companies, refusal to sell telephone equipment to non-Bell companies, and propaganda against competition. It also instituted a policy buying up the independents [Congress, 1976; Gabel, 1969].

Throughout the period of the patent monopoly and the early years of competition, the Bell System opposed government intervention and regulation of the telephone business. The efforts of the Interstate Commerce Commission (ICC), however, which had regulatory authority over telegraph, railroads, etc., to stabilize markets and price structures, eventually began to appeal to AT&T as possibly effecting stability in the telephone industry. Following 1907, the Bell System instituted an objective of a single monopolistic telephone system under government regulation, substituting regulation for the vigors of competition and in 1910 the Manne-Elkins Act conferred regulatory authority over interstate telephone companies on the ICC [Gabel, 1969].

In 1934 the Communications Act replaced the ICC with the Federal Communications Commission (FCC) and conferred greater regulatory authority upon it than had been vested in the ICC. The FCC was empowered to regulate the rates and services of interstate and international common carriers, including carrier acquisition of interstate facilities, carrier discontinuation of services, and all charges, practices, and regulations.

Following the Communications Act of 1934, Bell continued to solidify its position in the telecommunications industry. It established various departments and subsidiaries such as the Long Lines Department, Western Electric, and Bell Telephone Laboratories. Extension of services continued and the 1949 Rural Electrification Act assured the extension of service to even the most remote areas of rural America.

World War II had a significant impact on the telecommunications industry with the impetus the War gave to technological innovations to meet new telecommunication needs. The innovations fostered a renewed period of competition primarily along the lines of terminal interconnect manufacturing and transmission. The following pages trace some of the history of the development of competition in these areas, centering the involvement around the FCC.

2. Terminal Interconnection

New developments in terminal equipment and its manufacture in the 1940's gave rise to the terminal interconnect issue. Telephone recorders used by the U.S. military began to gain civilian popularity but AT&T refused to permit their connection to the telephone system. An investigation by the FCC resulted in a declaration that tariffs barring interconnection were unjust and discriminatory, but the FCC did agree with the Bell requirement of a telephone company provided interface device.

In 1949 the FCC upheld Bell's interconnect restriction as applied to the Hush-a-Phone, a small plastic device attached to a telephone headset to reduce background noise, and in 1954 it turned down a petition from manufacturers of electronic telephone answering devices to attach their devices on the grounds that there was no interstate demand for the product. In 1956, however, the Hush-a-Phone decision was overruled by the U.S. Court of Appeals which concluded that Bell's interconnection restrictions were an unwarranted interference with telephone subscribers rights to use the telephone in ways which were privately beneficial without being publicly detrimental [Hush-a-Phone Corp. vs. U.S., 238 F 2d, 1956]. The FCC subsequently implemented the Court's findings.

The Hush-a-Phone decision was used as a precedent when the interconnect issue again appeared. In the Carterphone Decision of 1968 [FCC, Carterphone, 1968], the FCC ruled against AT&T's tariffs which prohibited the interconnection of a private land mobile radio unit to the telephone network through the means of an acoustic coupler. The FCC contended that interconnection did not adversely affect the telephone company's operations or the telephone system's utility for others. The tariffs were particularly discriminatory when AT&T's own interconnect equipment was approved for use. The significance of the Carterphone decision was that it paved the way for the attachment of customer-owned terminal devices to the telephone companies lines and allowed customers to choose the kinds of terminal equipment they needed.

Some State regulatory commissions tried to bar telephone companies operating within the State from complying with Carterphone and related decisions. The FCC ruled in the Telerent case that State law could not frustrate Federal interconnection policies.

Bell's response was to submit tariffs which allowed terminal interconnection if a subscriber paid for a carrier-provided protective interface device to be inserted between subscriber equipment and the exchange network to protect the technical integrity of the network. In 1972 the FCC ruled to provide a certification program as an alternative to carrier-supplied connecting arrangements [FCC Docket 19528] establishing FCC equipment registration standards and certification of customer-owned equipment. In 1976 the FCC extended the registration program to other types of customer-supplied equipment, when used in conjunction with appropriate FCC registered protective circuitry [FCC, MTS-WATS, 1976].

The pro-terminal interconnect decisions of the FCC opened up the terminal equipment market for new entrants and saw a proliferation of new companies in the telephone equipment manufacturing industry. New and innovative products have been introduced by terminal equipment companies while the established telephone companies have been stimulated to produce their own innovations and improved devices.

3. Transmission

Microwave radio as a communications carrier was developed during the Second World War and was extended to civilian use. Petitions were made to the FCC to permit the development of private microwave systems in competition with common carrier supplied services. In the Above 890 Decision [FCC Above 890, 1959] of 1959 the FCC made some frequencies available for use by privately operated communications services on the grounds that there were uses that could not be met by the established common carriers and that the economic impact on the common carriers would be insignificant. This was the beginning of private line competition to the established carriers.

AT&T responded to this competition with Telpak, a bulk rate tariff which offered substantial discounts to large users of private lines. The discounts ranged from 51-85% below the rate for private line circuits. After a lengthy study, the FCC in 1964 determined that the Telpak rates were discriminatory against the small user of private lines, but held that the rates would be justified for the larger users so long as AT&T could prove they were compensatory [Congress, 1976; Raynor, 1974]. The FCC wanted proof that AT&T was not subsidizing its competitive offerings with monopoly service revenues by predatory pricing or pricing the competitive service below cost. The result was further study of the issues which has carried on to the present and has implications for rate regulation under either monopoly or competitive industry structures [FCC Docket 18128].

The initial private line competition introduced by the Above 890 decision was followed by the Microwave Communications Inc. Decision (MCI) in 1969 [FCC, MCI, 1969]. In this decision the FCC finally approved, after a six year controversy, the first application to build and operate specialized common carrier microwave facilities, servicing interplant and interoffice communications between St. Louis and Chicago. The FCC reasoned that the provision of private line microwave services by carriers other than AT&T would allow more efficient use of the spectrum, would bring small businessmen new services and fulfill public needs, while not posing a threat to the established common carriers.

In 1971 the FCC handed down its landmark Specialized Common Carrier Decision [FCC, SCC, 1971] which authorized the entry of special service carriers to the market. It was believed that there was an unmet need for specialized services in the interstate business and data transmission market and the increased competition would provide a wider range of specialized services. At the same time, this would not significantly affect telephone industry revenues or the rates of basic telephone services. It was also argued that competition in the specialized communications field would enlarge the equipment market for manufacturers and stimulate innovation and the introduction of new techniques, by both new entrants and AT&T itself. Competition would also afford some standard for comparing the performance of one carrier with another [Congress, 1977].

The Bell System vigorously opposed the SCC decision and in 1973 filed its Hi-Lo tariff in response to competition from specialized inter-city carriers. Bell contended that the new carriers were "cream skimming" by entering the most lucrative routes, concentrating on services between large cities with high-capacity, low-cost facilities. Bell filed a three-tiered rate structure which would reduce the rates up to 40% for some customers [Raynor, 1974]. The FCC rejected the tariff on the grounds that Bell had not shown the tariff to be compensatory.

Competition in domestic satellites came next. The Domestic Satellite Decision (or DOMSAT) of 1972 [FCC Docket 16495, 1972] extended the multiple entry policy regarding communication common carriage to the licensing of domestic specialized common carriers seeking to utilize satellite systems. It was argued that such entry would result in efficient, low-cost services and encourage technical innovations.

Another significant development in specialized common carrier services was the Execunet Decision [FCC, Execunet, 1976] and its subsequent reversal by the Courts. In 1974, MCI, a company representing an affiliation of specialized common carriers offering private line service, filed a tariff for Execunet, a class of metered-use service which permitted a subscriber to access any telephone in a distant city served by MCI via MCI's network. AT&T complained to the FCC that MCI was offering interstate long distance message toll service (MTS) under the guise of Execunet and this competed with AT&T's interstate monopoly. The FCC agreed that MCI had not been authorized to offer any service that was equivalent to MTS or WATS and forbade MCI to offer Execunet. The U.S. Circuit Court of Appeals, however, reversed the FCC decision in 1977, allowing Execunet to continue, on the grounds that the FCC's previous decisions (i.e. Specialized Common Carrier Decision) did not preclude MCI or other SCC's from offering services which the FCC did not forsee at the time those carriers had been authorized to construct facilities. The Court, however, said that the FCC could restrict future service offerings if it was found that such restrictions were in the public interest, but that such a finding was not contained in the SCC decision.

In response, the FCC launched in 1978 an inquiry to determine whether the public interest required that MTS and/or WATS should be provided on a monopoly basis. In its report in August 1980, the FCC concluded that competition in MTS-WATS was in the public interest thus removing another entry barrier in the provision of telecommunications service and opening virtually all interstate telecommunications markets to competitive entry [FCC Docket 78-72, 1980].

4. Other Areas of Competition: Data Processing; Switching; Resale; Sharing

Convergence of computers and telecommunications has brought a number of innovations such as the value-added network services or VANS and with them some new areas for competition and complex problems for regulatory agencies [Berman, 1974; Norwood, 1970]. VANS involve companies which lease transmission lines from other carriers and add additional equipment of their own to offer specialized communications services [Criner, 1975; Criner, 1977; Gamble, 1978]. Examples are the Electronic Funds Transfer System (EFTS) which replace paper transactions, and the "packet switchers" which divide information into packages and send them over computer selected routes of both AT&T and SCC's and reassemble the information upon receipt.

A 1956 Concent Decree barred AT&T from engaging in any business other than providing common carrier communication services. In its Computer Inquiry of 1971 [FCC, Computer & Comm. 1971] the FCC ruled that a communications common carrier could not offer data processing services or equipment unless a separate subsidiary was established for this purpose. AT&T, however, was barred from offering data processing services even via a separate corporate subsidiary. The decision was based on the determination that data processing was an unregulated, competitive market, and mixing regulated services could lead to cross-subsidization and predatory practices [Congress, 1976].

In March, 1976, the FCC rejected AT&T's proposed "Dataspeed 40" computer terminal on the basis that it performed data processing functions. This is a very complicated area, however, as technological innovations have made it increasingly difficult to distinguish between data processing and communications services [Cowan and Waverman, 1971; Dunn, 1969], and the complications led the FCC to re-open its 1971 Computer Inquiry with Computer Inquiry II in 1976.

In its tentative decision on Computer Inquiry II on May 17, 1979, the FCC proposed to divide common carrier communications services into three categories: (1) "voice" - the traditional oral communications; (2) "basic non-voice," pure transmission of data or information, and (3) "enhanced non-voice," those services employing computer processing to act on the form and context of the information transmitted [FCC, Docket 20828, 1979]. Communications Common Carriers, as under Computer Inquiry I, providing (1) and (2), could provide (3) only through a separate corporate entity. In its first decision on Computer Inquiry II in April, 1980, the Commission reclassified communications into (1) basic communications - message and circuit switching, and (2) enhanced communications - covers all other telecommunications, and deregulated all but basic communications. In addition, terminal equipment was totally deregulated. The Commission, however, agreed to review requests from AT&T for waiver on the ban on ownership of trans-

transmission facilities for enhanced services [FCC Docket 20828, April, 1980, October, 1980].

In 1976 the FCC adopted a policy favoring the unlimited resale (subscription to communications services and facilities by one entity to be reoffered to the public with or without value-added) and sharing (non-profit arrangement by which several users collectively use communications services and facilities provided by a carrier) of common carrier private line facilities and services [FCC Docket 20097, 1976]. Common carrier tariffs had traditionally prohibited the resale and sharing of private line facilities and services, with exemptions from such restrictions for certain customers designated by the common carrier [Congress, 1977]. The FCC contended that restrictions on the subscribers resale and sharing were unjust and unreasonable, and believed that unlimited resale and sharing of private line facilities would serve the public interest. Under the FCC ruling, entities which resold communications services would be considered common carriers and be regulated under the Communications Act just like any other carrier.

Competition was similarly fostered in the area of international telecommunications. In 1976, after years in which no new entrants had appeared in international telecommunications industry, the FCC authorized two domestic value-added carriers, Telenet Communications and Graphnet Systems, to provide new international data services. At the same time, the FCC granted the existing carriers the authority to provide overseas data-phone services.

PART II

MONOPOLY vs. COMPETITION: AREAS OF ANALYSIS

The established telecommunications industry in the U.S., notably AT&T and its affiliates, vigorously opposed the trend toward increased competition in the industry over the past decade. In the debate regarding competition vs. monopoly in the telecommunications carriers industry, attention has focused on three general areas; namely:

1) the technical and operational integrity of the telephone network
2) costs, rates, and basic telephone services
3) the independent telephone companies.

1. Effects on the Technical and <u>Operational Integrity of the Network</u>

The argument has been made that harm to the network will come from the interconnection of customer-owned terminal equipment. AT&T contended that the terminal equipment produced by its competitors was inferior to its own and that interconnection would impair the technical and operational integrity of the communications network and endanger Bell personnel. This argument was presented extensively in the FCC considerations of the Hush-a-Phone and Carterphone cases and the period thereafter.

AT&T argued that the network must be integrated technically and that parts installed today must work with parts installed ten years ago which must work with parts installed ten years in the future. It feared that the telephone companies would have no assurance that customer-provided equipment would be properly installed or maintained and would threaten the technical integrity of the system [Congress, 1977].

The pro-competition stand has been that there is little evidence from experience that the integrity of the Bell system would be harmed from terminal interconnection. Historically, railroads, pipelines, and electric utilities furnished their own service and terminal equipment, all of which were interconnected to the Bell System without a protective connecting arrangement. It was further argued that in many cases the equipment was identical to that provided by the carriers themselves, particularly in markets without manufacturing affiliates, and even in the case of integrated carriers, who purchase equipment from interconnect suppliers and lease them to customers [Congress, 1976].

The proponents of competition contended that even if there did exist a possibility that customer-provided equipment might interfere with the network, there were ways to prevent such interferences with programs of equipment certification and the provision of cheap interface devices. The landmark Carterphone decision that did permit terminal interconnection also provided a system of standards and certification [FCC, Carterphone, 1968].

Numerous instances were cited in which interconnect resulted in no evidence of network damage. In 1972 the New

York Public Service Commission permitted interconnection of certified equipment by an interfacing protective device for the Rochester Telephone Co. After two years, 94% of Rochester's customers had opted for it as opposed to Bell's more expensive equipment. A study showed fewer reports of trouble on these lines than on other company lines [Congress, 1976].

This is undoubtedly one of the weaker arguments put forward in the fight against competition. In Docket 19528 [FCC Docket 19528], the FCC heard the arguments from all sides and concluded that there was no evidence that interconnection does technical harm, damage to the network, or that it results in deterioration of quality of service. Furthermore, the program of terminal equipment registration assured certain standards for terminal equipment and would guarantee that no harm would result from interconnection.

2. Effects on Costs, Rates, Innovations and Services

The issue that increased competition in the telecommunications industry will produce general economic harm is a complex one, embracing many facets of the industry.' The issue concerns competition in both terminal interconnection and transmission. Basically it is contended that increased competition will erode the advantages and benefits resulting from the natural monopoly nature of the industry, will produce a loss of revenues for the telephone companies, a reduction in contributions, and a restructuring of rates, and the end result will be increased rates, particularly for basic household telephone services.

a) Natural Monopoly

A great amount of literature exists on the issue of natural monopoly and the economics of natural monopoly. The Proceedings of the Subcommittee on Anti-Trust and Monopoly regarding the communications industry includes a 15 page annotated bibliography of definitions of natural monopoly from 18 publications [Congress, 1974].

A long-held view is that public utilities or any regulated industry is a natural monopoly and competition in public utilities is uneconomical. More recently, however, it has been argued that the telecommunications field is unlike the traditional public utility, which generally provides one product, because telecommunications are endlessly varied [Congress, 1976]. Further countering the natural monopoly thesis, it has been contended that indeed few regulated industries appear at the extreme end of the monopoly scale (natural or pure monopoly). It has been argued that every monopoly is a product of public policy and that no present monopoly can be traced back through history in a pure form [Nelson, 1966]. In the case of common carrier, they are licensed (and regulated) under State and Federal regulatory bodies. They are assigned territories in which to render telephone services to the public — a grant tendered on the premise that competition is inefficient, wasteful, and unworkable. In return, carriers are obligated to serve all users at nondiscriminatory rates, submitting costs and revenue requirements for public scrutiny. Regulation is required to ensure that monopolies do not abuse their power and that they operate in the public interest in return for their monopoly status.

Technology also contributes to a position of monopoly. Monopoly can be created by technology and can be destroyed by technology [Phillips, 1972]. When cables were the only means of sending toll messages it was economically wasteful to have a large number of companies laying cables across the nation. Consequently, AT&T Long Lines could be seen as a natural monopoly. But technology has brought a host of new means of common services and equipment. The established carriers and their affiliates no longer possess the sole expertise in the components that make up the telecommunications network and services [Irwin, 1971]. Today the message that previously could only be sent through cables can be sent by microwave transmission, or be bounced off a satellite. Technical innovation has destroyed the prior natural monopoly of AT&T Long Lines. It has also fostered competition in numerous other areas as new products enter the market as substitutes for existing products which previously commanded a monopoly position by virtue of non-substitutability [Nelson, 1975; 908, Charyk, 1971].

In their attack on competition and justification for monopoly, the anti-competition forces referred to the industry structures of telecommunications in Britain and other European countries as evidence that competition is not considered desirable. In many of these countries the telecommunications industry is a government monopoly, owned and operated by a single entity. The British system is owned and operated by the Post Office, which requires that all customer apparatus forming part of its public services should be supplied, installed, and maintained by the Post Office [Congress, 1976].

AT&T has used the traditional argument that the telephone industry was a natural monopoly, with the consequent efficiencies and low costs that economies of scale bring, and asserted that as a result of economies of scale, all of its intercity services were collectively a natural monopoly.

AT&T submitted various studies to demonstrate that economies of scale exist in long-haul distribution facilities and that the loss of any appreciable volume of interstate traffic would raise the unit cost of services carried by these facilities, and ultimately the rates of all users of Bell services [FCC Docket 20003, 1976; 1980]. Bell also attempted to illustrate that if Bell provided all the private line services during the period 1971-1991, it could supply an estimated 14% increase in total circuit miles of demand with only an 8% increase in costs. If, on the other hand, all private line services were operated by other suppliers, Bell claimed that the total costs would increase by 16%.

The question of economies of scale in the telephone industry has been examined extensively in both the U.S. [FCC, Docket 20003] and in Canada (during the CNCP Telecommunications Interconnect application proceedings [CNCP, Wilson <u>et al</u>]). The FCC was unable to analyse fully the Bell studies purporting the existence of economies of scale because basic data underlying the studies was not made available. Numerous parts of the study and the techniques used and arguments presented, however, were open to question. The FCC claimed that the relationship between economies of scale and market size had not been adequately developed and the short-run reductions in cost which accompany increased utilization which Bell was advancing in its argument were unrelated to long-run returns to scale. The FCC concluded that no documented evidence existed of economies of scale such that losses in private line traffic would increase the costs of intercity channels used for non-competitive services. Indeed if economies of scale did exist, Bell would be the lowest-cost carrier and new entry common carriers building their own facilities would be unable to compete successfully [FCC Docket 20003, 1976; 1980].

The efforts and numerous sophisticated models to test empirically for economies of scale had not, in the view of the FCC and its analysts, produced satisfactory results. The Commission concluded that the concept of economies of scale, with all of its problems of estimation, was of limited practical policy use in relation to the telecommunications industry. The major problem was that scale effects are intermixed with other effects such as those of technological change and cannot be observed in isolation. The Commission contended:

> It is altogether possible that in the telecommunications industry, rapid technological progress rather than scale is responsible for most of the economies. New technologies are an important means by which progressively larger scales of production are made possible. [FCC, Docket 20003, 1980].

Numerous other problems have been cited in various presentations and studies submitted to cast doubt on the existence of true economic economies of scale in the telecommunications industry [Congress, 1973; Congress, 1975; FCC Docket 20003, 1980].

It was also contended that there was no evidence that the manufacture of electrical equipment is characterized by such economies of scale that one firm could produce more efficiently than a number of firms. And even if such advantages did exist, one large firm would emerge as the winner from the competitive struggle and there would be no reason to impose restrictions on entry. This has not happened. Instead, it was argued that competition in terminal equipment had served to control user cost of equipment [Kuehn, 1975].

But while new firms entered the terminal equipment market, they operated at a disadvantage in competition with AT&T's huge subsidiary, Western Electric. The vertically integrated Bell System assured Western Electric of the entire Bell System market [Irwin, 1969]. The FCC considered this issue and ordered AT&T to open its market for terminal equipment to bids from outside producers.

The telephone industry has argued that it has provided the finest and cheapest telephone service but in a monopoly there is no standard by which to compare it. In a regulated monopoly industry it is difficult to determine if costs are being kept as low as possible, if benefits of innovations are promptly made to consumers, if technology is being developed as rapidly as possible, if consumers needs are being satisfied.

While regulation may prohibit undesirable practices, it cannot compel innovation and efficiency [Congress, 1976; Capron, 1971]. Competition may, however, achieve this and numerous arguments have been presented that competition has not been utilized as fully as possible as a form of regulation [Adams, 1958; Welch, 1975; Demsetz, 1968; Johnson, 1968; Shepherd, 1973].

b) Innovation

The primary argument presented in favor of competition in the terminal interconnect and transmission industries centered around the potential benefits to customers and the public. Competition would stimulate the development and improvement of new technologies, offer a wider choice, and reduce costs [Baer, 1977]. The information and material gathered, for example, by the Subcommittee on Anti-Trust and Monopoly [Congress, 1974] gives numerous examples where the existence of competitive alternatives afforded potential customers the option of buying terminal equipment precisely tailored to their needs.

AT&T has claimed that Bell has been the major innovator in telecommunications [Congress, 1973; Congress, 1977]. Yet some studies [Schnee & Gorkiewicz, 1976] have shown that important developments in satellite transmission, lasers, and optic fibres have come from other sources.

Competition was also required to force Bell to increase the rate at which it was developing technology and putting new innovations into use — a stimulus which regulation had not supplied [FCC Docket 19129; Congress, 1976]. There is ample evidence that competition had this effect. Indeed, the chairman of AT&T, John DeButts acknowledged this when he testified that competition produced by Carterphone caused AT&T to introduce three new switchboards each developed in just one year's time versus a previous average development time of six years [Congress, 1977].

Other examples have been cited of how innovation in the established industry was spurred by the stimulus of competition. These include more rapid innovations in the development of PBX's, key systems, data modems, automatic call distributors, hospital interphones, speaker phones, etc. [FCC Docket 19129; Schnee & Gorkiewicz, 1978]. In the private line services, the innovative all-digital network for data transmission introduced by the SCC DATRAN caused Bell to respond with its own Digital Data Service. The Transaction Network Service offered by Bell in some States was very similar to the packet switching services that were already offered on an interstate basis by the value-added carrier Telenet [Congress, 1977].

While particular instances can be identified where a significant innovation originated in a competitor and appeared to stimulate responses by AT&T, comprehensive reviews of the literature on the relationship between market structure and innovation [Kamien & Schwartz, 1975; Congress, 1977] could reach no conclusion concerning the relationship between firm size and market structure and the amount and kind of innovation done. Empirical studies on this question face tremendous problems in data availability and interpretation. Of the evidence available, however, little support has been found for the hypothesis that research and development (R & D) activity increases with monopoly power [Shrieves, 1978]. Relative R & D activity, measured by either output or input intensity, appears to increase with firm size up to a point and then level off. Rivalry appears to be a major element for R & D activity although there seems to be agreement among economists that small firms in highly competitive industries are not likely to engage in much innovative activity. A new empirically inspired hypothesis has emerged to the effect that a market structure intermediate between monopoly and perfect competition would promote the highest rate of innovation activity and some theoretical support for this has been advanced [Kamien & Schwartz, 1975]. But the question of just how much size and market concentration is ideal for innovation is unsettled. The answer probably varies depending on the industry concerned and the particular time in its development, together with the state of scientific knowledge upon which it bases its research activities.

The argument that competition would result in wasteful duplication has been questioned on the grounds that it is based on dubious assumptions, the primary one being that the market is static. The pro-competition argument is that the market is not static but is increasing rapidly and is showing a large diversity of demand. New entrants have been capturing a share of new markets rather than existing ones, and have been creating new markets, and the growth of entrants has not been as great as the growth of the market [Congress, 1976; FCC Docket 20003, 1976]. Duplication would occur only if the rate of capacity expansion by established carriers and entrants together exceeded the rate of market growth. The telephone companies dominate the industry by a wide margin with SCC's having only about 2.1% of the total private line market [Congress, 1976]. Combining the private line and terminal equipment markets — the two areas open to competition — the telephone industry received $4.1 billion or 95.5% of the revenue as compared with $194 million or 4.5% for the competitors [FCC, Docket 20003, 1976].

Failing to convince the FCC of the necessity of monopoly in the telecommunications industry, AT&T attempted to restore its monopoly position in all aspects of the industry through Congress. The result was the industry sponsored Consumer Communications Reform Bill of 1976 or Bell Bill [Congress, 1976].

The Bill, in essence, would have placed new barriers in the way of competition in private lines and terminal equipment; given jurisdiction over terminal equipment to the fifty State public utility commissions thereby hampering any competitive marketing on a nationwide basis; used incremental cost pricing which would have allowed the subsidization of AT&T's competitive services and products by its monopoly services; placed the proof on competing private line carriers to show that their proposed services would not be similar to, or duplicate, existing or potential services of the established industry; and would have exempted established common carriers from anti-trust laws thereby enabling them to acquire the facilities of other carriers [Congress, 1976]. The Bell Bill aroused considerable opposition from various sectors [FCC Report, 1976], and eventually slowly receding into the background to be replaced by the Communications Bill of 1978 before the House Subcommittee on Communications. The aim of this Bill was to repeal the Communications Act of 1934 and it departed from the Bell Bill almost completely. The main features of the 1978 Bill included the proposal to replace the FCC by a new more limited agency, to be called the Communications Regulatory Commission (CRC), and the establishment of new rate-making procedures. It stressed competition as much as possible as opposed to monopoly and regulation. It provided for vertical disintegration of the industry, stipulating that, after a period of three years, telephone companies could not provide monopoly services and also own an equipment manufacturing subsidiary. This would require the divesture of Western Electric from AT&T. In essence, the Communications Bill represented an approval of the trend of FCC policy on competition since Carterphone.

AT&T and its affiliates reacted by massing opposition to attempt to remove the manufacturing subsidiary divesture proposal, arguing that it would attract foreign competition from Europe and Japan and result in a major restructuring of the U.S. telecommunications industry.

Since the introduction of the 1978 bill to rewrite the Communications Act of 1934, two bills were introduced in the Senate (S. 611 and S. 622) with similar objectives, and the new Congress of 1981 will likely foster further attempts to bring the Communications Act of 1934 up to date.

c) Contributions

The greatest controversy among the anti- and pro-competition forces concerning the question of general economic harm, and one which fostered a considerable amount of analysis, centered around the contributions issue. The telephone companies contended that competition would reduce contributions and consequently residential telephone rates would rise. The FCC and pro-competition proponents dispute this and have even contended that there is evidence to show that residential telephone service, rather than being subsidized by contributions from other services, actually subsidize other services, including those subject to competition (private line services and terminal interconnection). The Bell system contended that its long-distance services and specialized services, such as private line services and terminal equipment leasing, have been historically priced to generate revenues which exceed direct costs and this excess have been used as "contributions" to help keep down rates of basic telephone services (cross-subsidization). Competition would lead to contribution losses as the over-priced, specialized services are lost to competition, and the price of subsidized services would consequently rise [FCC Docket 20003, 1976]. A variation of this argument is that all services contribute a certain amount of revenue greater than their directly attributable costs, in order to cover joint and common costs. Any reduction in revenue from one service must be made up by increased revenues from others. Long distance services and certain specialized services have traditionally been priced above their direct costs to produce a higher contribution to the common costs of the business than does basic exchange service.

The telephone companies argued that specialized common carriers would engage in "cream skimming." That is, they would concentrate on the high-density, high-profit, private line business used by large commercial companies and government. They conceded that this private line business was vulnerable to competition because the rates they charged were not based precisely on costs but on value of service. These rates were higher than cost accounting methods would dictate to yield revenues for cross-subsidization purposes [FCC Docket 20003, 1976].

A number of industry and industry-sponsored studies have been presented, particularly in FCC inquiry Docket 20003, to demonstrate that basic exchange rates for households would increase dramatically as a result of the loss of contributions that competition would produce.

In its Embedded Direct Cost (EDC) studies, AT&T attempted to demonstrate that local rates would increase by 70% if competition eliminated all revenue contributions above embedded direct costs (the fixed and variable costs which can be readily assignable to a particular service, calculated on the basis of historical book value) for interstate and intrastate services (FCC Docket 20003, 1976; Congress, 1975]. According to the AT&T evidence, the embedded direct costs of local telephone, intrastate, and interstate services represent 57% of total costs of operation. The remaining 43% are joint and common costs not allocable to any particular service. Each service is considered to make a contribution to cover these costs, the contribution being equal to the difference between revenues and embedded costs and any reduction in revenue is considered a loss in contributions, which would have to be made up by increased rates.

The U.S. Independent Telephone Association (USITA) also submitted a study by a consulting firm System Applications Inc. (SAI) which presented a number of scenarios projecting future impact. Its "most likely" scenario suggested that by 1985 competition would have caused a 60% raise in rates for basic home telephone service and a 56% raise for business service [Congress, 1976]. Numerous interventions were made by Congressional representatives and Commissioners of Public Service Commissions from the small, low-populated States such as Wyoming and Mississippi [Congress, 1976]. They expressed concern that loss of revenues to the Bell System from competition would result in a restructuring of basic exchange rates to the disadvantage of customers in those States. Traditionally Bell has averaged costs in setting rates, that is, it took the average of all costs (high costs to rural areas, low costs in other areas), and established a uniform rate on the basis of this average. Fear was expressed that this would likely be changed if the FCC continued its policy of selective competition, the fear being fed by statements from the telephone companies [Congress, 1976]. For example, it was argued that in the State of Mississippi, the cost to construct the plant to serve the rural customers was $3345 per customer. Monthly costs to the telephone company associated with this plant was $84, but the monthly rate paid per rural customer was $7.59. The rural customer was subsidized from the profitable operations of Bell and if competition were to reduce these profits from other areas, the subsidy would fall, and rates for rural customers would increase [Congress, 1976].

In response to the FCC's contention that specialized common carriers would have little effect on the Bell System because they were expected to provide new and innovative private line services (PLS), and that specialized services of the telephone companies provided a relatively small amount of revenues, the telephone companies charged that the competition centered on the voice grade, business communications market — a market that was well-served by the telephone industry, and that the FCC was ignoring the concept of cross-elasticity of demand. Cross elasticity refers to the fact that two services are substitutes for one another and therefore the demand for one is a function of not only its price but also the price of the substitute service. If a customer finds that the cost of private line service offered by a SCC is less than the cost of message toll service (MTS), the rates of which are a function of usage, he will opt for the former. The economic decision to use PLS or MTS or a combination of both depends solely on usage. PLS is consequently regarded by users as a direct substitute for MTS, and a shift to PLS offered by SCC's means a reduction in toll usage and a reduction in contributions from that service. And because of cross-elasticity the scope of the market opened up by competition is of much greater magnitude than is conveyed by the statistics describing the existing private line market [Congress, 1976]. Bell System analysis of its experience along the first SCC route, that of MCI between Chicago and St. Louis, showed that of the circuits gained by MCI, about 65% came from existing telephone industry private lines that were replaced by MCI services; about 20% came from the shift of existing MTS to the lower-priced MCI private line; and about 15% came from customers who subscribed to additional private lines and also shifted traffic from MTS because of prices. Bell's revenue loss as a result of reduced MTS usage by those groups who previously had Bell private line service and those who previously did not have private line service but subscribed to MCI was estimated at 25% for the first nine months [Congress, 1977].

The controversy over the contributions issue stems from the method used for allocating costs for the purpose of calculating contributions. The telephone companies employed the embedded direct cost (EDC) method in their studies, which showed residential rates increasing by as much as 70% because of loss of contributions. In the EDC approach, a service subsidy or contribution is defined as the difference between revenues and historically-valued direct costs associated with a particular service. The pro-competition forces have argued that this approach is wrong and produces economically meaningless results [Stone, T & E., Inc., 1977]. In the opinion of the FCC in

Docket 20003, the EDC study at AT&T was so flawed in its assumptions, data, and computational methods that it did not show that revenue from competing services even covered their own costs, let alone make a contribution toward covering the cost of basic telephone services [FCC, Docket 20003, 1976]. An intensive analysis by Technology and Economics, Inc. (T & E, Inc.) an FCC contractor, of the AT&T study and of numerous cost of services studies submitted to the FCC by State Regulatory Commissions, produced results which concurred with the FCC opinion in Docket 20003 [Stone, T & E, Inc., 1977]. From these studies it was shown that basic telephone services were contributing more than their proportionate share of revenue toward covering joint costs (expenses associated with the provision of two or more services which can be economically produced only in fixed proportions to each other) and common costs (costs of equipment used to provide two or more services when the provision of one of those services uses capacity which could otherwise be used to provide one or more of the other services). As a result it was concluded that basic telephone services may be helping to support competitive services [Stone, T & E, Inc., 1977], and if this were true, then the rates of basic telephone services could be lowered if the telephone industry were to lose business in the competitive market. The FCC and T & E, Inc. findings, in turn, were disputed by the industry and in studies prepared for the industry [SRI, 1977).

The different results stem from the fact that AT&T used the EDC method for calculating subsidies,. while the FCC and its allies used the fully distributed cost (FDC) method. According to the FDC approach, contributions are defined as the difference between the revenues and fully distributed costs, which include all direct costs, assignable joint costs, and a specified fraction of common costs (compared to only direct costs under the EDC approach). The FCC argued that, in failing to make any assignment of joint and common costs to particular service categories, AT&T's contribution concept failed to provide any meaningful assessment as to whether the revenues from each category was covering its properly attributable cost [FCC Docket 20003, 1976]. The FDC approach, on the other hand, allocates all plant investment and operating costs in proportion to the relative use the respective services make of facilities.

The controversy over the pricing principles involved in the EDC and FDC approaches goes back to early difficulties that the FCC had with rate structures. For many years the FCC, in its regulatory function, merely kept its eye on AT&T's overall rate of return. When Bell persuaded the FCC to allow it to increase its revenues to attain a higher rate of return, the decision on which rates to increase was left to Bell. With the advent of competition, however, the situation became more complex as it then became relevant not only how much Bell should be allowed to earn, but how it was to structure its rates to achieve whatever level of earnings the FCC had approved. The Commission needed to develop a means of evaluating rate adjustments of particular services to determine if they were fair in relation to the carriers total structure of rates if the FCC was to discharge its responsibility for policing the fairness of the new competition in telecommunications which it was authorizing.

In 1964, stemming from a charge by Western Union that Bell was imposing unreasonable low rates on some of its services, the FCC ordered AT&T to determine its investments, revenues, costs, and profits for each of its individual categories. The result was the Seven Way Cost Study (the study broke Bell's operations into seven categories of service) to determine the interrelationships between all of Bell's services. Previously, cost studies had involved one or more services but never all of them. In these individual studies, Bell had submitted fully allocated cost studies and therefore the FCC directed that Bell make its Seven Way Cost Study on that basis. It has been contended that the results of the study caused AT&T to object to the use of the fully associated costs concept as the study proved Western Union's charges to be true. Bell was clearly realizing substandard earnings on its competitive services (as low as a .3% rate of return) while earning well above its authorized rate of return on MTS and WATS (approximately 10% rate of return), its major monopoly services [Congress, 1975].

Other FCC inquiries into Bell's rates and operations followed. After trying varying theories to justify its rates, Bell settled on long run incremental cost (LRIC) [Froggatt, 1971]. But while AT&T argued in Docket 20003 that LRIC was the most appropriate methodology to establish the existence and determine the magnitude of contributions and service subsidies, most of its cost of service studies filed before State regulatory agencies were based on the EDC approach [FCC Docket 20003, 1976; Stone, T & E, Inc., 1977].

Long-run-incremental-cost (LRIC) includes all costs that can be causally allocated to the service in the long run and that can be expected to disappear if the service was eliminated. It excludes unallocable costs which would not change if the service was eliminated. Long-run-incremental-revenue (LRIR), likewise is the revenue that in the long-run would disappear with the elimination of the service. Contributions, using this approach, is the difference between LRIR and LRIC [SRI, 1977; Rostow, 1977].

The FCC adopted fully distributed cost pricing as the most appropriate method for costing services and measure of cross-subsidization, arguing that long run marginal cost pricing, while sound in economic theory, could not be adopted to the pragmatic world of telecommunications. The fully-distributed-cost approach to pricing, however, is also not without its shortcomings. There are problems involved in allocating joint and common costs to various services and danger of distorting allocations under this method.

A widely referred-to study on contributions, and which the FCC included as evidence in Docket 20003, was the New York Public Service Commission (NYPSC) Terminal Equipment Cost Study of 1975, and updated in 1976, of the New York Telephone Co. The findings showed revenue deficiencies (defined as (costs-revenues)/revenues) of 61% on an embedded cost basis and 104%

on a current cost basis [FCC Docket 20003, 1976; Stone, T & E, Inc., 1977; Congress, 1976]. In other words, the revenues from the telephone company's terminal equipment offerings fell dramatically short of the embedded costs (and current costs) of its terminal equipment and was consequently making negative contributions.

As a result of these findings, the NYPSC authorized a 25% increase in terminal equipment rates.

In general, the essence of the controversy over contributions is the lack of precise information and data. It is difficult to determine the economic cost of providing a particular service to a particular user. Much of the industry's plant, equipment, and personnel is used in common to provide many different kinds of services to many different classes of users. In the past the rate structure of the varied telephone services has only been vaguely related to the cost structure, being the product primarily of vague rate making principles, demand considerations, and historical development.

d) Unfair Competition

Another of the arguments against competition was that the FCC had not fostered true competition, when competitors were left free to serve where it would profit them most, whereas the telephone common carriers were legally required to serve every customer in their territory, regardless of the cost to them [Congress, 1976]. In addition, AT&T charged that the FCC tended to protect new competitors by hamstringing the established carriers in their attempts to make competitive responses [Congress, 1975; Rostow, 1977]. Warnings were issued that competition would force the telephone industry to stop thinking in terms of the total market and concentrate on the profit centers in both equipment and service [Congress, 1977].

A case can be made for a framework which would allow a new entrant to pursue its goals and at the same time allow the established companies to respond. But established companies such as AT&T must be prohibited from using their position to prevent successful competitive entry by underpricing their services or engaging in other discriminatory practices [Raynor, 1974; Johnson, 1976; Eger, 1977].

Indeed, in response to AT&T's charges that the FCC was fostering unfair competition, the proponents of competition cited numerous practices by AT&T which they considered as unfair and which were later cited in the U.S. Department of Justice anti-trust suit against AT&T initiated in 1974. In its anti-trust suit the government charged AT&T with such abuses as making interface devices needlessly complex, increasing the cost and the risk of malfunction; not making enough such devices, thereby delaying attachment of competitive equipment; improper installation by Bell personnel, refusing to allow non-Bell couplers to be used; and denying technical information about protective devices to competitors [Schnee & Gorkiewicz, 1978]. Bell was also accused of abuses on pricing and tariff policies, such as selective price cuts, predatory (below cost) pricing, two-tier pricing, etc. Examples were cited where AT&T lowered its prices on products in competitive areas while increasing prices in other areas. In 1971 Pacific Telephone increased its rates by 19% except for competitive equipment and Illinois Bell and Ohio Bell similarly increased prices by 20% except for the prices of competitive equipment [Schnee & Gorkiewicz, 1978]. In its charge that many of AT&T's competitive prices were below cost, the government made reference to such studies as the NYPSC study on revenues and contributions. It was also charged that unfair competitive practices by AT&T contributed to the bankruptcy of DATRAN [Congress, 1976].

As part of its anti-trust case against Bell, the government's objective is to reduce the degree of vertical integration by attempting to separate Western Electric and Bell Laboratories from the Bell System. This would open up competition in terminal equipment as Bell relies solely on Western Electric for its supply and separation would force Western Electric to compete with other suppliers.

3. Effects on the Independent Telephone Companies

The independent telphone industry in the U.S. consists of approximately 1600 separate companies, serving 18% of the nation's telephones involving approximately 51% of the country's geographical area. In terms of operating revenues, they provide 15% of basic exchange service and 16% of interstate long distance service.

There are two segments to this industry:

1. The holding operations of companies such as General Telephone & Electronics, United Telephone, Central Telephone, and

Midcontinent Telephone which serve the bulk of the independent stations and are affiliated with manufacturing and allied service firms.

2. Small family-owned firms, co-operatives, and mutuals, serving single exchange or low subscriber density areas with multiple, small, local offices. They own no manufacturing affiliates and are dependent on their operations in these local communities.

It is contended that this independent industry is vital to the telecommunications network but will be adversely affected by the competition that will come from terminal interconnect suppliers and from specialized common carriers [Congress, 1976].

a) Effects from Interconnect Suppliers

Competition in the interconnect industry could threaten the revenues of independents in various ways: (i) independents with equipment manufacturing affiliates may lose sales in the affiliate as competitors enter the terminal equipment market; (ii) there may be a reduction in revenues from rental payments for equipment by customers who may begin to purchase rather than lease equipment; and (iii) there may be a reduction in tolls on co-operatively produced services resulting from jurisdictional separations [Congress, 1976; Dittberner, 1973; Baer & Mitchell, 1975]. The first affects only the large independents with affiliates, the latter two affect all independents.

With regard to the possible loss by manufacturing affiliates of equipment sales, the pro-competition argument is that there is no sound reason why manufacturing affiliates should be protected. If their products are in demand and are priced competitively, they should not lose sales. If rival products are more appealing to users, the market share of the manufacturers will and should be diminished as the market forces dictate [Congress, 1976].

Loss of revenue from lease of terminal equipment resulting from customer purchases from new suppliers would also, it was argued, be determined by customer evaluation of rival offerings. In this regard, independents have a locational advantage in serving local customers with equipment. They are able to provide prompt service, whereas outside suppliers would have to bring in maintenance personnel from long distances at a cost of time and money. Examples were cited where some such early outside firms providing interconnect equipment experienced this operational problem and went out of business [Congress, 1976]. It was therefore concluded that loss of revenue from a reduction of rental payments from leasing equipment was not likely to be great.

It was conceded that there was a possibility of some loss of revenue from the current separations agreement and toll settlements (described in the following pages). The amount of the loss, however, would depend on the amount of the terminal equipment market lost to interconnect suppliers but if such losses are small as indicated earlier, this effect will be minimal [Congress, 1976].

Attempts were made to show that allegations of harm to the independents arising out of interconnect competition was not confirmed by historical data. In the six year period (1962-67) prior to the Carterphone decision (1968), investment by independents in station equipment increased by 71%. During the six year period (1968-74) after Carterphone, independents' investment in station equipment increased 79%. During the period 1967-76, the interconnect industry managed to obtain only four-tenths of 1% of the market. The growth of the market has far outstripped the growth of new entrants. In the period 1969-74 the revenues of the Bell System increased by $1 billion from the sale of PBX's and Key Systems services, while sales of interconnect suppliers rose by less than $100 million. In other words, growth by these suppliers was less than 10% of the growth of the market [Congress, 1976].

b) Effects from Specialized Common Carriers

The telephone operating companies, both Bell and independents, and AT&T Long Lines are interconnected into a single nationwide telephone network for the provision of interstate services. Since the telephone network is used in common for both interstate and intrastate, and multiple carriers participate in the interstate services, some process of allocating costs and revenues between these services and between the Bell System and the independent companies is required. These procedures are called "jurisdictional separations" and "settlements" procedures [Stone, T & E, Inc., 1976; Gable, 1967]. Tolls from interstate message toll services (MTS) and wide area toll services (WATS) are placed into an interstate revenue pool which is then allocated, by formula, to interstate and intrastate categories (jurisdictional separations) and to the independents (settlements).* These are often referred to as "indirect contributions." The revenues are distributed to each associated company in accordance with (1) interstate expenses (originating, handling, and terminating costs) and (2) the amount of net plant investment allocated to interstate by jurisdictional cost separations. For example, a portion of the terminal equipment costs of a local operating company is allocated to the interstate market and revenue equal to that portion flow to the local by way of toll settlement from the interstate pool. Under current jurisdictional separations procedures about 20% of terminal equipment costs are allocated to the interstate market [Stone, T & E, Inc., 1976].

The method of dividing interstate revenue is very important to the question of whether other common carrier (OCS) and interconnect competition will affect the operations of independents and local exchange rates.

Up to 60% of revenue received by local telephone companies come from the interstate toll revenue pool and therefore any reduction in this pool will affect local company revenue through the reduction of indirect contributions [FCC Docket 20003, 1976; SRI, 1977].

The telephone companies and the independents feared that

---

* "Separations" is the process of apportioning direct, joint, and common costs among different services. For the allocation of costs between interstate and intrastate services the procedures which have been developed are termed jurisdictional separations principles [Stone, T & E, Inc., 1976].

competition would reduce the level of the interstate toll pool through the diversion of MTS and WATS traffic to private line services. If interstate traffic was shifted from the switched telephone service to private line services offered by SCC's, the separations effect would be a reduction in the local plant investment and expenses allocated to interstate services. If message toll traffic was diverted, measured toll usage would go down, and, since the message may appear as a local exchange call, measured local usage would go up. The compounded result would be a reduction of the revenue transferred through the separations procedure [Congress, 1975]. Another reduction could come through a downward repricing of MTS and WATS in an attempt to prevent the diversion of traffic.

AT&T claimed that revenue losses from private line competition would result in an increase in MTS and WATS rates, while the U.S. Independent Telephone Association (USITA) took the opposite position — that AT&T would be forced to reduce MTS and WATS rates in order to prevent diversion of traffic [FCC Docket 20003, 1976]. A USITA study attempted to show that SCC's would capture 10% of the MTS and WATS market by 1985 [FCC Docket 20003, 1976]. In the case of many small independents, where over 75% of total revenues came from long distance sources, even a small reduction percentage loss of long distance revenues could have a substantial impact on overall operations [Congress, 1977].

The FCC, after a lengthy study of the issue in Docket 20003, concluded that there would be little if any diversion or repricing of future MTS and WATS due to competition and no reduction in interstate revenue from these very large and growing services. The FCC reasoned that the SCC do not duplicate MTS or WATS and it was FCC policy not to permit them to do so. The FCC showed how it had consistently rejected SCC proposals to try to enter the MTS and WATS market. Another reason advanced by the FCC as to why MTS and WATS traffic would not be lost to SCC's was that SCC's did not have the financial potential to serve any significant portion of this market, given the large initial investment required [FCC, Docket 20003, 1976]. Besides any diversion of MTS and WATS traffic was equally likely under monopolistic or competitive conditions so long as the telephone industry itself offered private line services, the revenue of which did not go into the interstate revenue pool available for settlements. And private line services offered by the telephone industry have historically been priced below cost, making them attractive alternatives to MTS and WATS users desiring to shift.

It was contended that the very nature of independents protected them from direct competition from SCC's. They are small, serving small areas so that their toll business is not enough to warrant the installation of SCC's or satellite or microwave systems [Congress, 1976].

Up to this point the argument against competition on the grounds that it will harm the independents as a result of a reduction in MTS and WATS traffic and therefore the interstate revenue pool appears weak. The FCC concluded, that on the basis of the evidence submitted in the Docket 20003 inquiry, private line competition from SCC's was not likely to have any significant effect on the overall level of the interstate revenue pool available for separations and settlements. A basic factor behind this conclusion was that SCC's would not be authorized by the FCC to enter the MTS or WATS market to directly compete with Bell Long Lines and the independents.

A major turn of events, however, came with the reversal of the FCC Execunet Decision by the U.S. Court of Appeals in 1977. When MCI, an SCC offering private line service, applied to the FCC to offer Execunet, which permitted a subscriber to access any telephone in an area served by MCI via the MCI network, the FCC agreed with AT&T that Execunet was really an interstate message toll service (MTS) which competed with AT&T's interstate monopoly. The FCC consequently rejected Execunet. The Circuit Court of Appeals reversed the FCC decision and allowed Execunet to continue.

AT&T argued that the decision could have far-reaching implications. Other specialized carriers, including domestic satellite carriers, would insist that they be allowed to offer the same services as MCI, and the result would be a major impact on the Bell Systems revenues, with losses estimated at hundreds of millions of dollars, a revision of the rate structures, a major reduction in contributions, and a reduction in direct contribution to the independents through jurisdictional separations and settlements [Schnee & Gorkiewicz, 1977].

The Appeals Court did not consider or base its decision on whether Execunet was a private line service or not, or whether Execunet was desirable or undesirable from the point of view of the public interest. The case was decided on narrow technical grounds. The Court reviewed the FCC decisions and argued that the FCC had not taken the steps required by the Communications ACT of 1934 to restrict the services of MCI; it had not defined the limits of what MCI could or could not do. The Court recognized that the FCC had the authority to restrict the type of service which MCI or any other carrier could offer, but it had to follow certain lawful procedures to exercise this authority; and it had not. Furthermore, in its decision the Court argued that AT&T had never been granted a legal monopoly over MTS and WATS and did not have any legal basis for its supposed monopoly in the MTS and WATS fields [Schnee & Gorkiewicz, 1977].

In response to the Court's action, the FCC in 1978 opened an inquiry into the MTS-WATS market structure to determine whether the public interest requires that MTS and/or WATS should be provided free from direct competition. After two years of intensive analysis and hearings, the FCC concluded that competition was feasible in the provision of MTS-WATS service [FCC Docket 78-82, August 1980). As in its prior inquiries and decisions on Specialized Common Carriers, Domsat, etc. which found competitive markets with reduced entry barriers to be in the public interest, so in this case the Commission reasoned that competition would produce lower costs, reduction of waste, make carriers more responsive to the needs and desires of customers, and make them respond more rapidly to technological change and innovation [FCC Docket 78-82, August 1980].

PART III

CONCLUSION

The issue of competition vs. monopoly in telecommunications carriers and services is extremely complex. The trends in recent years in Canada and the United States has been toward increased, selective competition. The trend has been fostered and supported by the CRTC in Canada and the FCC in the United States and has generally been upheld by the Courts. The established industry in both countries has vigorously opposed this trend. Volumes of literature and documentation in the form of books, journal and magazine articles, submission to and reports of regulatory agencies, legislative committee hearings and evidence, conference papers, etc. have been generated on this subject.

The volumes of material presented against competition has generally been matched, or exceeded, by volumes of material brought forward by pro-competition forces. In response to the various contentions that competition was undesirable and would have an adverse impact, including adverse effects on services and rates, and result in general economic harm, a counter-argument can be found, together with arguments that a considerable amount of economic good would be generated from increased competition, including improved services, more rapid innovation, greater efficiency, etc.

There appears to be no concensus concerning the issue of monopoly vs. competition in telecommunications carriers and services. Arguments have been presented favoring a variety of market structures ranging from pure monopoly, to the regulated monopoly, to various degrees of competition and regulated competition. Strong theoretical arguments, based on varying assumptions, can be made in favor of either monopoly or competition in telecommunications services. However, no conclusive empirical evidence can be found determining which market structure is most likely to produce the most efficient, low-cost-high-quality services, which is most likely to provide the most responsive services to market demand, and which is likely to offer the greatest incentive for innovation and technological development.

In certain areas of the telecommunications industry the arguments and evidence tend to weigh in favor of competition. These are areas where it has not been shown that competition would result in harm to the basic nationwide communications network, and at the same time there exist strong a priori arguments that both users and the telephone industry as a whole could benefit from the introduction of competition, and empirical evidence indicates that they have benefitted from the competition permitted. Examples include areas where economies of scale are doubtful or questionable (terminal equipment manufacturing), where it appears desirable to stimulate innovation and technological development, and where it appears there is a need for facilities and services (terminal equipment, data transmission and processing) to meet public demands and changing public needs, which do not appear to be adequately met by the existing structure.

In general, the case against competition on grounds that it has or will, in the future, have a substantial adverse impact on telecommunication services in general, and on certain classes of users in particular, has yet to be made convincingly. Indications are that the trend towards increased, selective competition in the production and delivery of telecommunications services will most likely continue.

TELESAT CANADA'S MEMBERSHIP IN TRANS-CANADA TELEPHONE SYSTEM: A CRITIQUE

JANET YALE

University of Toronto

I. INTRODUCTION

The Canadian Radio-television and Telecommunications Commission issued a decision on August 24, 1977[1] announcing that it did not approve the proposed Agreement under which Telesat Canada would become a member of the Trans-Canada Telephone System (hereinafter referred to as TCTS). The Commission's decision was based on a consideration of those regulatory and public policy issues which it felt were relevant to a determination of where the public interest lay. Further, the CRTC's deliberations on these matters were aided by evidence received from affected interests at a public hearing held over the period April 25 to June 2, 1977. Despite the Commission's careful reasoning and informed judgment, the Governor in Council on November 3, 1977 by Order in Council[2] varied the decision of the CRTC and approved the Agreement.

This paper will explore the competitive and regulatory implications of the Cabinet decision. It will be seen that the Agreement contradicts the original Parliamentary intention to create an autonomous domestic satellite system and that the current problems in determining the appropriate form for marketing satellite services are an inevitable by-product of the novel ownership structure designed for Telesat. In addition to indicating the historical nature of the problems, the paper will examine the marketing conflicts created by Telesat membership in TCTS, as well as the effect of the Cabinet decision on the competitive and regulatory environment in the telecommunications field. Finally, measures to minimize (or undermine) the competitive impact of the decision and to maintain the integrity of the CRTC will be explored.

II. HISTORY OF TELESAT

Telesat was created in a political environment unanimously committed to the establishment of a domestic satellite communications system. This general feeling is best summarized by a statement in the introduction to the government's White Paper on Domestic Satellite Communications:[3]

> "In brief, it is the Government's conclusion that a domestic satellite communications system is of vital importance for the growth, prosperity, and unity of Canada, and should be established as a matter of priority."

The impetus towards rapid development of a domestic satellite system came from two sources.[4] First, it was felt that delay in the establishment of the facility would jeopardize the availability of desired satellite locations in outer space; if these orbit positions were occupied by United States satellites, Canadian communications would become dependent on foreign provision. Secondly, negotiations between France and Quebec for provision of French language satellite service also threatened Canadian control over telecommunications. While there was general agreement regarding the objective, the questions of design, operation and ownership of the new corporation were much more contentious. It might be useful to indicate these conflicting interests so as to understand the composition of the structure which was finally adopted.

The private carriers argued that their experience in the provision of communications services and expertise in satellite earth station technology made them ideal candidates for assuming responsibility for the satellite system.[5] In fact two groups submitted design and construction proposals for the new facility: the Power Corporation and Niagara Television, and the Trans-Canada Telephone System and CNCP Telecommunications. This latter group was willing to accommodate government participation in the project provided such ownership was limited to the space segment of the system. Furthermore, the carriers contemplated that the satellite service would not be accessible to their customers directly; rather, the system would be operated as a 'carrier's carrier', thereby ensuring that satellite and conventional communications services did not compete with one another. In addition the carriers supported their case for integration of satellite and territorial facilities on the grounds of elimination of waste and duplication, ease of coordination, and subsidization of uneconomical services by users of other facilities.

The government, on the other hand, was not prepared to relinquish control of the satellite system owing to a variety of non-commercial political objectives (most of which had nation-building overtones)[6] which it wanted to pursue through satellite technology. The vast Canadian territory with its concentration of population along the southern boundary has long been recognized as inhibiting the development of a Canadian national identity. The use of public enterprise as a means of developing an integrated political community makes good sense in the context of telecommunications. The extension of live television service in both languages to all parts of Canada may be viewed as an instrument cohesion, both in terms of reducing the isolation of northern communities and maintaining French culture outside Quebec.

In terms of integrating the North, satellite technology represented the first opportunity to provide services comparable to those available in the rest of Canada in terms of quality and reliability. The alternative communications systems are not suited to northern conditions; for example, microwave relay is prohibitively expensive to provide in remote areas, having regard to the need for repeater stations every twenty-five to thirty miles, and the harsh terrain, vast distances and sparse population. The cost of satellite services, on the other hand, is insensitive to distance and environment, since the signal is beamed directly to antennas located close to the viewing communities. In addition to reducing the isolation of northern communities, it was hoped that the availability of satellite services would encourage the development of the Canadian North. By improving the quality of life and providing the requisite infrastructure, it was anticipated that both people and projects would be attracted to these remote areas.

Another important benefit of satellite services relates to the spillover effects of enhancing the Canadian communications industry in terms of the development of space systems technology.[7] Given the government's commitment to scientific research and specifically its previous investment in the aerospace and electronics industries, public enterprise was seen as a necessary vehicle to ensure maximum

Canadian content in the design and construction of satellite facilities. Furthermore, such support would enable Canada to become a leader in a new high technology industry thereby improving the country's export position and assuring Canadian participation in world space programmes. A related concern was the outflow of skilled Canadian manpower was felt that the availability of challenging positions in a developing Canadian industry would halt the migration of such individuals to the United States

A final rationale for public enterprise was the need for government representation in international affairs.[8] The frequency spectrum and desirable orbit positions are limited resources over which negotiations with foreign governments are required. As mentioned previously, sovereignty and national security considerations also dictated a government-owned satellite system owing to the desire not to be dependent on foreign governments for carriage of Canadian military and diplomatic communications. While Canadian private enterprise was willing to provide these services, the government's unwillingness to allow private control must have reflected in part an underlying concern about the potential for linkages with American communications carriers and consequent undermining of Canadian control.

It can therefore be seen that a variety of social, political and cultural objectives motivated the establishment of a domestic government-controlled satellite communications system. Furthermore, it may be argued that regulation of private carriers would not have provided the flexibility necessary to implement government policy in this field.

Trebilcock and Prichard[9] maintain that the novel technology and uncertain regulatory issues indicated the need for a form of government intervention that allowed for continuous policy adjustments to meet these evolving objectives.

The opposition in Parliament to the government's proposal of mixed public/private ownership centred on the level of commitment to these non-economic policy goals: a Conservative member argued for complete private control given the willingness of private enterprise to finance the service while the NDP saw private sector input based on profit considerations as inconsistent with the ability to attain social and political goals.[10] While the government was concerned to maintain control of the satellite system to ensure its development according to national priorities, the massive saving in public expenditure associated with carrier participation and financing was probably too attractive to resist.[11] Further the government felt that carrier participation would facilitate integration of the satellite facility with the terrestrial communications network and would allow for use of accumulated expertise in the field.[12] Finally, partial public ownership was seen as symbolically demonstrating the accrual of benefits to all Canadians from the establishment of a domestic satellite system: the general public would reap a share of the returns from a profitable public expenditure, that is, the creation of sophisticated satellite technology.[13]

Thus Telesat Canada was established in June 1969 under tripartite ownership, the federal government, the telecommunications carriers, and the general public each receiving a share.[14] At the same time various legislative devices as well as the voting structure within Telesat were designed to ensure government control.[15] For example, both expansion plans and negotiations with foreign powers were made subject to ministerial approval. In addition to the potential for furthering non-economic objectives including northern development, national unity, and maximum Canadian content in the design and construction of the facility, the chosen corporate form also addressed several competitive concerns. First of all, government control would assure competition in appropriate areas, specifically among potential suppliers of parts of the satellite system. As the White Paper puts it: "[I]n particular it would ensure competition (for contracts) between the manufacturing subsidiaries of the common carriers and those manufacturers who are independent of the carriers."[16] In addition to the cost and price benefits of competition made possible by government control, the separation of satellite and terrestrial facilities would ensure that economies of of satellite communications would be passed on to satellite users and not be retained by the common carriers. The separation of costs for the various services would also allow for explicit recognition of any need for subsidies and an appropriate adjustment of rates.[17] Finally, the grounds for rejecting the carriers' suggestion that the earth and space segments be separated related to the planning and operating complexities which would ensue and the potential for conflicts of interest between Telesat and the carriers regarding the optimal location of earth stations.

The carriers failed not only in their bid to obtain control of either the space or earth segments of the system but also in their attempt to restrict Telesat service provision to themselves. A TCTS - CNCP amendment to the legislation was proposed which would have prohibited Telesat from selling its services directly to end-users: the carriers argued that absent this constraint they would be helping to create a competitor.[18] The government rejected both the amendment and its supporting justification on the groundsthat this would amount to a surrender of control of Telesat to the carriers and a restriction of the satellite's use to serve particular interests. Further it was felt that since the carriers were not assuming an obligation to purchase Telesat's services, they should not be granted the sole right to use the facilities. In addition partial carrier ownership of a system designed to serve purposes other than those of the terrestrial communications network was not seen by the government as participation in the creation of a competitor. Finally, in terms of the attainment of non-economic objectives, the government was concerned that the carriers might refuse to provide unprofitable satellite services required for national unity or northern development purposes. Thus, the only limitation on user access was a minimum rental requirement of a whole television channel or its equivalent.

This historical survey of the issues surrounding the creation of

Telesat indicates the early desire on the part of the carriers to control the satellite s:stem or at least access to it. More importantly it reveals the Parliamentary intention to establish a single autonomous corporation for the space and earth segments of the satellite system, with private and public minority ownership interests but clear government control.

III. THE AGREEMENT

Before examining the marketing conflicts of interest created by Telesat membership in TCTS it may be useful to detail the terms of the Connecting Agreement and the attached Memorandum of Agreement entitled Schedule A.[19]

The Agreement itself provides for creation of a Board of Management composed of one representative from each TCTS member, including Telesat, whose duti would include the establishment of terms and conditions for TCTS service provision as well as the setting of rates and appointment of revenues.[20] The Agreement also specifies that decisions of the Board must be unanimous thereby ensuring that each member has a veto power. Finally, the Agreement refers to the Memorandum of Agreement for details regarding the use of satellite services, financing and revenue-sharing arrangements, etc.

The first part of the Memorandum confirms Telesat membership in TCTS on an equal footing with other TCTS members in terms of rights, obligations, duties, responsibilities and voting power.[21] Paragraph six indicates the intention to form an integrated terrestrial and satellite communications system with specific provisions for the establishment of satellite performance requirements in accordance with TCTS procedures and compatibility of satellite design with TCTS service plans. The paragraph also encourages the 'timely' exchange of satellite design concepts and other information to facilitate TCTS planning and ensure optimum use of satellites by TCTS members.

The next several paragraphs[22] deal with earth stations. Site selections are declared to be the sole responsibility of those Regulated Canadian Telecommunications Common Carriers (hereinafter referred to as RCTCC's) that are members of TCTS, but locations are subject to Telesat approval. Telesat is also permitted to retain title to earth stations, although the sites themselves need not be under Telesat ownership, provided acess, use and other arrangements are satisfactory. In addition, TCTS members have a first option to purchase Telesat earth stations and to design, own, and operate support facilities and services for earth stations.

The Memorandum prohibits Telesat from building, operating or owning terrestrial transmission facilities within a TCTS member's territory except if the service relates to the space segment of the system or to experimental activities unconnected with TCTS business.[23] Telesat is also restricted in terms of service provision to those RCTCC's willing to lease one or more whole RF channels; users requiring portions of channels must contract with those in the aforementioned group. It is also explicitly provided that rates will be the same for TCTS members and other qualifying carriers "where the same terms and conditions apply with regard to communications capability and contract duration."[24]

The financial arrangements between TCTS and Telesat, detailed in Part B of the Memorandum, provide a guaranteed rate of return to Telesat which, except for the years 1977 through 1980, is to be the after-tax weighted average rate of return on common equity earned by the federally regulated TCTS members. For the four years 1977, 1978, 1979 and 1980, the guaranteed rate would be 6%, 7%, 8%, and 9% respectively.[25] Paragraph 18 provides for equal division of Telesat operating revenues between Telesat and TCTS to the extent that an excess over the guaranteed minimum rate of return is earned.

The last part of the Memorandum, Part C, provides for TCTS assurance of extensions to the present satellite capacity to achieve a stated minimum communications capability.[26] These commitments ensure the construction of the Anik C satellite series utilizing 14/12 GHz technology.

The Agreement and the Memorandum clearly entail a significant erosion of Telesat autonomy, the competitive and regulatory effects of which will be explored in the next sections. The government, however, had indicated its willingness to support Telesat membership in TCTS prior to the CRTC hearing provided certain conditions were met.[27] These assurances included equal status for Telesat in TCTS, direct service provision by Telesat for government and experimental programmes, non-discriminatory rates for and access to satellite services for non-TCTS carriers, and inclusion of CNCP and other affected interests in planning new satellite services. In addition, and perhaps most significantly, the government wanted a commitment on the part of TCTS that the economic viability of Telesat would be maintained without public expenditure and that there would be no interference with federal legislative and regulatory authority over satellites.

The 'paper' assurances provided by TCTS in the Agreement nominally address the government's concerns; however, as noted by the Commission in its decision on the matter, they are not sufficient to eliminate the likelihood of preferential treatment for TCTS. While the government wanted to have the best possible arrangement, that is the right to control Telesat without any corresponding obligation to assume financial responsibility, it was unrealistic to expect that the Agreement's financial arrangements and provisions constraining the independent behaviour of Telesat would not entail competitive advantages for TCTS. Furthermore in terms of the economic and financial assurances themselves, the justifications for Telesat membership in TCTS provided by Mme. Sauve were no different than those raised by the carriers and <u>rejected</u> by the government at the time of Telesat's creation.[28] These alleged benefits of integration were seen to include procurement of satellite services using 14/12 GHz technology without delay, Canadian occupation of desirable but limited North American orbit positions, job creation in the Canadian aerospace industry, and preservation of Canada's leading position in a developing high-technology industry. The Commission, on the other hand, was not convinced that the proposed level of integration was required to attain these benefits and moreover, was concerned that implementation of the Agreement would jeopardize rather than ensure the realization of these objectives.

At the time of creation of Telesat the government clearly indicated its commitment to Telesat autonomy and its conviction that this militated against integration of terrestrial and satellite communications and carrier control of the earth segment of the system. The changing technological environment which renders Telesat a viable competitor to the terrestrial carriers provides even more justification for ensuring the autonomy of the satellite system. The next sections of the paper will explore the consequences of the Agreement in terms of the marketing conflicts of interest and the competitive and regulatory impact of Telesat membership in TCTS.

IV. MARKETING CONFLICTS OF INTEREST

The erosion of Telesat autonomy may be expected to result in Telesat marketing behaviour and planning decisions which favour one customer, TCTS, over others. Another way to describe this result is in terms of a conflict of interest between TCTS and Telesat regarding the facility to be used to meet service requests. The Agreement condones the subordination of Telesat to serve the interests of TCTS, while the financial arrangements which tie the rate of return earned by the former to the financial success of the latter, further distort Telesat incentives. This section will demonstrate the ways in which these marketing conflicts of interest manifest themselves.

One of the concerns expressed by the Commission in its reasons for not approving the proposed Agreement was the potential for subtle hard to detect discrimination in terms of the information provided by Telesat.[29] In fact both CNCP Telecommunications and Broadcast News in their memoranda of evidence submitted for the current CRTC proceeding into TCTS rates and practices, have expressed dissatisfaction with the speed of release of tariff and technical information on Telesat services.[30] In the case of CNCP these delays have prejudiced their ability to compete effectively with TCTS for several contracts to distribute programming.[31] TCTS, on the other hand, would have no such disability given the Agreement's information pooling arrangements and provisions for cstablishment of rates by the TCTS Board of Management.[32] Indeed, it is difficult to accept that Telesat was providing non-discriminatory access to such information and that there was no underlying conflict of interest motivating discriminatory behaviour since Telesat acted on behalf of TCTS in securing CRTC approval of the CBC/TCTS agreement.[33]

Not only do CNCP complaints relate to slow and incomplete responses to requests for information on rates and service availability for existing satellite services, but it is also alleged that preparation of a market plan for satellite services using 14/12 GHz technology has been hindered by lack of information on proposed tariffs for these services.[34] Similarly, the submission by Canadian Press and Broadcast News maintains that unresponsiveness on the part of Telesat to their requests for information on tariffs and possible service restrictions have made it difficult to determine the feasibility of using satellites instead of terrestrial facilities to meet service requirements.[35] A related concern is the confidentiality of information regarding CNCP's service requirements submitted to Telesat.[36] Despite assurances by Telesat that disclosure is in the broadest terms, the necessary involvement of Telesat in TCTS planning activities raised the potential for use of this information by TCTS to secure a competitive advantage in service provision to customers. The need for formal protection against release of information is especially significant in an industry with relatively few end-users. Knoweldge of CNCP's proposed services would allow TCTS to steal customers by offering slightly more attractive terms of provision. As noted by the Commission in its hearing into the proposed Agreement, Telesat would have a greater stake in the financial success of TCTS than of CNCP thereby creating an incentive to engage in activities (including disclosure) which would benefit the former.[37]

These examples suggest that the Agreement has resulted in differential access to information for TCTS members as opposed to all other satellite users. If the autonomy of Telesat was not eroded by membership in TCTS one would expect impartial treatment of all RCTCC's to ensure competitive bidding on distribution contracts involving satellite service. Similarly, one would expect Telesat to engage in aggressive marketing strategies that encouraged novel uses of satellite services; the lack of responsiveness to Broadeast News' interest in switching from terrestrial facilities demonstrates the conflict of interest between TCTS and Telesat.

Furthermore, disclosure of confidential information and delay in the release of rate and service information would not be in Telesat's interest if both TCTS and CNCP were merely customers; rather, the recent behaviour of Telesat indicates an alliance of interest with TCTS with the result that CNCP is forced to discuss service plans with a potential competitor.

Before leaving the topic of information, it is worth mentioning that one of the stated benefits of Telesat membership in TCTS is the free flow of information between the two carriers. While the above illustrations demonstrate that information has flowed from Telesat to TCTS, it is less clear that Telesat will receive the full benefit of TCTS knowledge and experience owing to the desire not to create an effective competitor.[38] Thus the information pooling arrangements involved in joint planning and integration of facilities may not produce anticipated benefits for Telesat, such as encouraging innovative uses of satellite services or allowing for accurate forecasts of satellite demand and efficient utilization of facilities.[38a]

One of the most contentious issues in the marketing of Telesat services reltes to the volume discounts provided for in the proposed Tariff CRTC 8001. CNCP alleges that the combination of bulk discounts and the TCTS/CBC Agreement effectively preclude competition with TCTS in the provision of satellite services to end-users.[39] The discounts as presently formulated are based on the number of full period RF channels supplied to a particular customer, with a maximum discount for customers leasing eleven channels or more. The present capability of the satellite system is such that only one customer, TCTS, can take advantage of the full discount. In particular, of the expected TCTS demand for 6/4 GHz full period RF channels, 5 out of the 11 would be leased pursuant to the CBC/TCTS Agreement. Thus, CNCP is rendered uncompetitive in reselling 6/4 GHz services to customers owing to the preference accorded TCTS in the proposed rate structure.

Telesat has justified its scale of rates on the basis that utilization of satellite services would be encouraged if customers were charged less per channel the more channels they lease; it is submitted by Telesat that an average rate per channel would not produce these same incentives.[40] This argument may be criticized on its own terms. While price discrimination is often suggested as an optimal pricing strategy in industries with high fixed costs, the volume discounts are generally based on 'blocks' of service; that is, rate reductions apply only to incremental units of demand to reflect the lower consumer surplus on the marginal units.[41] It is in this sense that an average rate is sub-optimal: it does not capture the surplus on the inframarginal units while at the same time it eliminates the sale of those units on which revenues would at least cover marginal costs. Furthermore, since average rates do not provide total cost recovery where marginal costs are declining, under the proposed Tariff satellite users would have to be subsidized by users of terrestrial facilities. Thus, declining block prices, designed to create an incentive to utilization at the margin, do not imply rate reductions on infra-marginal units; on the contrary, it is the higher rates on these units which allow recovery of total costs.

In terms of the proposed Tariff, the recipient of the largest discounts, TCTS has demand requirements for satellite services which are not likely to be very sensitive to per channel rates. Aside from fixed CBC service needs, TCTS leases full period RF channels in the 6/4 GHz band for telephone message service. Since this is an activity in which TCTS members have a monopoly and on which the rate of return of each member is separately regulated, there is no incentive to cut costs and prices. Inelastic demand for TCTS monopoly service highlights the inappropriateness of lower average rates per channel to TCTS than to other users. Not only might it be argued that in these circumstances pricing should be based on declining blocks (to capture the consumer surplus) but further that Telesat should be able to take advantage of the market segregation achieved by its membership in TCTS to price discriminate, that is charge higher prices to TCTS. The literature on vertical integration suggests that this strategy (of market separation through integration) is one means of allowing the introduction of price discrimination where the elasticities of demand among various classes of users are different, and the possibility of resale is precluded.[41a] The bulk discounts in the proposed Tariff on the other hand, coupled with the resale provisions (discussed infra) do not realize this potential; rather, the rate structure promotes optimal marketing for TCTS and not for Telesat.

The main benefit which accrues to TCTS from these rates (i.e. bulk discounts) is the ability to market partial channel service to end-users at lower rates than those available to other RCTCC's whose full channel leasing requirements are lower. In this context, it should be noted that price discrimination, whereby discounts were applied to the additional unit leased, would probably not alleviate CNCP's concerns of undue preference to TCTS. Since end-users do not lease satellite services directly from Telesat, and since total TCTS demand is higher than that of CNCP, TCTS would simply attribute the last units, on which the discount was the highest, to those customers thereby ensuring resale of partial channel service at lower prices than those charged by CNCP (in fact, with a maximum price just below that which may be charged by CNCP). While the resale profits may serve to ensure recovery of the full costs of satellite service from satellite users (despite bulk discounts), Telesat should be reaping the benefits from service provision to end-users through a structure of rates which ensures, even absent direct provision by Telesat, that resale profits are not diverted to TCTS from Telesat or other potential providers of satellite services.

Telesat argues that,

> "the increased revenues accruing to the company, as result of encouraging utilization, provide a strong "backbone" of stable revenues which, in turn, permits Telesat to charge lower rates to other customers than would be possible without this support."[42]

While it may be true that rate reductions across the board may be made possible through efficient utilization of the space segment, the rate differentials to different customers are still not justified. An alternative rate schedule suggested by CNCP would base volume discounts on total aggregate use of 6/4 GHz full period RF channel service.[43]

While this form of sliding rates does not conform with the usual notion of block pricing based on customer demand, it would still produce a preferred outcome to either the proposed Tariff or average rates per channel.

Concerns for efficient utilization of satellite service were seen not to be the motivation for the proposed Tariff. Similarly, the policy of leasing only whole RF channels to RCTCC's demonstrates a marketing conflict of interest between Telesat and TCTS. The unavailability of partial channel service directly from Telesat is a consequence of paragraph 13 of Schedule A which provides that while the RCTCC's have the right to market portions of channels, the lease of partial channels must be arranged with TCTS. It is submitted by CNCP that the effect of this requirement is to put CNCP at a competitive disadvantage in terms of provision to end-users.[44] Not only is confidentiality of CNCP's service plans from its main competitor destroyed, but it is also alleged that partial channel service may be provided more cheaply by TCTS. The latter result occurs despite non-discriminatory rates on the leasing of partial channels owing to the potential for TCTS to service this demand as a subset of its own full period whole channel requirements. Since the rates for services are lower for TCTS than CNCP as a result of greater bulk discounts, the end-users are better off dealing with TCTS. In other words, the rate structure and service restrictions benefit a particular Telesat customer, TCTS.

In a memorandum of evidence [45] prepared for the current CRTC proceeding into TCTS rates and practices, Telesat notes that statements made by the Minister of Communications in 1969 supported the policy of leasing not less than a complete television equivalent channel. However, as noted in a corresponding submission by CNCP,[46] the government's position on this policy had changed by 1977. In a statement made in November 1977 by Mme. Sauve,[47] (Minister of Communications at the time) it was made clear that this leasing policy should be reconsidered, in part to address concerns raised by the government's approval of Telesat membership in TCTS. Not only did Telesat choose to ignore the more recent ministerial statements in its submission, but its leasing restrictions entail a misinterpretation of the support for leasing complete television equivalent channels. As noted by CNCP,

> "advances in satellite technology in the last decade make it possible for two complete, undivided television channels to be transmitted on one satellite RF channel. Therefore, Telesat's long-standing policy of leasing only full transponder capacity seems to be outdated not only with respect to government policy but also with respect to satellite technology".[48]

Despite these criticisms of Telesat leasing policy, it is clear that the Connecting Agreement and Memorandum prohibit any other practice; it is less clear that both Telesat and TCTS benefit from the restrictions. In terms of efficient utilization of satellite services one would have thought that Telesat would want to market its services in any form which would facilitate access to the satellite facilty.

While it may be argued that the range of leasing options in the proposed Tariff meets this concern, there is still a minimum leasing requirement of one RF channel for at least one-half hour per occasion,[49] a restriction which is of most benefit to TCTS for the reasons previously detailed. The policy prohibiting users other than RCTCC's from dealing directly with Telesat raises similar objections. While it would not be in the interests of any of the terrestrial carriers to support the elimination of this restriction, it would appear that Telesat's ability to promote novel satellite uses and to market its services effectively would be improved by direct access to end-users. With Telesat playing the roll of a 'carrier's carrier' (contrary to the original intention of Parliament) the division of customer demand between terrestrial and satellite facilities is controlled by the terrestrial carriers.[50] Indeed, as previously discussed, one of the motivations for an autonomous satellite system was the concern that integration of the two facilities would result in limited use of the satellite.

While direct user access would mitigate the problem, the connecting Agreement further constrains competition between TCTS and Telesat for customers in that satellite performance requirements must be established in accordance with TCTS procedures, and satellite design must be compatible with TCTS service plans.[51] In other words the joint planning inherent in an integrated communications network may serve to ensure TCTS control of satellite development even absent user restrictions.

Nevertheless, the current discrimination against non-carrier satellite users (including customers of Telesat prior to the Agreement) should be eliminated so as to encourage competition between Telesat and the terrestrial carriers. It should be noted however that since the RCTCC's comprise a significant portion of overall satellite demand, there will be some remaining ability on the part of the carriers to control utilization of satellite facilities.

Another marketing conflict relates to TCTS control of many aspects of the earth segment of Telesat pursuant to the Connecting Agreement. Provisions include sole right to select sites for earth situations, the option of designing, owning and operating support facilities for the earth sations, and first option to purchase any stations that Telesat decides to sell.[52] These arrangements confer a competitive advantage on TCTS relative to other carriers and provide a means of profiting from the satellite communications system.[53] Investment expenditures on earth facilities would likely be eligible for inclusion in the rate base and would therefore earn the allowed rate of return, whereas the cost of leasing these same facilities would be included in the revenue requirement but only as a reimbursable expense, unless capitalization of leases were permitted.

Furthermore, the configuration of earth stations which would best serve the TCTS network may be very different from that desired by Telesat to meet the requirements of all its customers. The preferential options on design of support facilities and purchase of earth stations are inconsistent with Telesat's interest in awarding contracts for services or sale to the highest bidder. In addition, while it may be most profitable for Telesat to divest itself of certain earth stations, TCTS, as opposed to end-user ownership, would remove an important source of access to customers.[54] In this context it should be noted that in an address to the Royal Society of Canada in November 1977 Mme. Sauve indicated government support for a review of its policy of ownership of earth stations especially in light of Telesat membership in TCTS.[55] Thus TCTS interference with marketing of satellite service through control of earth stations was recognized by the government as a potential outcome of the Agreement. These considerations imply not only the need for removal of TCTS' first option to purchase but also freedom to designate sites, and to construct and own earth stations.

The financial arrangements contained in the Agreement merit discussion in terms of their effect on efficient marketing (utilization and rates) of satellite services. The guaranteed rate of return on common equity was the main financial consideration motivating the Agreement from Telesat's perspective. It was argued at the hearing that a stable revenue base and guaranteed return "coupled with TCTS' commitment to share the financial risks associated with the planned expansion in the satellite facilities would provide accessibility to financing and reduce the cost of capital."[56] The Applicants maintained that these arrangements were necessary to assure availability and expansion of Telesat facilities including the development and marketing of services using 14/12 GHz technology.

The Commission was not convinced that Telesat's current financial position warranted support by and membership in TCTS.[57] Evidence revealed that services using 6/4 GHz technology would continue to be offered absent membership in TCTS, that projections of Telesat's rate of return (assuming non-membership in TCTS) were at least as high as that guaranteed with TCTS membership, and that Telesat's financial performance to date did not indicate cause for concern. As noted by the Commission, "Telesat has repaid it's government loans, has no outstanding long term debt, has been paying a modest dividend and, by virtue of its monopoly on the provision of satellite services in Canada, has reasonable market prospects".[58]

The introduction of services using the new technology was the only area in which the Commission did not reject the need for financial input from TCTS.[59] However, the Commission questioned both the need for 14/12 GHz technology given current user demand, and the appropriateness of telephone subscribers or users of other facilities subsidizing the establishment of these services. Furthermore it was suggested that Telesat policies on earth stations ownership and full channel leasing were the key limitations on increased satellite utilization. As noted earlier in the essay, the incorporation of these policies into the Agreement has provided the opportunity for TCTS carriers to control marketing of and access to satellite services in a manner inconsistent with Telesat's optimal development. Finally, it should be noted that the evidence indicated that, given long run demand forecasts for

14/12 GHz services, the revenue impact on TCTS from establishing these services would be positive. In other words, the financial arrangements may be such that the subsidies are in fact flowing away from Telesat to TCTS, a possibility which merits attention.

The government had indicated its willingness to support the Agreement provided it received assurances that public expenditure on satellite facilities would thereby be avoided. It was therefore in the interest of the Applicants to exaggerate the dire financial position of Telesat as a means of emphasizing the financial commitment being undertaken by TCTS, and corresponding burden being removed from government responsibility. The effect of the revenue-sharing arrangements is to allow a form of rate base averaging whereby the rates for services do not depend on the facility used, but rather are based on the revenue requirement determined by the total costs of all facilities.[61] As noted by Beigie, this plan to finance Telesat, if necessary, through subsidies from telephone subscribers is inappropriate in the sense that such implicit subsidies will distort resource allocation in terms of both the choice of facility (satellite or terrestrial) and the quantity purchased where communications services are used as inputs in the provision of other services such as data transmission.[61a]

On the other hand, while the guaranteed rate of return operates as a minimum commitment from TCTS, returns in excess of the guaranteed rate are shared equally by Telesat and TCTS.[60] Thus, if the satellite services are in fact relatively profitable, the revenue sharing arrangements ensure that the benefit accrues (in the form of a subsidy) to telephone subscribers as well as users of satellite facilities. The speed of passing on economies of satellite use, however, determines the rate of increase in utilization. Since efficient use of satellite facilities depends on high volume, which is in turn a function of lower rates, the sharing of excess revenues inhibits the expansion of satellite use and provides another means of ensuring that Telesat's services are not marketed efficiently.

Optimal marketing strategy on the part of Telesat should not involve subsidizing rate reductions on terrestrial services. However, the financial success of TCTS is of concern to Telesat after 1980 since the guaranteed rate of return is tied from then on to that earned by TCTS.[62] At that point there is a reduced incentive for Telesat to maximize its own rate of return and Telesat's optimal marketing strategy may be to ensure highest possible revenues and rates of return for TCTS. Since everything earned by Telesat above the TCTS guaranteed rate is shared equally, competitive policies which reduce TCTS revenues hurt both Telesat and TCTS. Note however that if satellite services comprise a minor percentage of total TCTS revenues then the effect on the rate of return earned by TCTS (and derivatively Telesat) of this strategy of maximizing TCTS' profit on resale activities may be minimal.

A related marketing conflict concerns allegations by CNCP that rates charged by Telesat for the provision of 6/4 GHz earth station services to TCTS do not cover total expenses.[63] Telesat has responded that there is no cross-subsidization of TCTS earth services by users of the space segment of the satellite; rather, this shortfall relates merely to part of the rate of return on equity and is not made up. It is understandable that with a guaranteed rate of return, rates too low to meet revenue requirements are not of concern to Telesat. However, such discrimination in favour of TCTS is very relevant to terrestrial carriers competing with TCTS and would be of concern to an independent corporation regulated on a'rate base rate of return' basis. Furthermore subsidization of TCTS, through higher rates to other customers for use of satellite services, clearly represents a marketing policy which is not in Telesat's best interests in terms of efficient utilization of its facilities.

A final aspect of the marketing issue involves the reduced competition in the telecommunications industry resulting from the Agreement. Specifically, in terms of the relationship between Telesat and TCTS, it was submitted by intervenors at the hearing into the proposed Agreement that,

> "Restrictions on leasing r.f. channels, on the construction, ownership or operation of terrestrial transmission facilities in TCTS members' territory, and on providing satellite services separately from the Agreement except for specialized space activities unrelated to the business of TCTS were examples of provisions which tended to put Telesat in a position where it could not effectively compete with TCTS." [64]

The leasing restrictions regarding customers and portions of channels have already been discussed. The provisions limiting terrestrial construction and service provision apart from the Agreement are related in terms of the allegation that the integration of satellite and terrestrial networks precludes the emergence of competition between the two facilities. The intervenors recognized that the 6/4 GHz technology was too expensive to emerge as a competitive means of service provision; however, they claimed and the Commission agreed that "14/12 GHz technology offered a real possibility of competition for long distance non-telephone traffic."[65]

The Commission evaluated the likelihood of competition separately for traffic relating to TCTS' monopoly on telephone services and other traffic.[66] While it was recognized that the TCTS monopoly guaranteed control of use of the satellite for telephone traffic, the potential for satellite competition was seen to exists on non-telephone traffic, "particularly the transmission of data and video communications over substantial distances and of TV signals from single point to multipoint."[67] Thus the merger of competing technologies and the prohibition on construction of terrestrial linking facilities within the operating territory of TCTS members were seen as mechnisms for restricting competition in long-haul data, video and other private line services. Given the financial support being provided by TCTS to Telesat pursuant to the Agreement, it is not surprising that TCTS would not want to be put in the position of supporting the creation of a competitor. It was precisely such concerns, however, which formed one of the Commission's public interest grounds for refusing approval of the proposed Agreement.

Another mechanism for inhibiting competition in long-haul transmission relates to the pricing policies adopted by TCTS. The Commission noted in its decision that the rates for long-haul services would be based on an averaging of terrestrial and satellite costs rather than on the costs of the technology involved.[68] At the same time, since satellite costs are distance insensitive, it is in respect of such services that the potential for cost savings from satellite use are greatest. Thus, preventing the realization of economies of satellite use through rate averaging allows TCTS to protect its competitive position vis à vis Telesat.

V. OTHER COMPETITIVE EFFECTS

A discussion of the Agreement would not be complete without mentioning the other effects on the competitive environment in the telecommunications field. Although the majority of these concerns have been alluded to in the discussion of the adverse effects of the Agreement on efficient marketing of satellite services, two of these effects merit explicit attention, specifically the competitive disadvantages for both CNCP (the only RCTCC that is not a member of TCTS), and the non-member non-carrier users of satellite services.

One of the grounds on which the Commission based its decision for denying Telesat membership in TCTS was the undue preference to TCTS and corresponding unjust discrimination against CNCP flowing from the provisions of the Agreement.[69] As noted earlier in this paper, advantages accruing to TCTS by virture of the Agreement include its exclusive right to select earth stations locations, requirements of compatibility of satellite design with TCTS economic and performance requirements and service plans, and timely release of satellite design concepts and other information to facilitate TCTS' planning activities. In addition, CNCP argued that there were other competitive advantages inherent in Telesat membership in TCTS, specifically, "early notice of new network facilities, preferred treatment in gaining access to the network system, better cooperation in clearing up network faults and favoured treatment through subtle biases in network design which would favour the equipment and service offering of TCTS members."[70] Furthermore, participation by Telesat and TCTS in joint planning activities would enhance the expertise of TCTS members, creating a competitive advantage in terms of competition with CNCP for non-carrier customers. Other disadvantages in terms of competition for end-users included restrictions on leasing less than whole RF channels, TCTS pricing policies such as bulk discounts and rate averaging, discriminatory release of tariff and technical information, and confidentiality from TCTS of information submitted to Telesat. A final benefit conferred by the Agreement related to the sharing of excess revenues earned by Telesat: CNCP alleged that this constituted an unfair price rebate for a single satellite user.

The combination of these explicit factors was seen by the Commission as creating a substantial likelihood of undue advantage to TCTS. Moreover, it was felt that subtle forms of discrimination in favour of

TCTS, against which it would be difficult to devise adequate precautions, were likely to occur within an integrated satellite and terrestrial network. Finally, the benefits of joint planning claimed by the applicants were rejected as sufficient justification for Telesat membership in TCTS given the undue preference involved and the evidence that the realization of these benefits did not necessitate the proposed level of integration.

The non-carrier non-member users of satellite service were denied direct access to Telesat under the proposed Agreement.[71] This restriction, covering many of Telesat's customers at the time, as well as potential users such as pipeline concerns and cable television companies, would result in a reduced choice of suppliers and would put these customers in the position of having to discuss service plans with a potential competitor. The rationale for restrictions on access to Telesat is not simply a desire to capture the profits on resale of satellite services to end-users (discussed supra) but lies also in the long-term competitive implications of controlling access to the satellite facility, that is the possibility of forecolsing competition. Specifically, the development of novel communications services which by-pass the terrestrial carriers altogether renders it important to TCTS both to ensure long-term control of its competitor, Telesat, and to be able to contribute to the formation of Telesat marketing policies which operate as barriers to innovative uses of satellite services.[72a]

Another competitive disadvantage relating to leasing provisions concerns the prohibition on marketing portions of RF channels for all but RCTCC's. Considering the significant advantage conferred on those carriers listed in Schedule A and given the evidence received at the hearing, the Commission did not feel these restrictions were justified.

A final competitive advantage accruing to TCTS from Telesat membership in TCTS relates to construction and design of parts of the satellite system.[72] As mentioned in the historical summary, one of the reasons for not permitting integration of the satellite and terrestrial networks was to prevent a lessening of competition between manufacturing subsidiaries of carriers and non-carriers for these construction contracts. Similarly, the Agreement integrating TCTS and Telesat facilities would be expected to produce a disadvantage for potential suppliers other than TCTS insofar as advance notice of expansion plans, as well as greater access to information on technical requirements, would allow TCTS subsidiaries to submit bids and develop proposals in advance of its competitors. While it may be argued that rules regarding timely disclosure of information might be instituted, the enforcement costs associated with determining whether or not Telesat was complying with these rules might be prohibitive.

VI. REGULATORY ISSUES

A final problem created by Telesat membership in TCTS concerns the effect on the Commission's ability to fulfil its regulatory responsibilities in the telecommunications field, specifically the regulation of rates and the adjudication of complaints of unjust discrimination.[73] In terms of determining the justness and reasonableness of rates, it was argued at the hearing that the guaranteed return and revenue sharing arrangements would render separation of the costs of satellite and other services difficult to achieve. Thus, it would be impossible to assess the extent to which satellite customers were receiving the benefit, in terms of rate reductions, of any economies of satellite use. Similarly, the Commission would be unable to evaluate whether the rate structure was hiding subsidies by telephone subscribers to satellite customers so as to ensure that Telesat received its guaranteed rate of return. At any rate, the inability to identify the costs and revenue requirements of each facility would serve to discredit the Commission's determination as to the reasonableness of rates.

The effect of the leasing restrictions imposed by the Agreement similarly undermines the Commission's responsibility to review the rates proposed by Telesat. As stated by the Commission,

> "Since it would in effect become a carrier's carrier, the tariffs for telecommunications services to the public partly or wholly based on the use of satellite facilities would be filed by the terrestrial carriers with their respective regulatory agencies for approval. ... No single agency could therefore fully review the principles underlying these rates, their related costs or their overall effects."[74]

The absence of such review makes it difficult to ensure that the cost advantages of satellite communications, which are especially significant in long-haul transmission, are being passed on to satellite users in terms of appropriate rates. Rate averaging for terrestrial and satellite services would be hard to detect given the integration of the facilities and associated cost separation problems noted previously.

The Commission was also concerned that Telesat membership in TCTS would eliminate the adversarial nature of Telesat rate proceedings.[75] The incentives to effective intervention by TCTS are significantly reduced by the splitting of excess revenues while the non-carrier users, prohibited from dealing directly with Telesat, would not be in a position to intervene in a useful way. The reduced quality of intervention, in terms of disputes over appropriate rates for satellite services, renders the Commission's task of determining the reasonableness of proposed rates that much more difficult.

The Commission's ability to make adjustments in Telesat's revenue requirements and to set the allowed rate of return are undermined by the guaranteed rate of return provided by TCTS. Transfer payments from TCTS would ensure that Telesat earns the minimum rate of return set out in the Agreement and not the lesser one deemed appropriate by the Commission. Similarly, to raise the return above the guaranteed level, the Commission recognized that it would have to "set rates high enough to take account as the fact that half the surplus over the amount approved in the Agreement would have to be paid to TCTS."[76] These cash flows are clearly inconsistent with the agency's responsibility for regulation of rates, especially given that the basis for the chosen guarantees would not be subject to examination and review by the Commission.

The effectiveness of the regulatory process as a forum for adjudication of complaints of undue perference in also open to question as a result of the Agreement.[77] The integration of the two facilities has been noted as creating the potential for subtle forms of discrimination in favour of TCTS. Furthermore, the circumstances allowing these advantages to accrue also make specific instances of undue preference hard to detect, and complaints or allegations difficult to prove.

VII. INSTRUMENTS OF POLITICAL CONTROL

Despite the concerns raised in the preceding pages, the Cabinet varied the decision of the CRTC and approved the proposed Agreement. The Minister of Communications, speaking on behalf of the Government shortly after the decision, indicated that despite some potential regulatory difficulties arising from the intergration of Telesat and TCTS, the regulatory powers required by the Commission for effective regulation in the new situation already existed.[78] At the same time it was conceded that " it might be necessary to devise new regulatory techniques to meet the complexities introduced by the association."[79] The preceding section has demonstrated the truth of this statement given the extent to which the Commission's ability to carry out its regulatory responsibilities has been impaired by the Agreement. The Minister argued that increasing integration and associated regulatory complexities were inevitable in the telecommunications field; however, it seems clear that is was the level and not the fact of integration which the Commission sought to control. The Commission indicated in its decision that any of the desired benefits were attainable without the proposed degree of integration. For example, joint planning committees might be established to consider means of eliminating duplication of facilities and to ensure compatibility of designs for linkage purposes.[80]

The competitive and financial considerations mentioned by the Minister as grounds for approving the Agreement have already been discussed.[81] The government felt that the realization of these objectives would be prejudiced if Telesat were left to the "vicissitudes of the competitive domestic environment";[82] however, the Commission had examined the financial position of Telesat and concluded that there was no demonstrated need for financial support. Moreover, as noted previously, the Agreement would tend to inhibit the development of new satellite uses owing to TCTS' control of the choice of facility and of Telesat marketing policies. Thus, it seems that the Government's decision was based on considerations which, in the Commission's opinion, had been demonstrated at the hearing to be unfounded.

This discussion highlights the extent to which Cabinet appeals represent an inappropriate interference with the regulatory decision-making process. Mme. Sauve, in an Address to the Royal Society of Canada shortly after the Cabinet decision, stated,

> "By approving the Agreement, I believe that the Government has reaffirmed its confidence in the professional skill and competence of the Commission and its staff to deal with the regulatory matters, within its jurisdiction." [83]

It is difficult to understand how respect for the Commission's skill and competence would justify varying one of its decisions; rather, such action tends to undermine the credibility of any agency and in the particular circumstances, reduces the effectiveness of rate proceedings and the Commission's ability to fulfill its duties regarding the regulation of rates and the prevention of unjust discrimination.

It is not being suggested that agencies remain immune from outside influences; however, coordination between government and tribunal activity, and government input into regulatory decisions, are best achieved through ex ante policy directives as opposed to ex post facto Cabinet appeals. These two policy instruments differ in their underlying view as to the appropriate role of both agencies and government in the regulatory process and specifically, the proper balance between control of agencies by elected officials and the need for autonomy in day-to-day adjudication by tribunals to ensure continuity, experienced decision-making, impartiality and insulation from the political process.

The binding ex ante policy directives proposed in the Economic Council of Canada Report entitled Repsonsible Regulation have the following characteristics: [84]

> (i) Such Policy Directives would be issued by the Governor (or Lieutenant Governor) in Council.
>
> (ii) They should be tabled in the legislature (hence facilitating debate) and published in a gazette.
>
> (iii) They would apply only to general policy questions and not to individual cases.
>
> (iv) A Policy Directive could not be issued once an SRA has begun proceedings on a specific case.
>
> (v) Policy Directives should be preceded by public hearing if possible.

The advantages of holding a public hearing and producing a public report prior to the issuance of a policy directive include direct participation by affected interests, open proceedings and decisions based on information presented and publicly available. Moreover, this method of policy formation adopts the perspective that regulation is a political process whereby brokerage among competing interest groups should determine regulatory outcomes.[85] Thus, decisions are most responsive to and best match the public interest if all affected interests participate in the process. Furthermore, access to the policy-making process is enlarged while the ability of powerful interests to influence the decision is reduced, both of which imply a more equitable outcome.

Government participation in the policy hearing provides the requisite input for coordination purposes and ensures that government views are openly stated and consequently not formed in response to undisclosed pressure from particular interests. At the same time the requirement of final Cabinet adoption/modification/rejection of a directive arising out of a public hearing recognizes that the ultimate policy choice is a political decision and as such should be made by elected officials who are accountable to the legislature and ultimately the electorate. Finally, policy directives are likely to enhance agency autonomy and credibility, insofar as carefully reasoned decisions, based on evidence received at a public hearing, are no longer subject to reversal by government.

Political appeals to Cabinet, on the other hand, since they occur after a decision is made, are disruptive and render a waste of resources consumed in the hearing process.[86] At the same time accountability is selective at best and may be more imagined than real; intervention tends to occur in response to narrow interest group pressures and tends to reflect a concern for short-term political ends. Other disadvantages relate to the demoralization of the agency owing to the possibility of arbitrary reversal by Cabinet. The undermining of agency credibility may be expected to result in reduced quality and limited scope of agency decisions as tribunals grow lazy or try to second-guess the politicians.

A final set of considerations relates to the lack of procedural regularity in the appeal process.[87] The Cabinet is not limited to issues considered by the agency nor are there requirements of participation by affected interests, disclosure of departmental submissions or publication of reasons. While the Inuit Tapirisat[88] decision imports 'fairness' into the process, it should be noted that such concerns are really only relevant in an adjudication of rights, which is the function most appropriately left within the sphere of agency autonomy.

The benefits of political appeals relate to the incorporation of broad public policy concerns into the decision-making process.[89] Accountability and legitimacy are ensured through ministerial responsibility to Parliament and the electorate no matter how remote the regulatory decision may be from ultimate election issues. However, the self-interest of politicians and differential access to them are two potential sources of distortions in the responsiveness of regulatory decisions to the public interest. In the absence of other safeguards, Cabinet appeals may ensure against abuse of procedure and excess of jurisdictions; however, the courts and not the government would seem to be the appropriate forum for redress against arbitrary action. Finally, the distributive consequences of appeals are inequitable in that organized well-entrenched interests tend to receive preferential treatment.

Cabinet appeals, as compared to policy directives, therefore represent an inferior means of ensuring accountability and an inappropriate mechanism for political control over and transmission of government policy to regulatory agencies. In the context of the particular decision to allow Telesat to join TCTS, it would appear that policy directives might have provided a useful forum for articulation of the government's priorities in the telecommunications field. For example, the introduction of new satellite technology, the protection of employment in the Canadian aerospace industry, the maintenance of an autonomous satellite corporation providing services on a commercial basis, and the preservation of Canada's position in a high technology field illustrate relevant policy concerns which might be embodied in directives. At the same time a determination of the merits of particular applications would be left to the Commission, which would perform its adjudicative function in light of these policy directives.[90] In this manner, respect for agency experience and expertise in evaluating conflicting technical evidence would be assured and agency credibility enhanced.

Thus, to the extent that there is a "range of factors ... far wider than that which the CRTC could reasonably have been expected to consider",[91] the government may ensure through directives that these broad policy issues are brought to the Commission's attention before it makes a decision. In particular, since the government's stated policy concerns were considered by the Commission in its review of the proposed Agreement, the original CRTC decision would now be in effect and Telesat would not be a member of TCTS.

VII. CONCLUSION

This paper has attempted to demonstrate the adverse consequences of the Cabinet's approval of the Agreement providing for Telesat membership in TCTS. Various dimensions of the effects of the Agreement on the telecommunications industry have been explored: the impediments to efficient marketing and development of satellite services; the undue preference provided to TCTS and consequent impact on CNCP and non-carrier users of satellite facilities; and finally, the effect on the regulatory environment in terms of the impaired ability of the Commission to carry out its regulatory responsibilities and the potential for Cabinet appeals to undermine the agency's credibility and autonomy. Given this catalogue of undesirable aspects of the Agreement, it seems appropriate to devote the conclusion of the essay to an evaluation of one proposal for minimizing and perhaps nullifying its application.

The Agreement itself provides in paragraph 24 "[t]hat nothing shall be binding in this Agreement which may override or conflict with any Act of the Parliament of Canada or any Province thereof." Moreover, it is expressly stated in paragraph 2 of the Memorandum of Agreement that it too "shall be considered a part of the TCTS connecting Agreement". That same paragraph provides that Telesat, on joining TCTS, will become a party to the TCTS Connecting Agreement and will abide by the 'terms, conditions, processes and arrangements' laid down by it.

These arrangements suggest an examination of the relevant regulatory legislation to ascertain whether in fact the terms of the Agreement and Schedule A violate some statutory provision; by the Agreement's own terms, its provisions will not be binding where such a conflict is found. Sections 265, 320 and 321 of the Railway Act[92] provide just such a possibility. Section 321(1) requires that telegraph or telephone tolls be just and reasonable and that equal charges be assessed for all traffic provided on simil terms and carried over a particular route. Subsection 2 prohibits a 'company' from making any unjust discrimination or conferring an undue preference in respect of tolls, services, or facilities. The definition of 'company' is contained in section 320(1) and includes a telegraph or telephone company or any company authorized to construct or operate such services. Since Telesat does carry telegraph and telephone traffic, it will be assumed for the purpose of this discussion that it is a 'company' within the meaning of s. 320(1) the Railway Act. Similarly, s. 265 requires all railway companies to afford all persons and companies reasonable and proper facilities for the receiving, forwarding and delivery of traffic. Note however that the section refers specifically to a railway company; it is therefore not clear that this particular provision applies to Telesat.

The statutory provisions dealing with remedies are found in ss. 45 and 46 of the National Transportation Act.[93] The Commission is granted jurisdiction to adjudicate complaints relating to violations of the Railway Act and to order compliance with its provisions and or prohibit the impugned practice. Specifically, the Commission[94] may issue a declaration indicating which provisions of the Agreement are in conflict with the Railway Act and declaring them inoperative and not binding upon the parties. In addition, the order may declare invalid any provisions contained in tariffs which embody provisions of the Agreement declared inoperative.

Turning to potential violations of the Railway Act, it has been indicated throughout this paper that the provisions of the Connecting Agreement and Memorandum allow Telesat's rate structure and conditions of service provision to favour TCTS and discriminate against other users of satellite services. The specific provisions of the Memorandum of Agreement open to challenge include those relating to compatibility of satellite performance requirements and designs with TCTS procedures and service plans (paragraphs 6(b) and 6(c)); information pooling arrangements between Telesat and TCTS (paragraph 6(d)); and TCTS designation of sites for earth stations and first option to purchase them (paragraph 7). Objection also may be taken to paragraph 8 giving TCTS first option to design and own support facilities for earth stations, and paragraph 10 providing TCTS with the first option for on-site maintenance operations at earth stations intended for use in service provision to non-TCTS members. Other provisions which may be invalid are those prohibiting Telesat from building, owning, or operating terrestrial transmission facilities (paragraph 11), the leasing restrictions relating to the type of customer and the nature of service offered (paragraphs 12 and 13), and the limitations on independent contracting and billing (paragraph 15).

In terms of rates, the integration of the satellite and terrestrial facilities is expected to be accompanied by pricing policies based on an averaging of rates and not on separate charges depending on the facility used. On this basis, it may be argued that Telesat membership in TCTS inherently entails a contravention of section 321 of the Railway Act; that is, Telesat's rate structure will necessarily discriminate in favour of TCTS because of the integrated network and the pricing policy of the TCTS Board of Management based on rate averaging. On this basis, the proposed Tariff CRTC 8001 which applies these discriminatory provisions, would seem to be inoperative.

This discussion suggests that it may be possible to have the major provisions of the Connecting Agreement and Memorandum declared to be in violation of the Railway Act and therefore, by the Agreement's own terms, not binding upon the parties. In light of the considerations raised in this paper, it is to be hoped that a successful application for an order to this effect is brought in the near future.

FOOTNOTES

1. Telesat Canada, Proposed Agreement with Trans-Canada Telephone System (1977), 3 C.R.T. 265.

2. Order in Council P.C. 1977 - 3152, November 3, 1977.

3. Canada, White Paper on a Domestic Satellite Communications System for Canada (Ottawa, 1968), p. 10 (Hereinafter referred to as the White Paper) Found in Tupper (23), p. 86.

4. Tupper (23), pp. 84-85.

5. Ibid., pp. 82-83. Dalfen (6), pp. 199-200.

6. The following arguments are based on Tupper (23), pp. 88-92; Dalfen (6), pp. 187-189.

7. Dalfen (6), pp. 192-193, Tupper (23), p. 81, 92.

8. Dalfen (6), p. 194, Tupper (23), p. 84, 92.

9. Trebilcock and Prichard, (21)

10. Tupper (23), pp. 95-97.

11. Tupper (23), p. 93; Dalfen (6), p. 203

12. Tupper (23), p. 93; Dalfen (6), p. 203.

13. Dalfen (6), p. 202.

14. Booth (3), p. 25.

15. Dalfen (6), p. 204, 206-209; Tupper (23), p. 88.

16. Dalfen (6), p. 201.

17. In fact, Schwartz (20) argues that carrier participation in Comsat allows subsidization of users of terrestrial facilities by satellite customers and that this rate structure precludes efficient (high-volume) use of satellite services.

18. Tupper (23), pp. 93-95; Dalfen (6), pp. 205-206.

19. Found with the Telesat Canada Telecom Decision, 77-10 (25).

20. Connecting Agreement (25), paragraph 3, 4.

21. Memorandum of Agreement (28), paragraph 2.

22. Ibid., paragraphs 7 to 10.

23. Ibid., paragraph 11.

24. Ibid., paragraph 13.

25. Ibid., paragraph 17.

26. Ibid., paragraphs 23, 24.

27. Telesat Canada (25), pp. 14-15.

28. Jeanne Sauve, Satellites Are For People (24).

29. Telesat Canada, (25), pp. 36, 39.

30. CNCP Telecommunications (26); Canadian Press and Broadcast News (27).

31. CNCP Telecommunications (26), pp. 4-5.

32. Connecting Agreement (28), paragraphs 3, 4 and 6.

33. CNCP Telecommunications (26), p. 5.

34. Ibid., pp. 4-5.

35. Canadian Press and Broadcast News (27), pp. 9-10.

36. CNCP Telecommunications (26), pp. 11-13.

37. Telesat Canada, (25), p. 39.

38. Schwartz (20), p. 479 makes this argument in the context of American terrestrial carrier participation in Comsat. Note that the information economies expected to flow from vertical integration will not hold in this context owing to the potential for Telesat to operate as a competitor as well as a supplier to TCTS. See Carr and Halpern [5a].

39. CNCP Telecommunications (26), pp. 6-7.

40. Telesat Canada (29), p. 15.

41. Ontario Hydro Electricity Costing and Pricing Study, Vol. 6, "Alternative Objectives for Pricing", (1976).

41a. Halpern and Carr

42. Telesat Canada (29), p. 15.

43. CNCP Telecommunications (26), p. 7.

44. Ibid., pp. 8-10.

45. Exhibit No. Telesat -80-10 (29) p. 3.

46. CNCP Telecommunications (26), pp. 8-9.

47. Satellites are for People (24), p. 10.

48. CNCP Telecommunications (26), p. 9.

49. Telesat Canada (29), p. 15.

50. Similar objections were raised by Schwartz (20) regarding the prohibition on Comsat dealing directly with ultimate users.

51. Memorandum of Agreement (25), paragraph 6.

52. Supra, n. 22.

53. These considerations are mentioned by Schwartz (20), p. 447 as reasons for the interest of U.S. terrestrial carriers in earth stations ownership.

54. Schwartz (20), p. 469.

55. Satellites are for People (24), p. 10.

56. Telesat Canada (25) p. 48.

57. Ibid., pp. 49-50.

58. Ibid., p. 50.

59. Ibid., pp. 50-52.

60. Memorandum of Agreement (25), paragraph 18.

61. The following argument is based on Schwartz (20), pp. 471-473.

61a. BEigie [2]. See also English [9].

62. Memorandum of Agreement (25), paragraph 17.

63. CNCP Telecommunications (26), p. 11.

64. Telesat Canada, (25) p. 53.

65. Ibid., p. 54.

66. *Ibid.*, pp. 55-56.

67. *Ibid.*, p. 56.

68. *Ibid.*, p. 30.

69. *Ibid.*, pp. 34-41.

70. *Ibid.*, p. 36.

71. *Ibid.*, p. 41, 53.

72. Schwartz (20), p. 474 notes similar advantages for the Americam terrestrial carriers who are all major equipment suppliers.

72a. Carr and Halpern [5a]; Kahn [14] Ch. 4, *Proposals for a Communications Policy for Canada* [5], p. 22; Irwin, *The Telecommunications Industry* [10] Ch. 7; *Kayson and Turner* [15].

73. *Telesat Canada*, (25) pp. 26-33.

74. *Ibid.*, p. 29.

75. *Ibid.*, p. 30.

76. *Ibid.*, p. 32.

77. *Ibid.*, p. 41.

78. *Satellites are for People* (24), p. 11.

79. *Ibid.*,

80. Schwartz (20), p.

81. *The Agreement* supra, pp.

82. *Supra*, n. 78, p. 11.

83. *Ibid.*,

84. Economic Council of Canada (8), p. 65.

85. Doern (7); Reich (18), Trebilcock et al (22).

86. Janisch (12).

87. Economic Council of Canada (8), ch. 5; Janisch (12).

88. *Inuit Tapirisat* v. *Leger* (1979); 24 N.R. 361 (Fed. Ct. of Appeal); leave to appeal to the Supreme Court of Canada granted February 9, 1979.

89. Economic Council of Canada (8), Ch. 5; Janisch (12).

90. Any lack of flexibility imposed ontthe regulatory process as a result of having clearly articulated policies would be justified; however, any concerns for undue fettering of discretion might be met by a reserve clause willingness to consider the exceptional. See Janisch (12).

91. Statement by the Minister of Communications, Jeanne Sauve in Respect of Order in Council PC 1977 - 3152; November 1977 in Janisch (12), p. 69.

92. R.S.C. 1970, c. R-2.

93. R.S.C. 1970, c. N-17.

94. Commission in s. 2 of the National Transportation Act refers to the C.R.T.C. and not the C.T.C., National Transportation Act, R.S.C. 1970, c n-17 as amended by S.C. 1974-75, c. 49, s. 18.

/

ON THE GENERAL IMPOSSIBILITY OF COMMUNICATIONS MONOPOLY

ALAN BAUGHCUM

Charles River Associates

This paper briefly reviews the recent history of communications monopolies in the U.S. and other countries. That review documents the growth of competition in equipment supply, in communications transmission, and in new service offerings. The forces leading to the introduction of this competition are identified. The paper argues that those forces are so pervasive and so powerful that there is simply no room left for traditional monopoly provision of services in communications. In particular, provision of telephone services at the local level, the distribution of the mails, and over-the-air transmission of television signals face an evermore competitive future. Even if regulatory authorities seek to maintain and protect traditional monopolies, their efforts are likely to be largely in vain. Regulatory authorities will find themselves unable to stem the tide of competition. Of the greatest public concern is the possibility that regulatory intervention may distort the movement of telecommunications systems toward their most efficient organization. One of the ways in which this distortion may well occur would be the imposition of particular institutional structures on the providers of telecommunications services. While the paper will not offer definitive judgements on the advisability of imposing "separate subsidiary" structures, divestiture, or other organizational proposals, it will suggest that there are dangers to replacing organizational responses to market forces by regulatory fiat. The paper closes with a call for research into whether regulatory policy is even required in the area of organizational structure; in particular, the wisdom of proceeding with specific policy measures in this area without such research is questioned.

Consider the typical American household in 1960. Communications were available in the form of over-the-air broadcasting of television and radio signals, newspaper distribution, distribution of personnal and financial correspondence as well as book and magazine distribution through the mail, and through access to the telephone network. In the 20 years since 1960, technological changes in communications have made it possible to multiply the services available to the household and have made the delivery of those services possible through largely electronic media.

The American household is no longer merely a passive recipient of over-the-air broadcasting signals; two-way cable makes it possible to respond to those signals. Programming options available to the household have increased as cable, MDS, and STV have become realities. In the future, the availability of low power television broadcasting signals point to even greater freedom of choice for the consumer. Additionally, the presence of video cassette recorders and video discs will increasingly allow the consumer to be his own programmer.

The increasing availability of cable also makes it possible to forego traditional, labor intensive, distribution of newspapers, magazines, and books. Indeed, I recently learned of the availability by way of two-way cable of a new encyclopedia offered by Arete Publishing. This availability is at the moment limited to the QUBE experiment in Columbus; as other urban areas in the U.S. award franchises to cable operators, availability will expand. One can hardly read any issue of the communications trade press without seeing reports of experiments by which newspapers are offered over cable systems. Similarly, Time magazine has plans for offering an experimental electronic magazine.

The success of Tandy Inc. and Apple in offering home computers combined with cable penetration offers great hope to an entirely new industry offering data bases of one sort or the other. Dow Jones offers the home consumer programs which will assist in managing investment portfolios. Attorneys can access data bases containing case law from Mead Data, owner of LEXIS.

I am now free to purchase my own telephone equipment and am not required to depend on the Bell system to lease me telephone sets. Types of equipment that I can purchase from non-Bell suppliers or lease from the Bell System have multiplied. Answering devices, call forwarding systems, wireless telephones, etc., are now available. Additionally, I have the choice of several companies in addition to the Bell System over whose facilities I may make interstate phone calls. (These companies are Southern Pacific Communications, Western Union, MCI, and the USTS.) Additionally, I may pay my phone bills out of my checking account with a local bank. The availability of services and new companies offering such services is, if anything, even more dramatic in the business sector. Indeed, one consultant's recent report indicated that the majority of Fortune 500 companies in the U.S. would by the mid-1980's have their own electronic mail systems. The availability of long distance communications networks through General Telephone and Electronic, Xerox, and SBS, has been widely reported.

In the excitement over the availability of new and innovative services currently available and speculation about future developments, it is important not to lose sight of a basic fact. That basic fact is the gradual lessening of dependence on the monopoly form of industry organization for the delivery of communications. The American household in 1960 was dependent on over-the-air broadcasting established via spectrum allocation at the FCC which effectively meant that most areas of the U.S. would have three television signals available to it. Cable TV was not widely available largely due to regulatory barriers. The postal system has long had a monopoly on delivery of most of the correspondence to the home. The monopoly element in telephony was largely unchanged from the early part of this century until the 1970's.

It is the primary thesis of this paper that the effective weakening of these monopolies was no accident. Nor should these developments be viewed as the result of the inexorable forces of technology which develop and change economies in a fashion exogenous to the economic system. Technological change, while it occurs for a variety of reasons, has caused the considerable restructuring of the U.S. economy and will cause greater restructurings in the future due to the prospect of the availability of large profits to be earned by implementing new technologies.

By concentrating on the availability of profit and the desire of new and existing firms to appropriate those profits, it becomes possible to analyze the development and impact of technological change. Were we to accept technological change as entirely exogenous and a result of factors totally out of our control, it would be difficult to draw useful lessons of public policy from such technological change. However, linking technological change to the desire to appropriate profits makes it possible to make some general statements about the future and about appropriate policies.

The first lesson to be drawn is that the U.S. economy is not unique. The desire to appropriate profit is apparently present in virtually all of the developed countries in the world. Certainly there appears to be no shortage of U.S. businessmen who are willing to invade other countries in order to demonstrate their own dedication to the profit motive. Japanese businessmen appear to be quite aggressive throughout the world in a variety of industries. The Japanese penetration of communications markets in this country and elsewhere is well documented. Similarly, American businessmen have returned the favor by seeking to break into the domestic Japanese economy; recent negotiations between the U.S. and the Japanese governments regarding telecommunications equipment appear to make this increasingly easy in the future.

The desire of new companies to appropriate profits leads them to seek ways of invading the sales territories of existing monopolists. Existing monopolists, also motivated by the desire for profit, seek ways to expand their product lines and are increasingly led into more competitive arenas. This has occurred to some extent in Canada. Economic pressures in England have led to the restructuring of national telecommunications in that country that offers the hope for some greater competition in telecommunications there. Similar forces appear to be at work in Germany. An interesting element of the pervasive effect is that regulatory authorities may be able for a while to use the threat of competition from the U.S., for example, as a lever to force changes in the behavior of their own domestic industries. For example, it is conceivable that European PTT's could use the threat of entry as a means of coaxing lower rates for telecommunications service or lower prices for telecommunications equipment from manufacturers. However, there is the longer term danger that regulators will attempt to regulate on the basis of their forecasts of the means by which technological change and less restricted entry might impact the economy. To the extent that the regulator misforecasts, that economy will incur costs as technological change is channeled in directions that market forces might lead it.

The second lesson to be drawn is that society is generally better off when it chooses to rely on industry structures other than those of monopoly. In a recent book entitled The Economics of Competition in the Communications Industry, I and my coauthors argued that this was certainly the case with regard to telephony in the U.S. Various public policy goals including encouragement of technological change were promoted by less restricted entry and freedom to price experimentally. Furthermore, it seems clear that the goal of universal service and static efficiency would not be adversely affected by multiple supplier structures. Finally, it seemed likely that more desirable rate structure characteristics would have no deleterious effect on income distribution objectives.

With regard to the postal service, it's clear that a variety of users have found satisfaction by using their own delivery system or by relying on carriers other than the U.S. postal system. The United Parcel Service is one example of a major success story being built on offering superior service and generally lower rates than those charged by the postal service. Other examples would certainly include express mail services offered by such firms as Federal Express, Emery Express, and others. Additionally, various users of the postal service have experimented with adoption of their own delivery systems for bills or periodical distribution.

With regard to home entertainment, American consumers are beginning to experience alternative programming from cable operators, pay cable programmers, MDS, STV, video discs, and video cassettes. While the audience ratings and shares for the three major networks in the U.S. have been dropping over the past several years, that rate of decline has been modest and profits do not appear to have been severely affected at this point. Much of the monopoly element in home entertainment has stemmed from regulations promulgated by the FTC with regard to spectrum allocation and programming requirements. As these restrictions have been loosened, the consumer is facing increasing diversity of home entertainment possibilities. Indeed, programming which once was viewed as unprofitable will now be offered by firms who are not dependent on government funding. Both ABC and CBS as well as other companies have announced their intention to fund cultural and performing arts programming to be offered on cable TV.

A third lesson to be drawn is that remaining monopoly elements in the communications industry will in all probability be eroded by market forces over the long run. Even where those monopoly elements appear to be the strongest, in local telephone exchanges, the word monopoly is already inappropriate. The existence and pervasiveness of CB radios constitutes at least one very direct substitute for local telephone company operations. I don't wish to pretend that such radio communications are as of yet a perfect substitute nor do I wish to be pinned down on just how close a substitute they are. Nonetheless, the existence of CB radio, the availability of two-way cable, experiments with cellular radio, and the existence of direct commercial substitutes such as those offered by SBS and XTEN all point to the inappropriateness of considering local telephone company operations as a communications monopoly.

The postal service monopoly may also be doomed in the long run. Despite the considerable political power of postal worker unions and various users of postal services, the election of the Reagan Administration may very well mark the beginning of the day in which public tolerance of subsidized industries is being rapidly diminished. If the deregulation fever is sweeping most other sections of the regulated American economy, there seems no reason to believe that the postal service will remain immune over the long run. Indeed, one might view the desire of the postal service to break into electronic mail systems as evidence of their own perception that public tolerance for subsidization will disappear. Electronic mail as a diversification alternative offers the postal service a way out of its currently labor intensive operations while maintaining a stronghold on future revenue streams.

The fourth and final lesson which I would share with you might well be termed a "concern" rather than a lesson. This concern arises because I note little or no enthusiasm or interest on the part of state and local regulators for deregulation of communications. Indeed, the traditional preoccupation of state and local regulators with limiting the profitability of regulated utilities seems to continue unabated. The notion that substitutes might exist in sufficient number and variety as to make possible the loosening of regulation for local telephone companies or local cable companies for that matter is not an idea that has penetrated the collective psyche of the state and local regulatory community. This is critically important because of the diametrically opposed interests at stake. If technological change in telecommunications is being driven by the profit motive and if regulators are motivated by the desire to restrict the profits earned by regulated utilities, then there is the very real danger that technological change will be impeded and perhaps seriously distorted. As the tenor of this paper suggests, I don't believe that technological change can be stopped; as a matter of practical and economic reality I think most strongholds of telecommunications monopoly, regulated or not, will diminish in importance overtime. Nonetheless, the regulatory community has shown time and again an ability to impose inappropriate restrictions upon regulated industries that tend to cost the consumer more than any apparent benefits which derive from such policies. ICC regulation of rail and trucking, DOE regulation of energy allocation, and CAB regulations of airlines are clear examples of innapropriate regulation.

The question of conflict between regulatory interest and the ability of the industry to adapt to technological change is critical in telecommunications because it is technological change and the dynamic nature of the industry that is characteristic. The standard static economic model of a regulated natural monopoly is not the appropriate image for this industry. If that image is rejected in favor of the image of an unfolding movie, then regulators should be particularly modest in attempting to forecast technological developments by the end of the movie based on scenes in the first few frames. It is not at all clear that regulators will be so modest.

In particular, regulators seem to be enamored of imposing separate subsidiaries on telephone carriers in order to secure the advantages of competition while avoiding the dangers of anticompetitive behavior. Separate subsidary proposals seem to have the advantage of being politically acceptable. Yet, I know of no substantial body of research which would document the advisability of such structural impositions on public policy grounds. I should hasten to add that I know of no substantial body documenting great harm either. Yet if one views technological developments in telecommunications with great optimism and sees the industry moving in an increasingly competitive direction, then the imposition of particular structures upon firms in the industry by regulatory fiat seems to be a move that should be undertaken only after great analysis.

If a vertically integrated and tightly controlled firm is one that has developed as a result of market forces as the best vehicle leading to profit maximization at particular risks levels then the imposition of an arm's length subsidiary may well impose more costs than benefits. Even if the vertically integrated monopoly structures which we traditionally observe in telephony have arisen for reasons other than market forces, then why not rely upon market forces to push the firms in the directions of needed adjustment. Regulation simply may not be needed to anticipate the best approach to the applications of existing and pending technological change in an environment of unblocked entry.

Let me briefly sketch one way in which separate subsidiaries may be an unwise public policy. The direction of technological change in telecommunications is toward the merging of a variety of competing and diversified forms of communications into one digitalized stream. Traditional distinctions between voice and data, between entertainment and telephony, between newspapers and television, etc., are simply going to be inapplicable in the future. The ability of telecommunications firms to respond to this merging will depend importantly upon their ability to move across a variety of transmission technologies. In particular, telephone companies should be free to utilize cable TV where appropriate. Similarly, cable TV operators should be free to offer switched communication services. Barriers to transmodal ownership should be largely eliminated. Yet the imposition of separate subsidiaries, even in the absence of barriers to transmodal ownership, may very well restrict the ability of existing firms to utilize capital resources and access to R&D that would benefit the consumer by achieving such cross-technology combinations. Inappropriate regulatory strictures offer the prospect of imposing costs at least as great as the benefits which they purport to achieve.

In summary, my series of lessons amount to a suggestion that we should let the market operate. If problems develop and if those problems are of sufficient magnitude to call forth regulation, then I would not be opposed to such regulation. Imposing particular structural forms on rapidly developing markets seems to me to be inappropriate. Inappropriate because such imposition takes place on the basis of inadequate or incomplete or nonexistent analysis and inadequate simply because no one has the ability to forecast what's likely to happen in the rapidly developing telecommunications sector.

Revised Apr 29/81

Should the traditional service monopoly of the Telecommunication Administration be restrained ?

Jürgen Müller

1) Introduction

The field of telecommunication is an interesting area of study for economists worried about the extent of competition and regulation in the economy. I am speaking about the question of organizational designs : setting the rules and incentives for players in a sector in such a way that the available technology and resource potential is efficiently utilized, the boundaries of technology are pushed forward and the actions of users and producers are constrained to a minimal amount (to secure user and producer freedom). To raise this question of organizational design in the telecommunication sector is especially relevant because of the unusually strong government influence on the publicly owned service suppliers (the PTTs) on one hand, and the rapid change in technology on the other. The latter seems to make the traditional concept of the service monopoly obsolete.

Increased competition in the telecommunications sector, at its most pronounced in the USA, is also in prospect for Canada, the UK, and, to some extent, France and West Germany. The effects of deregulation, which originated in the USA, are certainly being felt in Europe as well. Traditional legal monopoly legal concept, the dominant

factor in telecommunications, is therefore no longer the universally accepted concept. What can be learned from the experience in the USA, and what lesson can be drawn for such policy changes as are contemplated in Europe?

I am referring here to competition in two areas :

- Terminal equipment provision
- Networks provision - either on service competition (on the basis of the existing public-owned network structure) or facility competition (to allow physical entry into network provision).

2) Private Terminal Attachment Provision.

This has been possible in a number of countries for a long time. In Germany, for example, private terminal equipment (mainly PABX systems) has been offered in competition[2] to Post Office supplied equipment since 1900. In other countries, depending on the restrictiveness of the attachment policy, some competitive activity in terminal equipment supply has also been possible. But change has been most pronounced in the USA. Initially, the dominant Bell system, and the independent companies, pursued a very restrictive terminal equipment policy[3]. There had been earlier attempts (after the second world war) by the FCC to give specific rights to customers to own and operate their own equipment, if it caused no harm to the network. Examples include the Recording Device Decision in 1947, the Hush-a-Phone Decision[4] in 1959, and the well-known Carterfone Decision in 1968[5].

In all these cases, the very narrow and restrictive interpretation of the Bell system concerning terminal equipment policy was challenged by the FCC and the courts, without making the public or the competitive industries really aware of the fact that the legal monopoly of AT&T or the other common carriers could be successfully challenged in the marketplace. It was only with further policy pronouncements which regulated this new freedom of attachment in the form of the Equipment Registration Programme in 1977-78, and a successful denial of AT&T's Primary Instrument Concept in 1979 by the Supreme Court, that the boundaries of the legal monopoly were narrowed and competitive offerings of terminal equipment became truly possible.[6]

3) An Assessment

Analysis of these developments is clouded by the short time horizon and by the considerable legal uncertainty associated with the US regulatory environment. This uncertainty may to some extent have hindered entry and expansion of the new "interconnect supply" industry.[7] Nevertheless, some interesting insights into the effect of increased competition in the telecommunications equipment market can be obtained:

- The much discussed harm to the network has not taken place.[8]
- The product spectrum has been very much improved over the past few years, with many more features now available not only to sophisticated business consumers, but also to private households.[9]
- The behaviour of the traditional suppliers is changing.
- Innovative activity has increased and the adoption of new features has been accelerated.
- The price-cost margin for the equipment in question has apparently been reduced and moved closer to local cost conditions through unbundling of tariffs.

- The structure of the industry itself is changing considerably. The market shares of the interconnect supply companies in terminal equipment markets have reached 30-40% in selected sectors (such as small PABXs, hotel and motel PABXs, etc[10]). The market share of the leading monopoly is decreasing, but it is still in a dominant position.
- As a result of this competition and competitive entry, additional regulatory signals are becoming available, which help the regulatory agency to determine the kind of price and rate revisions which should be pursued. In addition, the competitive threat causes a change in the firms' behaviour, which regulatory agencies alone would not be able to bring about. The point that competition is an important instrument of regulation is of particular significance in the Western European context.

4) Network competition

With network competition, the degree of uncertainty is equally high. Nevertheless, here too competitive offerings by so-called 'interconnect' companies[11] point towards the strong competitive potential in this market.

Just as in the market for terminal equipment, network competition was opened up with an incidental decision. At the time, it did not seem to challenge the underlying major philosophy of one unified telephone monopoly. The Above 890 Decision in 1959 was designed to give micro-wave transmission privileges only to firms whose own demand was large enough to justify such investment and where common carriers had not yet been able to provide such services. But already, with the Specialized Common Carrier Decision in 1971, the monopoly was further reduced to allow entry of data transmission networks. Furthermore, with the Domestic Satellites Decision (the Open Sky Decision) in 1972, in which AT&T was prevented from owning domestic satellites and other firms were encouraged to enter the domestic satellite business, some form of intermodal competition was set in motion. This type of competition was considerably increased with the Shared Use and Resale Decision. The FCC allowed customers of leased lines to resell their services to achieve both structural reform and to have the resale ability act as a price regulating mechanism. With the Execunet Decision in 1978, which required interconnection at the local loop, such private 'interconnect' carriers were able to offer network services which were almost equivalent to the telephone service offerings of the traditional companies.

Thus, the position changed from one in which few exceptions to the legal monopoly (namely private facility service) were permitted by the FCC, to full competition, not only as a regulatory tool in network transmission, but also with the Execunet Decision, as a substitute for regulation. The Second Computer Inquiry Decision continues this trend in a forceful way. Today, the traditional common carriers are not only faced by network competition from the interconnect carriers but also from other common carriers who have diversified, such as GTE (with Telenet), Continental, (with American Satellite Corporation and Western Union), and ITT (with USTS), as well as from large office equipment suppliers now entering (such as IBM, XEROX etc).

The results of these changes in policy have to be

interpreted with the same caution as those concerning terminal equipment. Business uncertainty remained quite high. Like other decisions, the Second Computer Inquiry Decision is bound to be in the courts for many years. With further entrants lining up (like Exxon) and market participants jockeying for position, the market remains in a state of flux. Nevertheless, apart from the uncertainty and legal obstacles surrounding it, the results of this exercise in increased competition are encouraging.

- Cream-skinning has become the order of the day. As a result, tariffs which had been set on a uniform basis independent of cost across the country are becoming cost oriented. While the potential for cross-subsidisation is significantly reduced, as some of the surplus is eroded by cream-skimming entrants, the redistribution between densely and thinly populated areas is thought to be minimal [12].
- As a consequence of entry, price-cost margins are reduced and move closer towards local cost conditions.
- In addition to standard telephone services, innovative interconnect companies started to offer some new services, which increase the available products features to network users as well.
- Again, prices, products and qualities offered in the market provide important additional regulatory signals to the regulators. Furthermore, the traditional carriers, who still supply the bulk of the national services, have modified their behaviour to take better account of market demands. Such

a change could not have been achieved with traditional regulatory tools alone, but is a direct result of competition.

But it is important to note that the FCC already went much beyond the concept of competition as a regulatory tool. In essence, it introduced competition as a substitute for regulation. Except for continued surveillance over AT&T (as long as it maintains its current monopolistic position), competition is seen in the long run as a substitute for regulation. A similar language is contained in the various Bills before the US Legislature.

If one goes to full network competition, an important question remains: what to do with the local loop, where economies of scale are still dominant and where a legal monopoly, in a regulated form, will still lead to a more efficient provision of services? But here, too, technology is changing. Mobile radio, cellular systems, and direct broadcasting satellites are offering competitive technical alternatives, whose widespread application may not be too far off.

5) Lessons

Before drawing any definite conclusions from the US experience, one must be cautious and keep in mind the uncertainty in which the market penetration by interconnect firms has developed. There has been much uncertainty because of legal cases in the courts; the potential of cross-subsidisation and predatory pricing by the common carriers, who have protected legal monopolies; and finally the fact that there has been no experience in providing competitive telecommunications services.

These three factors must have reduced the potential of the market to act, according to Hayek,[13] 'as an efficient search process for potential new services. Nevertheless, three conclusions seem possible:

- Efficient competition is possible in both terminal equipment and network provision. There are as yet no signs of duplicative, inefficient resource use. It can be argued that the potential of entry already implies a strong enough discipline force even in areas where some economies of scale prevail, but entry is easy[14].
- Customers have benefited through increased service offerings, both in terminal and network facilities. At the same time, the regulatory procedures surrounding this increased competitive environment[15] have ensured that no harm to the network took place. In other words, basic access to the network, which is of vital concern with regard to the public role of the telecommunications network, has been maintained throughout.
- Dynamic efficiency has increased and points towards significant long-term gains. Because new entrants do not have to provide universal service, their investment risk is lower and the potential of the market to act as a successful search process is considerably increased.

While these three conclusions support a move towards more competition in the telecom sector, there are a number of issues which, especially during transition, deserve special attention: predatory pricing, access charges, and redistribution. We have hinted at these issues already.

- Predatory pricing exists when the dominant firm in the market prices to deter entry, or to weaken or to drive out competitors. It requires considerable resources, which may be obtained by cross-subsidisation, i.e. charging above cost in other, perhaps legally protected monopoly markets. This has certainly been an issue in the USA, as indicated by the recent MCI Case. To prevent cross-subsidisation, the FFC is requiring AT&T (and GTE) to provide terminal equipment and enhanced services only through separate arms length subsidiaries (Computer Inquiry II) as well as requiring an improvement in accounting method.[16]
- Access charges refer to the price which interconnect carriers trunk operators must pay to originate and terminate their competitive offerings. Since the costs for the local loop represent the largest single item of investment in the network (roughly 75 % followed by 9 % for switching and 16 % for trunk transmission[17]), those who access it should pay a cost covering contribution. Only this way can established carriers compete on an equal footing. So far, the SCCs have been at an advantage in the ENFIA agreement -paying only a portion of the full cost as compared to other common carriers.

  This issue has not been settled at all and the proposal of the FCC in the MTS-WATS proceedings differs significantly from the legislative proposals before Congress.
- The distributive issue is directly related to access charges, as they may also have to cover certain subsidies for inter-regional inequality. At the moment, local service is being

subsidised by long distance revenue (for example, through the separation and the settlements procedure). Uniform national tariffs also imply cross-subsidisation between dense and less populated regions. With the introduction of network competition, these tariffs become differentiated according to local cost conditions, leading to rising costs for local calls and a lowering of long distance tariffs. This later trand is being amplified by technical changes (reduction of long distance costs), increasing the redistributive aspect of a move to competition. Furthermore, tariffs will become differentiated between low and high density routes, leading to a further change in the tariffs structure. Establishing this difference in rates in the presence of significants cost variations will have an important signalling effect in the service market. On the other hand, retaining a uniform national tariff in the face of significant cost variations would go against the principle of costs being sufficiently reflected in market prices. It would hinder an effective search process by the market.

6) <u>Competition with state-owned enterprises</u>.

While assessing the US deregulation experience as a successful policy move, it is not obvious that this path should be followed in Europe. In most cases, European countries do not have private enterprise provision of services in a regulated environment, but state-owned enterprises which provide the service. With such public ownership, the power to regulate and to carry out other policies rests directly with the state-owned enterprise, the PTT. Since regulation and execution go hand in hand, no information loss should exist which makes competition as a regulatory tool less attractive. In addition, by having a state-owned enterprise, the government has an additional policy tool at its disposal - the tool of a mixed economy.

The existence of state-owned enterprises offers other alternatives. For example, risks can be taken in the introduction and provision of new services, for which private companies may not find sufficient backing in imperfect capital markets. With state-owned enterprises and connected legal monopolies in the provision of services, cross-subsidisation between services is possible. It can be used to achieve certain redistributive aims directly, if other redistributive tools are less efficient or politically unfeasible.[18] Significant gains from competition must therefore be shown if the benefits of state-owned enterprises in a mixed economy are to be given up.

Some of those benefits are only theoretical, however. For example, state-owned enterprises tend to develop, just as private enterprises do, their own behavioural rules and these may lead to a divergence from regulatory goals. They therefore need to be supervised, and competition as a regulatory tool is an effective and desirable alternative. In addition, state-owned enterprises are often quite conservative in their investment and operating policies. They value the reliability of service and avoidance of interruptible service (poor quality would give them a poor public image) often higher than individual customers would. They may be slow to innovate, unless a product has proved itself beyond doubt, thereby reducing the potential

of the market to act efficiently in its search process. Furthermore, the benefits of being able to cross-subsidise between services may not be very large and may be outweighed by the costs. The costs are associated inefficiencies in investment and expenditure on substitutes for the services, which are priced above costs, and a reduced incentive to innovate for services which are priced below cost. Further inefficiencies are caused by lobbying efforts, given a strong preference for maintaining the status quo by those disadvantaged by technical change[19]. The threat of entry will not only reduce the available surplus for cross-subsidisation and make prices more efficient; it will also reduce the potential of politicians to tinker with the system in a politically opportune way. In summary, even with European-type PTTs, competition serves a useful purpose in the telecommunications sector, so that one may consider the following recommendations:

7) <u>Recommendations for the terminal equipment market</u>

If one starts with the proposition that the effective working of the markets'search process requires a maximum amount of producer and consumer freedom, then one should allow as much competition as possible. Nevertheless, some safeguards should apply to terminal equiment which is connected to the network, even if it means limiting producer freedom, somewhat.

- No first-party harm (to avoid injury to equipment operators or employers of the network operating company from malfunctioning equipment).
- No third-party harm (which would influence the communications activities of third parties).

- Some second-party harm should also be avoided. (Second-party harm refers to distortion or the inability to communicate between the caller and receiver due to the use of incompatible equipment or different communication protocols[20]).

The USA adopted the first two safeguards[21] with the registration programme of the FCC in 1978/1979. It is normally possible to complete the registration process of new equipment within two months. In Europe, this process usually takes considerably longer (especially if new features or different concepts from those mutually approved are proposed), often involving delays for innovative equipment suppliers of several years.

There is one additional policy issue of relevance here. If competition is possible, what is the future role of the PTT in such a demonopolized terminal equipment market? Even though some have argued against the participation of the PTT, I do not follow this line of argument, on the grounds of both economies of scope, dynamic side effects and political feasibility.

Economies of scope refers to benefits obtained when the joint supply of services is provided more cheaply than their separate supply. The PTTs may well enjoy some economies of scope either between network and terminal equipment or in relation to other activities, and it would not be efficient to prevent them from exploiting this potential in the public interest. Since we do not know where and how large these economies of scope are, we should leave that test to the market itself. This requires a proper framework for workable competition. For example, when allowing the participation of the PTT

in these markets, it may be necessary to separate the regulatory function of the PTT from the equipment sales division. In addition, since the PTT has access to revenues from its legal monopolies services, sufficient care must be taken to prevent the danger of predatory pricing. One possibility may be to extend the anti-trust laws to cover the activities of the PTT, another to improve the accounting procedure.

The second point, political feasibility, refers to the large presence of the PTTs in these sectors. It may not be feasible, at least in the short term, to reduce completely the activities of the PTT in this very dynamic segment of the market. However, to expose it to competition will leave the test of the size of the PTT in the competitive sectors to the market itself. In addition, important internal effects may be produced within the organization when it tries to compete in these markets, which may be much more substantial than those achieved via regulatory reform. In fact, one would expect a movement from a bureaucratic to a dynamic organization.

8) Recommendations for networks competition

Generally speaking, I see no reason why European countries should not move towards competition in this field as well, especially if one advocates both consumer and producer freedom. However, while the US experience of facility competition continues, and further evidence about its effect is being collected and analysed, it may be useful to await the results before restructuring the institutional environment in which network competition is to take place in Europe. For the moment, I suggest at least service competition. This involves the ability to resell and/or share leased-line services and to allow, on the basis of these, the establishment of value-added networks.

The effect of a shared-use and resale policy will increase consumers' freedom and allow a more effective use of the network. A whole host of user restrictions (especially concerning integrated terminals) could be lifted and the search for improved solutions be carried out by the market, compared to the current administrative search. One further effect, just as we have seen in the USA, would be the 'de-averaging' of uniform national tariffs - through cream-skimming by new entrants. However, with a response towards cost-oriented tariffs by the PTT, the tendency of entrants to survive on arbitrage alone will soon disappear. The only entrants to survive in this case will be the value-added networks (i.e. the supplier adds value to the service it rents from the PTT) or those entrants who are more efficient in network utilization than the PTT, or those who share or resell services which they themselves do not fully use. It is difficult to predict the exact amount of redistribution (from low density to high density areas and users) as a result of tariff revision. But if continued cross-subsidisation was politically required, one could tax the entrant equivalent to the cross-subsidisation contribution, which his entry loses the PTT, rather than restrict resale and shared use altogether[22].

This recommendation, which at present stops short of allowing facility entry, leaves one remaining problem: how can gold plating in the facility provision be prevented? In essence, the PTT should be allowed to operate alone all the physical

networks with only service competition, checking its network performance. But competitors operate only on the basis of leases from the existing networks. If the PTT chooses an excessively high level of quality or an excessively expensive product, for example in transmission provision, the competitive entry of interconnect companies will not be able to correct this technically inefficient choice by network operators. To avoid this, either a move must be made towards facility competition as well, or at least intermodal competition must be allowed with respect to long-distance transmission. One possibility in Europe may be the separation of ownership of satellite services from the PTTs towards an independent, perhaps European, satellite service.

FOOTNOTES

* The author is currently with CRG, Ecole Polytechnique, 17 rue Descartes, 75005 Paris, and ESSEC 95001 Cergy, France, on leave from DIW, Berlin, West Germany.

Part of this paper is based on work carried out in the Intervention and Adjustment Project at the Sussex European Research Centre. The financial support of the Volkswagen Foundation, FRG, is gratefully acknowledged. Additional insights were gained from a project on the role of competition in the telecommunications sector sponsored by the West German Monopolies Commission. That research is summarized in G. Knieps, J. Müller and C.C. von Weizsacker, Die Rolle des Wettbewerbs im Fernmeldebereich, Nomos, Baden-Baden, 1981. The thoughts of and critical discussions with the two co-authors of that report have also influenced the author in the preparation of this article. An earlier version of the paper was presented at the Third International Conference - Analysis, Planning and Forecasting for Public Utilities, Fontainebleau, June 1980. The revision and extension was made possible by grant from the German Marshall Fund.

1. Some networks in the USA are now owned by separate entities and expanded or contracted as each operator sees fit. Network competition can extend all the way from simple point-to-point provision of services to the offering of complete transmission and switching services as well as value added networks (VAN), which are equivalent to today's PTT offerings.
2. The competition, however, has been regulated by the PTT.
3. In some instances this went as far as describing in detail the plastic covers which could be sold privately to protect telephone books. FCC, first Report and Order, Docket 20003, 1976, p. 42.
4. This concerned an acoustic capsule which was placed over the speaking device to reduce background noice.
5. For details, see G. Knieps, J. Müller and C.C. von Weizsacker, Die Rolle des Wettbewerbs im Fernmeldebereich, Nomos, Baden-Baden, 1981.
6. The registration program became effective on Oct. 7, 1977, after the US Supreme Court declined to review a lower court decision supportive of the registration program, which in effect eliminated the need for protective coupling devices. At first, it covered only some simple interconnect products. The extension to the more import PABX and key telephone market came only on April 13, 1978.
7. This industry comprises private firm who sell terminal equipment to be connected to the public network by its users.
8. See the discussion in FCC, 2nd Report and Order, Docket 20000, Jan. 29, 1980, Appendix C.

9. Many of these new features are due to the microelectronic revolution, not just the effect of 'interconnect' competition.
10. The Outlook for the US Telephone Interconnect Industry, Impact Study, Arthur D. Little, Cambridge, USA, June 1980 ; nationwide, the market share of the interconnect supply industry is expected to reach 25 % by 1985.
11. These are private companies who supply network services to individual customers, such as MCI, DATRAN, SP Communications,USTS. ; see also para 1
12. NTIA, Deaveraging of Interchange Toll Rates due to the Introduction of Competition - Preliminary Estimates, Working Paper, Denver, June 1979.
13. F Hayek, Der Wettbewerb als Entdeckungsverfahren, Kieler Vorträge N°. Kiel, 1968, Weltwirtschaftliches Institut,
14. See also the theory of contestable markets, set out by J.C. Panzar and E. Bailey, 'The Contestability of Airline Markets during the Transition to Deregulation', in Law and Economic Problems, Vol. 84, N° 1, Jan. 1981,
15. i.e. the terminal equipment registration programme, the interconnect arrangements for specialized common carriers etc...
16. I have some reservations about the arms length subsidiary concept. It should be considered only as a measure of last resort, when even improved accounting procedures and an application of the anti-trust laws do not prevent the threat of predatory pricing.
17. Planck, R.L., 'Neue Techniken in der Geschäftlichen Kommunikation', paper prepared for Telecom 80 Deutschland, Cologne, October 1980.

18. R.A. Posner,'Taxation by regulation', Bell Journal of Economics and Management Science, Vol. 2, N°1, 1971 pp. 22-5(
19. See B. Owen and R. Braeutigam, The Regulation Game : Strategic Use of Administrative Process, Ballinger, Cambridge USA, 1978, especially chap 1.
20. The exact degree to which the avoidance of second-party harm needs to be specified depends essentially on transaction costs between private parties versus the cost of additional regulatory rule making. Probably both basic standards and minimum quality levels, which are necessary to ensure compatibility between equipment, may be issued by the PTT in many reasonable circumstances. But one would not consider any further concerns, like the specification of additional quality features or price regulations, or the requirement of a servicing monopoly by the PTT, as acceptable grounds for certification standards. Also trade policies should not influence the certification procedure.
21. Some claim that because of the missing third safeguard, the quality level of the US phone service may be deteriorating.
22. For details, see Knieps, Müller, von Weizsäcker, para 5.

REGULATORY PROCESSES AND SOCIO-ECONOMIC ISSUES

TELECOMMUNICATIONS AND THE LOCATION OF EMPLOYMENT: IMPLICATIONS FOR REGIONAL DEVELOPMENT POLICY IN CANADA

BARRY LESSER

LARS OSBERG

Dalhousie University

As Canada enters the 1980's, she is embarking on what many observers have called the "information-processing age". New technology in telecommunications, computers and computer/telecommunications combinations and a growing awareness of the potential of the existing technology in these fields holds the promise of a major source of increased productivity in the years ahead. One can, however, question how these new technological developments or new applications of existing technology will be distributed among the regions of Canada. In turn, given dramatic technical change and a government goal of redressing regional imbalances, new policy directions must be considered if the implementation of new technology is to produce desirable socio-economic results.

This paper begins by accepting the premise that the amelioration of regional economic disparities in Canada is a desirable goal for Canadian governments to pursue. Two corollary propositions are also accepted without question:

1. that, *ceteris paribus,* the goal of reducing regional disparities should be pursued by policies which are most consistent with the simultaneous goal of promoting economic efficiency, and
2. that, when there is an irreconcilable conflict between the goal of efficiency and the goal of greater regional equality, it does not necessarily follow that the efficiency goal should dominate; it is, however, important to have an appreciation of what the efficiency losses may be.

Our contention is that the revealed preferences of the Canadian polity indicate that the 'social welfare function' of Canada contains as arguments more than simply the individual incomes of individual Canadians wherever they might be; that Canadians are not indifferent to the decline of long established communities or to changes in the social structure of those communities. As a result, governments which seek to maximize 'social welfare' in Canada must seek to balance the goals of economic efficiency as indicated by market judgements with the political claims of regional and distributional equity. Vague and ill-defined as the latter concepts may be, the former has no unique claim to overriding social importance.

The existence of regional economic disparities in Canada, expressed in terms of relative incomes and employment is both well-recognized and well-documented. {1} So too is the pursuit of policies by governments in Canada which are intended to redress regional disparities. {2} It is, however, also well-recognized and well-documented that, to date, these policies have been of only marginal impact. {3} While the relative position of some "have-not" provinces has shifted during the 1970's, this has been primarily a function of Western Canadian energy resources coupled with rising energy prices and not the result of the government's regional development policy. Finally, it may fairly safely be asserted that the traditional focus of regional development policy in Canada has been a combination of (a) the use of various transfer payment mechanisms (b) the promotion of regional infrastructure development and (c) incentives to the location or relocation of goods-producing industries in Canada's less developed regions. Telecommunications has not been a focus of government-directed infrastructure creation nor have offices or office units within the service sector been emphasized in considering new economic activities to locate or relocate in less developed regions. {4}

In this paper, it is argued that computer and telecommunications technology offers the possibility of a redistribution of regional income and employment with, we anticipate, relatively little (if any) cost in terms of economic efficiency.

We do not, however, offer in this paper quantitative estimates of these effects; and figure 1 may help to illustrate why. Our contention is that the telecommunications industry in Canada is in the midst of a period of extremely rapid technical change -- if we graph the marginal cost of providing many types of services against time, we might represent Canada's current poisition as point c in figure 1. Published data is, of course, only available with a lag, hence available statistics might cover the period ac'. In most instances in economics, such reporting lags (1-2 years is not uncommon in time series on particular industries, 10 years is normal for input-output data) are relatively unimportant, but this may not be the case in the current telecommunications/computer field. While historic data over the period ab will provide reasonable estimates of technological trends over such a period, any errors or "noise" in the data over the period bc' will affect substantially any estimates of cost trends for the period bd.

The eventual social and economic impacts of technical change in telecommunications will of course depend on the ultimate levels of service costs attained (de in figure 1), the extent to which those changed costs of service are reflected in consumer prices and the extent and rapidity with which consuming firms and individuals alter their economic behaviour in response to changed prices. Even leaving aside the issues of the relationship of market prices to marginal service costs in such a highly regulated and monopolistic industry and the speed of technological diffusion in response to any change in prices, we

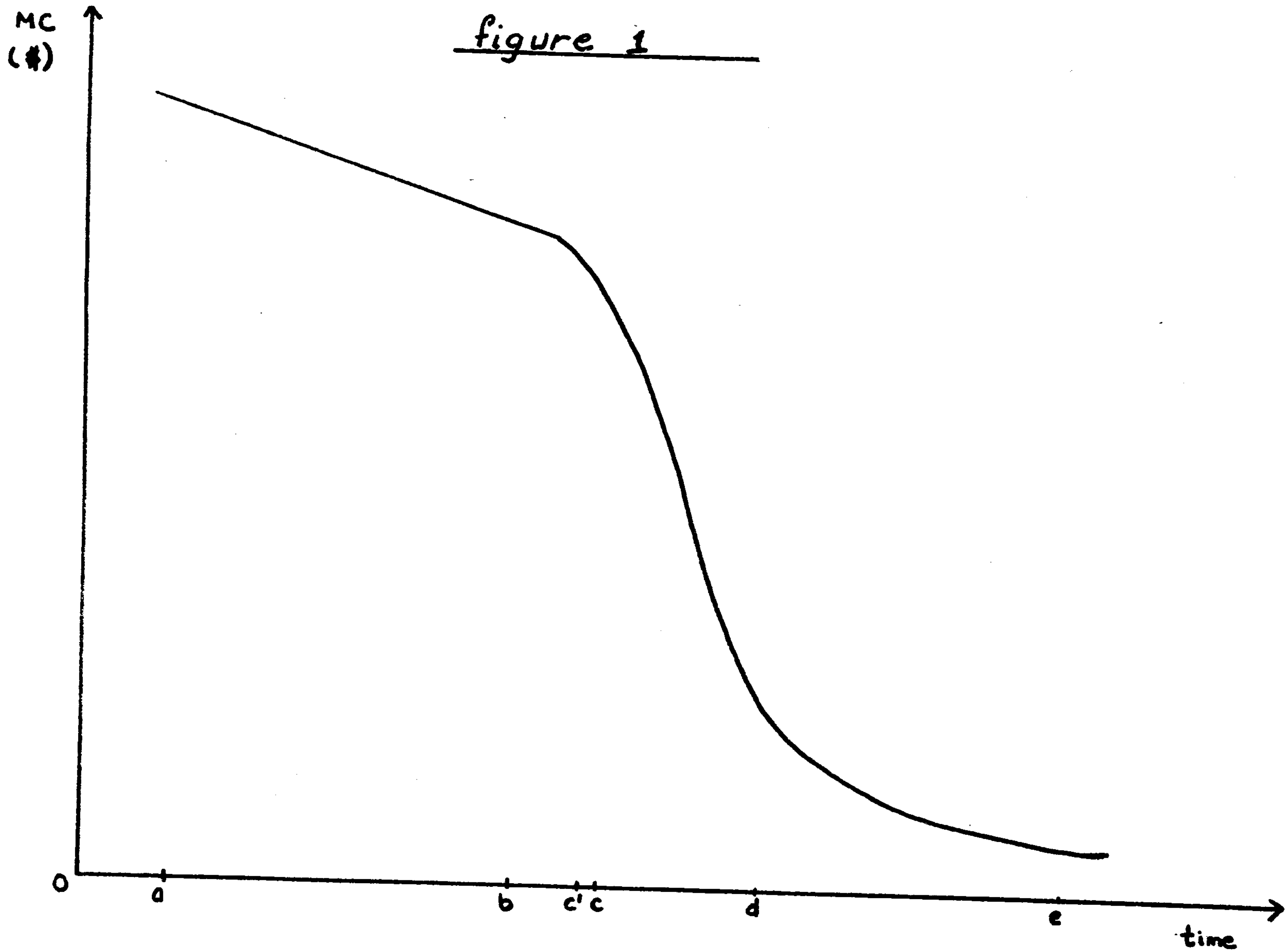


figure 1
MC
($)
0
a
b
c'
c
d
e
time

cannot see how historical data on the period ac' can enable us to predict with confidence costs of service over the period de.

Indeed, in many instances current technological developments are simply unprecedented. If one compares, for example, the marginal cost of delivering tele-conferencing facilities or live transmission of athletic events to Newfoundland outports via a micro-wave relay network technology or via current satellite technology, one is basically talking about the change from prohibitively expensive to moderately cheap. Government policy cannot wait until we have completely reliable ex post estimates of the trend of such technology over the entire cycle represented by ae in figure 1, since if it does so the implications, for good or ill, will already be in place. Government policy makers therefore have little realistic alternative to some degree of imprecise futurology; delay or inaction is in itself a policy.

Telecommunications Externalities and Regional Implications

Telecommunications is a two-way interactive information-transfer mode, which exhibits strong externality characteristics. The recipient of a call, as well as the originator of a call, may receive benefit from the call. Similarly, the existing subscriber population may benefit, because of greater calling capability, when a new subscriber joins the system. {5} When put in a regional context, this means that individuals and firms in one region may share in the benefits from telecommunications development in another region.

Suppose we have two regions, A and B, and that A is relatively disadvantaged in economic terms compared to B. Suppose, also, that A expands (upgrades) its telecommunications network, creating greater calling capability and a larger subscriber body, not only within A but between A and B. It is quite possible that B will receive more benefits from this expansion than will A. This, in turn, would mean that telecommunications expansion in A could worsen the relative inequality between A and B. Of course, if A received the greater benefit, then relative regional inequality would be lessened.

Which of these possible outcomes is, in fact, realized will depend on whether telecommunications on balance tends to promote a centralization of economic activity or a decentralization of economic activity. Region B, if it is the relatively advantaged region, may be presumed to be the central region, initially possessing a larger, more developed economic base. Region A, if it is the relatively disadvantaged region, may be presumed to be the periphery region. B will gain, in terms of employment, relative to A from telecommunications expansion if, on balance, telecommunications promotes a centralization of economic activity, while A will gain employment relative to B if, on balance, telecommunications promotes decentralization of economic activity.

Which of these outcomes is more likely to occur will depend on a number of factors including, inter alia, the "footloose" nature of the economic activity in question, the size of firms in the periphery region and hence their ability to employ the technology, the degree of innovation shown by government in employing the technology, and management practices within affected firms.

Telecommunications, potentially, could positively influence employment opportunities within a less developed region in several ways, affecting both labour supply and demand:

1. by permitting a relocation of existing firms or units of firms to less developed regions;
2. by making less developed regions more attractive locations for new firms to consider as initial location sites;
3. by expanding the effective market area of existing firms within less developed regions;
4. by enhancing the potential for greater social contact for workers in an isolated region with persons outside the region (relatives, friends), thus making isolated regions more attractive to outside workers and/or reducing rates of turnover (typically higher in such areas as Northern Ontario or Quebec than elsewhere);
5. by enhancing the delivery of educational and health care services in isolated/less developed regions, thereby

enchancing the "quality" of local labour supply in the long term and making the region more attractive to in-migration of outside workers (especially skilled workers) in the short term.

On the other hand, telecommunications might negatively influence employment opportunities in periphery regions, if there is a greater tendency towards centralization than decentralization as a result of:

1. permitting centre region firms to more effectively (efficiently) service periphery regions from a central location, thus eliminating such activities as inventory depots or sales offices in periphery regions;
2. permitting centre region firms to more effectively penetrate periphery region markets, increasing competitive pressures on local firms which, because of distance, have traditionally enjoyed a degree of "natural" protection;
3. permitting a centralization of management control over satellite operations, which control is maintained through a combination of telecommunications and computer facilities and eliminates that part of the management cadre previously located in periphery regions to monitor ongoing operations. (In practice, it appears that a combination of telecommunications and rapid air travel access is required to enable both

continual central monitoring and rapid trouble-shooting.);

4. permitting a replacement of clerical staff by central computer facilities, which are essentially "foot-loose" but may well be located in the centre region.

Sectoral Impact

If one views the economy as composed of a goods-producing sector, a distribution sector and an information sector, one can analyze the potential impact of telecommunications on employment in each of these sectors. If employment in the goods-producing sector is defined as those jobs which involve primarily the manipulation or manufacture of physical objects, it is unlikely that telecommunications will have a major impact on the location decisions of goods-producing establishments. In the distribution sector, the impact is likely to be negative for periphery regions. In the information sector, there are competing forces and it is here that we will suggest that government policy may play a key role.

For over 40% of goods producing industries the cost of relocation of production from current sites could be considered almost infinite. Agriculture, forestry, fishing, trapping, mining and construction are all fairly immutably tied to their resource or market base. Manufacturing industries differ widely in their flexibility for relocation, but for most establishments, the costs of physically transporting inputs to and from markets have traditionally dominated considerations of telecommunications costs and will likely continue to do so for the foreseeable future. Telecommunications could indirectly affect employment in periphery regions in goods producing industries if it assisted the growth of existing firms <u>over time</u> by decreasing the cost of the acquisition of information. {6} But there is no present evidence to indicate the magnitude of such an effect, if it exists.

Within the distribution sector of the economy, i.e. those jobs involving the transportation, wholesaling, and retailing of goods, one can expect, based on recent Canadian experience, that telecommuniations/computer technology is, on balance, likely to lead to greater centralization of inventory holdings which, in turn, will most likely decrease employment in periphery regions. {7} As the speed of order processing increases, it becomes feasible to service isolated outlets from a centralized depository rather than maintain a (necessarily larger) total stock at dispersed locations. The employment loss in warehousing jobs may, however, be mitigated if the now lower cost of establishing retail outlets produces an increase in their density of coverage of isolated market areas.

It is within the information sector of the economy, i.e., Bell Canada restricted its attention to DDD in terms of two mileage bands but regardless of TOD (also excluding holidays), duration and type of customer, and to person-to-person with the same restrictions, but without mileage bands differentiation. This led Bell Canada to use deflated revenue output measures. Only Fuss and Waverman (1981), in one of their attempts, consider duration explicitly. Finally, Bernstein (1980) uses messages as his output measure.

2.3 Review of the Demand Specification

In this section, we shall review the main characteristics of the various demand studies cited earlier to attempt to draw some conclusions with respect to the specification of a demand model.

2.3.1 Money Illusion

Taylor (1980)'s specification of the demand for telephone services allows for the possibility of money illusion. Only the earlier models such as Bell Canada (1976) and Dreessen (1977,1978) have maintained this format, all other studies rejecting a priori any money illusion. This will also be the approach adopted here.

2.3.2 Cross-price Elasticities

The only attempts at measuring cross-price elasticities are found in the work of Concordia University and in Dreessen (1977, 1978). Cross-price elasticities were estimated unsuccessfully in Bernstein et al. (1977) and Corbo et al. (1978, 1979). The problem follows from the lack of variability of relative prices, the degree of multicollinearity being very high, and from the aggregate nature of the local services series. Similarly, in the Intra-B.C. model, Piekaar (1980) abandoned Dreessen's (1977, 1978) previous attempts to introduce, in some mileage bands, the price of local services as an explanatory variable. If the price of local services is seen as an access charge, it should be deducted from the income variable and not be introduced as a price variable (Bernstein, 1980). Toll calls over other mileage bands are not substitutes, and their price level would act only through the income constraint. In other words, cross-price elasticities between mileage bands should not be a problem to worry about. Cross price elasticities between types of call and between different periods of the day or days of the week are still outstanding problems.

2.3.3 The Dynamic Structure of Demand

The most general linear demand model, in terms of its dynamic structure would be the transfer function:

(1) $$r(B)q_t = B^s s(B)p_t + B^u u(B)y_t + B^v v(B)x_t + e_t$$

$$\emptyset(B)e_t = \Theta(B)a_t$$

where r(B), s(B), u(B) and v(B) are proper rational functions in B which is itself the lag operator, i.e. such that $Bz_t = z_{t-1}$,

s, u and v are non-negative scalars which indicate the dead time, $\emptyset(B)$ and $\Theta(B)$ are proper polynomials in B and a_t is $N(o,\sigma)$,

q_t, p_t, y_t and x_t denote the quantity demanded, the price of the service, the income and other exogeneous variables, after proper deflation and transformation, as required.

This general model, without transformation of the variables and with $r(B) = (1-r_1B)$, $s(B) = s_o$, $u(B) = u_o$, $v(B) = 0$ and $\emptyset(B) = \Theta(B) = 1$ is the Houthaker-Taylor flow adjustment model.

As noted earlier, the standard application in modelling the demand for telephone services is the DL model, which implies that the variables are expressed as logarithms. Then we have the habit formation model; this model was used by Piekaar (1980). If in addition to the habit formation hypothesis, one also assumes that $\emptyset(B) = (1-\emptyset_1 B)$ while $\Theta(B) = 1$, i.e. if one introduces a correction for autocorrelation, then we obtain the model adopted in Corbo et al. (1978) and Fuss and Waverman (1981). The most extensive study of $\emptyset(B)$ can be found in Corbo et al. (1979), in which even $\emptyset(B) = (1-\emptyset_1 B-\emptyset_2 B^2-\emptyset_3 B^3)$ was investigated. However as all of these tests in Corbo et al. (1979) are applied to regressions which also contained cross-price elasticities and as it is unlikely that we can disentangle those elasticities from one another, the utility of these tests is limited. On the other hand, the attempt to go beyond a first degree polynomial in the specification of $\emptyset(B)$ is welcomed since its higher degree polynomials introduce the possibility of complex roots corresponding to cyclical movements.

Economic theory has nothing to tell us as to the proper dynamic structure of the model, and one must turn toward time series analysis. Box and Jenkins (1970) present a methodology to identify a transfer function, however, their methodology cannot be applied, at this stage, to our problem because the series are too short. Furthermore it has also been noted that different models may yield very similar summary statistics (Granger and Newbold, 1978). In this context, it seems wise to follow Box and Jenkins' parsimony principle, i.e. to select the simplest of the models which can reasonably be entertained. The testing, as indicated above, will remain rather ad hoc as long as we do not have longer time series.

2.3.4 The Seasonality of Demand

As noted earlier when referring to the construction of price indexes, seasonality affects the demand for message toll services. Again the time series framework presented in the previous section can accommodate this new dimension without any problem. The problem, however, is that the requirement on data is too much greater, hence, the practical application of standard time series procedure will have to wait for a few years, when, barring major structural changes, we will have sufficiently long series!

The seasonality question affects only the industry studies. There, the Intra-B.C. model is based on seasonally adjusted data, the seasonal adjustment procedure being the Bureau of the Census X-11, while Bell Canada (1980) utilizes seasonal dummies for the intercept. Cleveland (1972) has shown that the X-11 program can be approximated by a seasonal multiplicative autoregressive-integrated-moving average (ARIMA) model. It can also be shown that the use of dummy variables can be analyzed from the point of view of ARIMA models with common roots which were studied by Abraham and Box (1979). Courchesne, Fontenay and Poirier (1980) have developed a general analytical framework within which the use of the X-11 and the use of dummy variables both can be evaluated within the general ARIMA specification. They show that it is an empirical matter which of the dummy variables approach and the use of variables adjusted by the X-11 dominates. Hence, once again, as with the previous sections, we cannot derive a general rule.

Finally, even though it cannot be said exactly how many degrees of freedom are lost through a prior adjustment by the X-11 seasonal adjustment program, if one uses its ARIMA approximation as given in Cleveland (1972) or Cleveland and Tiao (1976), one can obtain a reasonable estimate. This correction was not done in the Intra B.C. model which treats the seasonally adjusted data as raw series.

2.4 Fuss and Waverman (1981) Demand for Toll Calls by Distance and Length of Call

As noted earlier, most approaches adopted in the demand for message toll services are tailored to the data base available to the author. A particularly interesting example is one of Fuss and Waverman (1981)'s models which they tailored around a 1977 Quebec interrogatory which made public the number of calls by duration and by mileage band for Bell Canada. Even though the econometric estimation, by the authors' own account, was unsuccessful, it is worthwhile to present their model.

those jobs, from bank-teller to university lecturer or government administrator, which involve primarily the manipulation of symbols and the creation, transmission and recording of information, that the largest potential exists for telecommunications/computer technology to affect the spatial distribution of employment. As relatively far back as 1967, work done in the U.S. indicates that 25% of U.S. GNP originated in the production-processing-distribution of information goods and services while, in addition, over 20% of GNP originated in the production of information services by the private and public bureaucracies for purely internal uses. By 1970, close to half of the U.S. work force was classified as "information workers", holding a job where the production, processing or distribution of symbols was the main activity. {8}

Within Canada, the same type of precise information is not readily available but by one very rough estimate, approximately 40% of the Canadian labour force could be regarded as members of the information sector of the present day Canadian economy {9} -- the fastest growing sector of employment over the past two decades.

The chief function of persons in this information-processing sector is the storage or manipulation of symbols. Not surprisingly, the main industries in which these persons are employed (e.g., banking, government) are currently the industries most dependent on the purchase of telecommunications services. {10} This does not mean, however, that such firms have already made use of telecommunications/computer technology to the best possible advantage. Not only might these firms have stopped short of maximizing the efficiency gains from use of the new technology because of, for example, traditional management reluctance to decentralize the workplace, but future technological advances (already in the offing) will undoubtedly create additional opportunities. (In addition, of course, one must face the possibility of 'market failure' where individual profit-maximization may not in aggregate produce a socially optimal result. This is particularly likely if, as we have argued in Canada's case, the 'social optimum' includes a degree of regional balance.) If we distinguish different levels or strata within information-processing firms, in terms of certain characteristics of the jobs involved, then we find that telecommunications/computer technology has the potential for influencing the spatial pattern of employment in different ways, depending on the job level in question.

To date, available evidence would indicate that there has been a tendency to centralize top-level management and highly skilled, specialized personnel. Work in the U.K. indicates that the "emergence of the so-called 'post-industrial society' . . . has resulted in an increase rather than an amelioration of regional inequality, especially in key occupations." {11} A 1976 Economic Council of Canada study has similarly indicated an under-representation of service employment in the out-lying regions of Canada. {12} This tendency, if continued, has implications for employment in periphery regions but, more significantly, for relative regional income levels.

Models of the wage structure of hierarchic organization suggest that firms can be seen as made up of levels of supervision. If the wage at each level of the hierarchy is proportionate to the total wages of those employed at the level below, Lydall {13} has demonstrated that the 'upper tail' of the distribution of earnings will follow a Paretian distribution (as is approximately true in practice). If the "span of control" of a supervisor increases, so too will the supervisor's relative wages. Thus, as telecommunications/computer technology enables the centralization of supervisors and other key personnel, given the greater ease and efficiency of supervision from afar, and an increase in the span of control of centrally located supervisors, one would expect to observe greater inequality of earnings for full-time employees at a national level (as has occurred in the U.S. {14}). Greater skewness in the national distribution would be accompanied by a truncation of the job structure of periphery regions and, since high wage jobs are increasingly centralized, *ceteris paribus,* one would expect increased inequality in the national distribution of earnings to accompany greater divergences in average incomes between regions. Politically, one might also expect to find that as the local population in periphery regions changes in character due to the change in job structure and that as "local" decision-makers get replaced by direct lines of authority from afar, regional alienation increases.

If one considers parts of firms, rather than necessarily the firm as a whole, however, it becomes clear that modern technology has made various "routine" office units independent of location, subject only to the firm's organization/supervisory structure and needs. The possibility of 'hiving-off' parts of a firm and relocating such office units to periphery regions is a very real one. While not motivated by economic concerns, this in a sense is precisely what the federal government is doing with its programme of regional decentralization of federal government offices. U.S. studies have shown that it may be highly profitable for certain companies to decentralize within a metropolitan area from the downtown core to a ring of suburban satellite offices. {15} While regional decentralization implies greater distances, such findings are not likely to change greatly in the regional case, especially if greater use of satellite technology causes the cost of telecommunications to become more independent of distance.

The potential exists in office units, where face-to-face supervision of parts of an organization are only really required on an occasional basis, for a regional decentralization of organizations which could yield significant private and social savings. Firms which locate in periphery regions can expect, on average, to incur considerably lower space and labour costs -- in 1976, for example, keypunch operators in Halifax-Dartmouth (a relatively high wage area for the Maritimes) earned 12.7% less than similar workers in Toronto; female file clerks earned 15% less. {16} More highly skilled operatives had lower wage differentials (partly because they tend to be more mobile workers) but it is precisely these "lower-level" information processing jobs which are most amenable to relocation. Goddard has suggested that the efficiency of certain types of information-processing functions might be increased through regional decentralization. {17} In addition, employment created by office relocation may substitute for other efforts to increase regional employment opportunities. Against these private and social benefits one must put the increased cost of telecommunications services consumed and the cost (which might be largely a psychic one) of restructuring the traditional organizational/supervisory structure of firms.

A widespread regional decentralization of office employment hinges on the possibility of substituting telecommunications for face-to-face meetings and/or the transmission of written messages. This, in turn, is a function of the elasticity of substitution between telecommunications and face-to-face communication and between telecommunications and written communication. In practice, however, attitudinal factors will play a major role in the success of such a decentralization programme. Traditional management style tends to foster a centralization of workers, to keep them "under the eye" of supervisory personnel, as it were. Many managers today have proven to be resistant to the use of new information-processing technology, due perhaps to irrational prejudices against "machines versus men", perhaps to an anxiety that their own skills may become obsolete and perhaps also to an attachment to such traditional social roles in the office as that of the secretary (who types and gets coffee) and the executive (who doesn't). {18}

Government Policy

Potential may exist for promoting employment in periphery regions through relocation of office units (i.e., a decentralization within firms), but research is required to quantify the feasibility of such relocation and education may be necessary to make these benefits widely appreciated. Such research, at this stage, is probably most usefully done in the format of case studies or 'pilot plant' experimentation, whose results have, in any event, a greater saliency among business decision-makers than aggregative results. Once office relocation has joined the toolkit of regional planners, its costs (in possible telcommunications subsidies) must be compared to the costs of alternative subsidy programmes for decentralizing employment.

On the negative side, however, the impact of new technology on regional income distribution arising from the probable general tendency to centralize key personnel, may create new stresses in the form of a more unequal income distribution and regional alienation. In addition, as communications/computer technology confers economies of scale on firms adopting such technology at the same time as the ability to adopt such technology on a stand-alone basis is itself a function of existing scale, small local firms in periphery regions may find themselves increasingly threatened, in terms of competitive market position, by their own inability to afford the technology. {19} It may thus be necessary for government to encourage and in some cases subsidize the diffusion of the technology to local firms in periphery regions by such measures as providing soft-ware counselling services, encouraging the greater availability of time-sharing operations, making efforts to ensure the supply of an adequately trained labour force to fill the new job requirements called for by the technology, and perhaps offering financial assistance to small firms to allow them to make use of the technology, i.e., to ensure access to the hardware.

As the potential grows for centralized, but essentially foot-loose, computer and telecommunications facilities designed to serve new market areas (the videotex market, for example), government policy makers will have to remain cognizant of the need to encourage a "dispersion" of such facilities so that some of these, at least, are located in periphery regions. Otherwise, a system such as Telidon could very well supplant local publishing houses, parts of local libraries, etc., and if principally located outside periphery regions, provide no substitute employment or industrial activity.

While the provision of telecommunications plant has, historically, been quite adequate within periphery regions in Canada, expansion of use along the lines implied by various of the above policy considerations might in turn require greater coordination in planning of future telecommunications plant between carriers and government and government intervention to assist in provisioning of plant capable of utilizing computer/communications technology. Exchanges still served by step-by-step switching equipment cannot, for example, accommodate data transmission.

If telecommunications policy is to become a focus of regional development policy, then regulators are going to have to be cognizant of this new focus and, ideally, receive some specific policy directives from government rather than being left to make policy themselves.

Beyond this, the concerns expressed above that (a) essentially "footloose" computer installations not gravitate, by inertia, to central areas of the country; and (b) that accessibility to both hardware and software facilities be emphasized, for small-scale local enterprise in periphery regions may require regulatory bodies to reconsider some aspects of the monopoly position of telecommunications carriers, permitting more private-line services and, perhaps, increasing access to the services of the public-switched network, i.e., allowing more interconnections. To the extent that new service offerings such as Telidon are brought within the regulatory umbrella, one must be aware that the employment creation it entails may be in quite different regions from the existing employment it displaces.

Through all of this discussion, it must be remembered that regional income redistribution is an existing government objective. In seeking to achieve this objective, government is already undertaking numerous programmes which, many would charge, have a high cost in terms of efficiency losses. Thus even if one believes that, on balance, the types of policies discussed here imply a net efficiency loss, these policies may still be superior on efficiency grounds to many of the traditional regional development policies of governments in Canada.

This paper has not sought to provide a definitive answer to the question of what impact telecommunications/computer technology has had or will have on the spatial pattern of employment and incomes in Canada. Indeed, to attempt to do so would have been well beyond the scope of a conference paper examining an area which is in the midst of very substantial technological change. What the paper has sought to do, however, may be summarized as follows:

1. the information processing sector, which has traditionally been largely ignored in regional development efforts, is now a potential area for government regional employment policies;
2. telecommunications/computer technology, both that which is presently available and that which is likely to emerge in the foreseeable future, is an essential ingredient to such an emphasis on the information-processing sector;
3. such an emphasis could both promote the goal of greater regional equality and greater aggregate efficiency in the allocation of Canada's scarce resources;
4. such an emphasis could see the emergence of a regional development policy based on technological improvement and innovation, centered on creating indigenous growth of output and employment in Canada's less developed regions -- a policy consistent both with those who advocate that technology in general is the key to Canada's future economic growth and those who advocate that indigenous growth in the less developed regions is the key to a permanent solution to regional inequality in Canada.

FOOTNOTES

1. See, for example, Economic Council of Canada, Living Together, A Study of Regional Disparities (Ottawa, 1977).

2. The Canadian federal government is at least nominally committed to redressing regional disparities.

3. Economic Council, op. cit.

4. The federal government's own government office decentralization programme is an obvious exception to this.

5. For a theoretical discussion of the externalities of telephone service, see R. Artle and C. Averous, "The Telephone System as a Public Good: Static and Dynamic Aspects," The Bell Journal of Economics and Management Science, Vol. 4,1973; L. Squire, "Some Aspects of Optimal Pricing for Telecommunications," The Bell Journal of Economics and Management Science, Vol. 4, 2, 1973; and Jeffrey Rohlffs, "A Theory of Interdependent Demand for a Communication Service," The Bell Journal of Economics and Management Science, Vol. 6, 1, 1975.

6. As an example, the recent increase in free tourist line information/reservation systems such as Nova Scotia's "Check Inns" system represents an attempt to utilize telecommunications technology to promote a more rapid rate of growth of the tourist industry in periphery regions. Basically, such systems reduce the acquisition costs of information for tourists.

7. Automobile parts distributors are a good example of an industry in which such a centralization of inventories has already taken place.

8. M.U. Porat, "Global Implications of the Information Society," Journal of Communications, 28, 1, 1978.

9. B. Lesser and L. Osberg, The Importance of Telecommunications to Regional Economic Development, Phase I Report, Government Studies Programme, Dalhousie University, 1978.

10. Statistics Canada, The Input-Output Structure of the Canadian Economy, 1961-1971, Ottawa, 1977.

11. J.B. Goddard, "Organizational Information Flows and the Urban System," Economie Applique, 1, 1975, p. 126.

12. Fernand Martin, Regional Aspects of the Evolution of Canadian Employment, Economic Council of Canada, Ottawa, 1976, esp. pp. 33-39.

13. See, for example, J. Nilles, et al, The Telecommunications - Transportation Trade-Off (Baltimore: John Wiley, 1976).

14. H. Lydall, "The Distribution of Employment Incomes," Econometrica, 27, 1959.

15. See, for example, P. Henle and P. Ryscavage, "The Distribution of Earned Income Among Men and Women, 1958-77," Monthly Labour Review, Vol. 103, No. 4, April 1980, pp. 3-10.

16. Labour Canada, Wage Rates, Salaries and Hours of Labour--Toronto, October, 1976; Wage Rates, Salaries and Hours of Labour -- Halifax-Dartmouth, October, 1976 (Ottawa: Supply and Services, 1977).

17. Goddard, op. cit., p. 144.

18. See, for example, R.N. Rappaport, Information for Decision Making: Quantitative and Behavioural Dimensions (New York: Prentice Hall, 1975); B.G. Champness, "Attitudes Toward Person-Person Communications Media," Human Factors, 15, 1973; and B. Christie and S. Holloway, "Factors Affecting the Use of Telecommunications by

Management," Journal of Occupational Psychology, 48, 1975.

19. See, for example, J.B. Goddard, "Technology Forecasting in a Spatial Context," Futures, April, 1980.

LIMITATIONS OF CONVENTIONAL APPROACHES TO REGULATION: THE PROBLEM OF SOCIAL REGULATION IN RATE AND SERVICE BASED DETERMINATION

LIORA SALTER

Simon Fraser University

Introduction:

Several years ago, a series of procedural reforms were introduced in telecommunications regulation shifting the regulatory emphasis, in part, from economic to social regulation. Although the CRTC acted in telecommunications questions under the relatively narrow jurisdiction of the Telecommunications Act,(drawing largely from the Railway and National Transportation Acts that had set limits for regulation by its predecessor agency in telecommunications) the CRTC sought ways of considering how rates and services would affect consumers, northerners, special interest groups and the practice of regulation itself. In doing so, it strained against the limits of conventional analyses of telecommunications regulation.

The reforms were based in part on a series of studies and recommendations from the Consumer Council of Canada, published between 1970 - 1972. (1) The Council's studies argued that regulation represented a process of negotiation of interests and, more significantly, that consumers were excluded from that negotiation by barriers of cost procedure. The Council examined broadcasting regulation, in general favourably, (2) but drew most of its recommendations from other regulatory agencies, calling for the funding of consumer representation, a practice not done in broadcasting.

The CRTC also drew upon its own experience. Broadcasting regulation is, in its first instance, social regulation and economic questions are considered, albeit not always fully, as means of implementing the social and cultural provisions of the Act. Nonetheless, the CRTC was convinced that social questions could be combined with economic consideration in a licence or rate hearing. The relatively accessible broadcast hearings, the non-legalistic tenor of the proceedings, the widely cast scope of discussion in the hearings were all considered points to be emulated in telecommunications regulation.

The new procedures were also introduced after the Telesat-TCTS hearing, although they were under consideration before that hearing. The representation of the Consumer Association and of Inuit Tapirisat in the Telesat-TCTS hearings had been central in the final decision of the CRTC, although the agency struggled against the limitations imposed by its mandating legislation. (3) The final disposition of the issue notwithstanding (the decision of the CRTC to reject the merger between TCTS and Telesat was eventually overturned by Cabinet), the CRTC remained convinced that questions raised in the Telesat-TCTS hearing by those representing advocate interests were central in any proper determination of public interest. While instituting new procedures could not prevent the further occurrence of Cabinet actions with respect to CRTC decisions, they could address directly procedural problems faced by the agency in the Telesat-TCTS or similar hearings.

.... 3

The reforms, then, were centered on two goals. The first was simply the inclusion of consumer representation in the regulatory process. Provision was made for better access to information, for funding consumer groups, for better public notice and for a number of measures that would facilitate the recognition and representation of client group interests.

A second goal, one much less well specified, was the encouragement of the participation of an active and diverse public in the determination of public interest. Although telecommunication hearings were not made significantly more informal, provision was made for regional hearings at which general issues could be discussed and the general public act as participants. It was suggested that these regional hearings need not be tied to the consideration of a specific application. They were seen more as a sounding board, alerting the agency to problems in telecommunication services.

Whether or not the CRTC recognized it at the time, the two goals were based in very different, albeit not necessarily conflicting, views of the regulatory process. The goal of increasing consumer representation was rooted in the perception of regulation as a negotiation of interests. From this perspective, it was argued that the interests of the applicants were, by definition, well represented. It was assumed that the interests of major users of service, of industries dependent upon telecommunications services and concerned about the rate structure, would also be represented. What was lacking was adequate and consistent representation of the individual clients of services. The consumer, northerner or member of a special interest group would not likely be represented unless his or her interests were aggregated and then brought forward by an advocate group.

Traditionally, the agency had been seen as the representative of those aggregate interests, in as much as they were assumed under the mantle of public interest. If however, as it was argued at the hearings on the new telecommunications procedures, the agency was to act as the arbitrator of those interests, it could not also properly be their representative. The inclusion of advocate intervenors, then, as full partners in the regulatory process funded through an assessment of costs, enabled the CRTC to remove itself from the conflicting roles, (if not totally from conflict of interest situations). Necessarily both financial and procedural assistance would be required if advocate intervenors were to take up the task.

The second goal, the encouragement of an active and diverse public in the determination of public interest, was rooted in the view that regulatory bodies were instruments of delegated legislation. From this perspective, the role of regulation was the administration of legislation. Necessarily, by specifying what was to be administered (in what way), the agency assumed a policy making function. The choice of making policy through a case by case consideration of applications or through policy statements and hearings could be left to the agency's discretion. The limits of the policy making role were set by courts, and in many agencies, including recently the CRTC, by Cabinet directive. Within the scope of its jurisdiction and mandate, however, and unless directed otherwise by Cabinet, the CRTC could use the hearing as a locus of policy determination, for the discussion of how public interest might be specified in practice.

Obviously, from this perspective, negotiation of interests was only one factor in determining adequate policy in the

.... 5

public interest. The representation of minority or individual interests, questions pertaining to the development of future services, the capacity of the agency to perform its regulatory functions and the impact of telecommunication services were all equally part of a proper consideration, as long as the agency did not exceed its mandate and jurisdiction. The Telesat-TCTS case is a good example, since all of these considerations were included in the agency's final decision on the application before it.

As a forum for the determination of public interest--and not simply the negotiation of interests--the hearings had to attract as many points of view as possible. The individual member of the public, other agencies or departments of government, other levels of government and even the agency itself were all legitimately actors in the process, regardless of the nature and the representativeness of their interest. Aggregate interests had to be distinguished from collective interests.

The Broadcast Act and the procedures adopted under it by the CRTC were well adapted to the second goal, but much less appropriate for the first. The informal hearings encouraged widespread participation from the lay public and from other governmental bodies. The debate like structure encouraged discussion about the nature and determination of public interest as applied to specific applications. The policy hearings held by the CRTC gave both individuals and group representatives the opportunity to make explicit their arguments for criteria to be applied in licencing decisions.

On the other hand, consumer and other aggregate groups seeking a proper negotiation of interests were frustrated

.... 6

by the informal procedures and demanded more procedural rigour, both before the agency and through the courts. (4) Negotiation of interests requires full disclosure of information by all parties, cross-examination, appeals to the courts. Consumer advocates argued for each of these procedures at the public hearing discussing the first draft of the proposed telecommunications procedures, fearing, rather than supporting, the transposition of broadcast style regulation into telecommunications. Of course, representation of aggregate interests in a negotiation process also demands high levels of expertise and often legal counsel. The price on adequate representation of these interests is often the exclusion of the general public as individuals.

The new telecommunications procedures, on the other hand, were well suited to increasing consumer representation, but ill suited to determining public interest through debate. In telecommunications, the enabling legislation defined the scope of agency actions clearly; the procedures, while costly and usually requiring legal expertise, allowed the hearings to be used effectively to elicit and debate information. The current demand for further reform of telecommunications procedures now centres only on how the cost allocation procedure (funding advocate groups on the basis of their contribution to the decision) should be implemented. What has made the procedures work for consumer representation weighs against their effectiveness for social regulation, for generating widely cast discussion as part of a determination of public interest.

As a result, telecommunications regulation deals with social questions, at best, only as an adjunct to economic ones, as an overlay of problems to be solved, if possible, within the structure of a hearing centred on costs and benefits. The reforms have meant greater regulatory capacity and a relatively

.... 7

clear allocation of the costs of regulation but their shortcomings are less easily measured by the tools available with economic regulation. Whatever the agency was not doing with respect to social regulation would go unnoticed if those for whom it was important to debate social questions were not participants in the hearings.

This situation has only recently become problematic. The distinction between broadcast and telecommunications procedures has served the agency well, despite its shortcomings, by tailoring the telecommunications regulatory process at least to the demands and expectations of its participants. Such is no longer the case. The shortcomings in procedure now make explicit and exacerbate a crisis in regulation. What is to be regulated, under which mandate, with whom as participants, in what kind of debate have all become decidedly unclear.

In this paper I will argue that telecommunications regulation has generally been viewed from too narrow a perspective, even given procedural reforms that broadened the scope of agency action and the role of the public. Several factors have caused the problems, specifically the participation of new groups, the growing impossibility of making distinctions between telecommunications and broadcasting and the dimensions of the new communication systems that defy enclosure within current regulatory practice. In this context, the reforms, which entrenched the division between broadcasting and telecommunications, actually compound the problems.

In light of recent developments in communications,, I will argue that telecommunications regulation must be seen as instrinsically social. As well, I will suggest that regulation in Canada is necessarily highly politicized, ill-defined and fluid. Thus, any adequate approach to solving current problems in telecommunications regulation necessarily takes account of both social and political dimensions of regulation. I will make a number of specific suggestions about how this might be done.

The Nature of Current Problems:

Recently, the CRTC held hearings in Vancouver to consider application by B.C. Telephone for a rate increase. The hearings lasted seven weeks and produced six feet of shelved documentary materials. The situation with respect to B.C. Tel's financial position was complex; the corporate relations, the history of the company's technological growth and investment, and the introduction of new systems all made the determination of just and reasonable rates difficult. What made the hearings so long, and so newsworthy, however, was not the complexities of the economics of B.C. Tel. Rather, the hearings were used to discuss the current labour management dispute in B.C. Tel, regional and rural-urban disparities, the effect of the introduction of new technologies on small communities and the employment practices of the Company.

In some senses, all these questions were outside the jurisdiction of the CRTC; those running the hearing had cause to believe the hearings were being used for a staging ground by actors on both sides of a political campaign. The fact remains, however, that the labour practices of the telephone company, the introduction of new technological systems, the use of capital resources for particular forms of expansion and even the employment practices of the company

.... 9

do have direct bearing on the financial position of the company. Thus they can be considered relevant in an assessment of an application for a rate increase. The pattern of service expansion, if indeed it discriminates against those in small communities as intervenors claimed, is within the jurisdiction of the agency and certainly questions about the quality of service were appropriately addressed in the hearings.

It was clear from the hearings that the CRTC had trouble drawing distinctions between what was applicable and what was not. The CRTC seemed not master of its own procedures by any means. The problem arose, not so much because of staff or the CRTC hearing chairman, but more because the points raised by intervening groups were <u>simultaneously</u> relevant and not relevant to the CRTC consideration at hand. Even had the intervening groups wanted to make a separation between what was within and outside the jurisdiction of the agency, it is unlikely they too would have known where to draw the line.

It would not be surprising if the B.C. Tel hearings were the first of many such hearings. Social questions simply can not be reduced to consumer issues within such hearings, nor are advocate groups always limited to consumer or aggregate representation of interests. The relatively easy access to public hearings afforded by the CRTC to labour groups did not itself account for their presence there. Nor did the regional hearings, in this case closely tied to an application before the CRTC, deal simply with matters within the purview of a telecommunications mandate. Increasingly, system expansion plans of telephone companies and the introduction of new technologies will force hearings beyond the traditional consideration of rate-related factors into the realm of social regulation. Even within a narrow mandate, agencies like the CRTC will be pressed to consider relevant broad questions relating to priorities set by major companies in their system expansion.

The second pressure forcing economic regulation into the realm of social regulation is structured into the nature of communication regulation itself. With the introduction of new information-based technologies and the design of complex systems for delivering information, the distinction between what constitutes broadcasting, thus regulated under the Broadcast Act, and what should be regulated as telecommunications becomes difficult. Responsibility, jurisdiction and control become difficult to assign.

Traditionally, the definition of broadcasting has three elements: broadcast regulation deals with "programmatic materials", broadcast between a point of origination and many receivers simultaneously ("point to mass"), using radio waves at some stage of the distribution of the signal. This definition, although tested in several court cases, has never been without controversy, of course. The prohibition of audio carrier services on cable, the distribution of signals via satellite (similar in function to a cable system, using radio waves, yet regulated as a carrier) and of course pay television all cross the jurisdictional boundaries between broadcasting and telecommunications. Yet the definition has been workable until recently.

The distinction between broadcasting and telecommunications disappears, at least in part, however, when a cableand videotext delivery system provide identical news teletype service and that service is also read as "news" by broadcasters. Cable, videotext delivery and broadcasting are regulated under fundamentally different mandates and with different rules and procedures.

Most of the Telidon type experiments have been put in place by the telecommunications carriers and it has been assumed that a Telidon system will function as a telecommunications service. Yet, although new Telidon type experiments promise two way communication, the information that subscribers access, at the present time, is on a continuous loop. Thus, the individualized information ("point to point" communication) is, in fact, mass produced and packaged. What is individualized, thus "point to point" communication, is simply the process of access -- when people get which information on the loop. In an important sense, Telidon is "point to mass" communication.

The production of what might be called national programmatic services, the children's programme Galaxie or the proposed national multi-lingual service to be carried on satellite, bring cablecasters into the traditional preserve of broadcasters. When broadcasters and others then argue that such an incursion represents unfair competition, cablecasters respond that they have been "broadcasting" for years, at the CRTC's insistance, on the community cable channel. Cablecasters, in turn, face the possibility now of rate of return regulation, regulation partially on the same basis as the telephone companies. Whatever the justification for rate of return regulation, and there are many, the new programmatic services produced by and distributed over cable confuse the regulatory picture.

.... 12

That picture is distorted further as the traditional carriers, the telephone companies, move into producing information. The information programming produced is certainly an element in the single system regulated under the Broadcast Act, although it is regulated as a carrier service when provided by the telephone companies and regulated as part of the single system of broadcasting when operated by cable.

The legal basis for federal jurisdiction over cable, and indeed for broadcasting as a whole, rests on the airborne distribution of the signal at some point in the process. Provincial governments have taken their demand for control over communications to court on the grounds that cable systems represent intra-provincial systems and only sometimes involve signals transmitted by herzian waves. Although provincial governments have lost in all court cases to date, the basis of the court decisions rendered has, in all cases, rested on the necessary integration of broadcast and cable services into a single system. Jurisdictional questions become confused when the systems for delivery of signals become more complex than either cable or broadcasting. When, for example, programmatic and non-programmatic material is beemed to satellite (regulated as a carrier) from both cable and broadcast controlled origination points, then distributed by either cable or transmission, probably on the basis of a pay per programme charge, the communication system has become so complicated that sorting out telecommunications and broadcast responsibility becomes impossible. The new systems may be licenced as "networks" but it is unclear whether such networks are simply distribution systems, thus carriers, or new "superstations" or, even the equivalent of new forms of "cable" systems.

.... 13

The situation is complicated yet further by the current federal-provincial dispute over jurisdiction of communications. One jurisdictional solution, although perhaps unworkable, to the above connundrum may well be to consider the whole system part of broadcasting since whatever is distributed over the new networks most certainly affects "the single system." Because provincial control over communications is being asserted primarily on the basis of a distinction between broadcast and cable/information services, this jurisdictional solution would likely prove unpalitable. The satelite's status as a carrier and the complexity of the system ensure such a solution would quickly be tested in the courts. Yet, the legal basis for a clear resolution of the issues involved simply does not exist. X

The third pressure bringing social questions directly into telecommunications regulation has to do with the nature of the new technologies themselves. It would be fair to characterize the new technologies as communication systems. The integration of a variety of separate technologies used to deliver any one service, the massive scale of technology involved in each case and the inter-dependence of one service and the system design as a whole mean that any one service or the introduction of any one new technology cannot be assessed in isolation. The B.C. Tel rate case is a good example, since what was being debated across the hearing floor was the <u>chain reaction</u> within the company and within the province following from the introduction of the new technology. Neither the intervenors nor the regulators could draw clear boundaries between economic or social effects since the economic effects shaded into social consequences that were certain to have both direct and indirect economic effects of their own.

The delivery of extensions of service into remote communities in the north is another case in point. The technological design of the system as a whole will set the capacity for that system to respond to social and economic pressures. The technological system, then, has a profound biasing effect, one felt both in current and future economic decisions and in the social impact in northern communities. The kinds of technologies involved in extension of service are both technology and capital intesive. Their main characteristic, once in place, is their rigidity. The cost of changing the capacity of the system, once installed, make response to new economic or social conditions exceedingly difficult. Thus, whatever capacity is not built into the system, as it is designed and first installed, is unlikely to be provided later. The technologies involved cannot easily be retrofit to new needs and conditions.

Of course, it is impossible to gauge potential needs and future costs accurately; moreover the cost of making provision for unused capacity - that may or may not be needed in the future - is often prohibitive and certainly not attractive to those investing in the system. The likelihood that the initial licencing of the new systems may be the last opportunity to shape the social impact of those technologies in a significant way has to be weighed against the reluctance of investors to accept the risk involved in building a system with unused capacity and potential.

The original licencing of ANIK services in 1973 in the north caused immediate backlash as it became apparent that no provision had been made for either intra-northern communication or significant northern origination of programming. The provision of the service by the CBC, operating with a budget that only allowed for the delivery of southern originated signals to the north, gave the CRTC little regulatory scope to redress the problem.

Current applications for extension of service to northern and remote communities are similar to the original ANIK plan. Simply put, they provide little opportunity for inter-northern service or for northern origination in a new national service. Clearly, the costs of providing northern based service are high and the risks involved to the applicant companies are great, given the current potential for northern originated programming of interest to all subscribers of the new service. The technological system involved in delivering service to northern and remote communities is capital intensive, major in scope and certainly poses economic risk to its investors. Thus, the flexibility within the system design to allow for northern service either now or in the future is not great. What is likely to be put into place as a new communication network in Canada is a highly rigid system.

In this context, the persuasive power of the regulatory agency may well be stretched beyond its limit. Yet the lack of regulatory action with respect to these potential services will result in their probable exclusion at any point in the future. The social implications of the decision to be made are immense, especially given the stated goal of the new service and the nature of the northern, and in particular native, communities to be served. The agency has few tools, even under the Broadcast Act (which only applies in part) to deal with the social questions involved.

How then are social costs to be assessed and weighed against economic costs? The problem, clearly, has assumed more urgency given the nature of the new technological systems, the breakdown of operational distinctions within regulation between telecommunications and broadcasting and the possibility that social questions will increasingly be raised in hearings by new participant groups. Yet little in the conventional analysis of regulation or in current approaches provides much guidance.

First Assumptions:

The new procedures in telecommunications regulation were based on several assumptions. First, they implied that social factors could be included in an assessment made by the agency on telecommunications regulation. Second, they assumed that social costs, once assessed, could be combined and assessed against economic costs. Third, they assumed that the public would contribute to the social assessment through their interventions. And fourth, they assumed that procedural measures facilitating public intervention would be adequate to permit the assessment of social factors. All these assumptions are problemmatic.

The term "social" has been used in this paper to cover a multiplicity of questions and variables that might be taken into account in regulation. This unsystematic use was purposeful, since it accurately reflects dilemmas in social regulation. In the B.C. Tel rate case, for example, social questions included the social and economic consequences for the workforce and for individual employees of making technological changes. They included the social-economic relations within the

.... 17

company that were clearly affecting both service and cost. They included the impact of the introduction of new technologies on small communities in British Columbia and the broader questions of access to telecommunication service regardless of location.

Indirectly, "social" extended to questions about the role and capacity of a federal regulatory agency to perform its functions within its mandate, particularly given the current discussion about possible reallocation of jurisdiction over the provincial telephone company.

Finally, on a more general level, social regulation includes those questions arising from the system expansion plans of B.C. Tel. In fact, the actions of any major company contemplating system expansion have direct consequences for a region or province. The telephone company, like other utilities in providing infra-structure support for development, is no exception. In other provinces, system expansion plans of major companies are usually subject to more than rate of return regulation and often more than the scrutiny of one agency. (5) In the B.C. Tel case, the CRTC was the only, if not the appropriate, locus for the debate of development related issues in telecommunications service.

Any agency has trouble with questions as broadly cast as these, including the new environmental assessment agencies that are broadly mandated. Unfortunately from the agencies point of view, the agency can neither set nor fully control the agenda for discussion. Experience now with energy, nuclear and even transportation regulation, all operating under narrow mandates, suggests that growing pressure will be exerted through the public hearing process to debate matters of broad social and political concern. (6) The determination of public interest inevitably includes a discussion of the necessity and desirability of system expansions plans if the public, often represented by counsel and acting easily within the formal procedures of the agency, choses to raise these issues. And perhaps it is desirable that they should.

What, then, about the combination of social and economic assessment? Obviously, broad social questions cannot be transposed easily into that which can be quantified or viewed in terms of a cost/benefit analysis. In fact, often such transposition alters the content of the social questions being addressed, to the point that those raising them protest loudly. No assessment technique, no matter how sophisticated (and social assessment techniques are in their infancy) can cope with social questions using a simple cost/benefit methodology. For one thing, the social landscape is altered not only by the introduction of new services or rates but also by the assessment process itself. For another, what people want or believe one day may be changed by the next; the intervening variable is the situational context within which the questions are asked and answered. Social factors are not, like cost factors, relatively stable nor do they easily respond to projection methodologies.

Third, although the public has a legitimate role in regulation, hearings often make a poor tool for social assessment. Certainly, in telecommunications regulation, viewing the public as a collection of aggregate or special

.... 19

interest groups each negotiating for their interest through the hearing, capsizes public demand into only one of its many components. Broad social questions become merely technical; the hearing is used inappropriately either as a referendum on public opinion or as a social social-scientific study. It is neither, particularly given the barriers and disincentives to participation, the easy accessibility of the hearings for political campaigns and the financial constraints of operating a full, localized hearing process.

Finally, the new hearing procedures has created a double-level process. On one hand, participation in the hearings demands a high level of expertise on the part of intervenors. On the other hand, the agency seeks and indeed requires participants from a much more diverse public. No means of integrating the contributions from each kind of participation was given in the new procedures or practices of the agency. Nor can social assessment data easily be fed into the formal hearing process, given the agency's and departmental unwillingness to testify in public on the basis of research they have conducted. The result has been a disjuncture between a highly formalized and an informal process. The point has been raised often by intervenors who note that much of the decision making activity, and certainly a good deal of the information used in making a decision, is generated and distributed outside the hearing process.

Extending the Limits of Conventional Analysis:

These difficulties have their genesis in the narrow perspective usually taken with respect to telecommunications

.... 20

regulation and social assessment within it. The analysis of regulation in Canada is heavily influenced by theoretical literature from the United States. (8) The central role in that analysis of monopoly regulation, the importance attached to means of introducting competition-like pressure in the market, and the fine-tuning of economic analysis of costs and rate structures all respond well to the regulatory environment in the United States. Without doubt, the analysis has applicability in Canada, and indeed is the basis for most papers given in this conference. The industries being regulated, the tools at the disposal of the agency and even the telecommunications mandate all bear direct resemblance to their American counter-part.

At the same time, significant differences exist between the regulatory environment in the two countries, differences that have not attracted sufficient attention in the past. First, of course, the "national interest" mandates of many Canadian agencies, and even the widely cast definitions of public interest, move regulation here directly into the political arena. Regulation in Canada is now and probably always has been an instrument of government policy, not so much with respect to competition, but more in terms of development strategies and social policies.

Second, a significant amount of regulation in Canada occurs outside the framework of agencies. As well, not all agencies hold hearings, establish procedural rights for intervenors, release their decisions or even do more than make recommendations. Thus, the regulatory environment is ill-defined and the process of regulation is both fluid and subject to a variety of pressures.

.... 21

Third, agencies in Canada have relatively little research capacity and operate in tandem with governmental departments that may not have much more. The role of the courts in scrutinizing or reviewing agency actions is limited and, although Cabinet often may issue directives to the agency, there is no formal process for legislative supervision of regulation. As well, social regulation in Canada is not simply environmental or occupational health regulation, separately developed and applied from economic regulation. Agencies mandated to conduct economic regulation may consider environment or health effects and vice versa. In short, the regulatory environment is more complex in Canada, shot through at every level with social and political constraints.

Fourth, unlike their American counter-parts, agencies in Canada are often remandated or retooled at frequent intervals. The relatively short tenure of any agency or of its mandating legislation suggests that regulation in Canada is tailored to the specific conditions in the industry to be regulated at the time the agency is put into place. The growth of new technologies, new relations within the industry or major technological systems cannot easily be encompassed and regulated within a process shaped by the dictates of a particular period. The CRTC is not atypical in this regard. Other agencies may have longer tenure, but the regulatory environment with respect to pipelines, potentially dangerous products or even public utilities . been altered significantly in each decade.

Finally, as noted above, regulation in Canada performs a number of conflicting functions simultaneously.

.... 22

Regulation mediates intra- and inter-industry conflict through the licencing process and barriers to entry and by rationalizing the introduction of new, potentially competitive services. The agency acts as the locus for a negotiation of interests, certainly including more than intra-industry interests. Regulation serves as a forum for public debate about the scope and definition of public interest. And in order to identify costs, agencies necessarily conduct some form of economic and social assessment. These functions place conflicting demands on the agency, its practices and procedures. The demands cannot be sorted out through procedural separation of functions: the actors are often the same and the issues, as noted above, almost always involve a complex of questions. As well, fine-tuning procedures has never been sufficient to permit an agency control over its own process.

In short, then, regulation in Canada is not always significantly different from other kinds of government action. Crown corporations, some forms of taxation and governmental subsidy programmes all allow government a directive role in shaping corporate activity, sometimes using a structurally independent decision making body and usually focussing on entry, rates and service. Neither public hearings, nor specific administrative procedures nor even the administrative status of the regulatory tribunal are essential in the performance of regulatory functions. Mediation, assessment, negotiation of interests and determination of public interest are functions performed by many bodies within government. In this context, it is unrealistic to assign regulatory agencies much greater capacity to handle social and economic questions than other government bodies have demonstrated.

.... 23

What makes some agencies unique, and the CRTC is a good example, is that their actions are potentially much more visible than those of other government bodies. The legislatively determined mandate, the publicly appointed decision makers, the possible hearing process, the specification (in many cases) of rules of procedure and the recourse, albeit limited, to the courts all bring the agency to the front lines of public debate. The increased availability of information on the process of decision making and on the decisions themselves makes explicit what is being done, by whom, in response to what data and under which pressures. Without doubt, then, public regulation has the potential to change the political equation. New groups become involved; different kinds of issues are raised; the often implicit routines and assumptions of those who make decisions see the light of day and are subject to scrutiny. Relationships between the players are altered as a result.

Under these conditions, the hearing process, public intervention and the kinds of procedural reforms introduced in telecommunications regulation by the CRTC all have indeed great significance. One author has attributed the disjuncture between the CRTC's and the Cabinet's decision on the Telesat-TCTS case simply to the pressure exerted by the new actors in the process, in part directly and in part by baring longstanding relations between the telephone companies and the Department of Communications and its predecessors. (9) The public hearings, and indeed the presence of advocate intervenors, does not necessarily mean, of course, that the resulting regulatory decisions will be different than if they had been made in another type of body. Simply, the decision making functions and practices can be identified relatively easily.

.... 24

By making the hearing central and intervention an integral part of the regulatory process in its new telecommunications procedures, then, the CRTC did not alter the functions of regulation so much as the pressures to which it would be responsive. The significance of strengthening the process of intervention was great. Yet the hearing itself was not necessarily central to the regulatory functions ~~necessarily~~ performed by the agency (except to the degree that the public process legitimated the highly controversial and publicly visible decisions made by the agency) Neither the hearing nor the intervenors could carry the burden of ensuring adequate resources would be made available for those regulatory functions unrelated to a publicly visible process. Other new means of making assessments, of ensuring negotiation and even of defining public interest were necessary in addition to the hearings if regulation was to be really reformed.

The new procedures focussed attention on the hearings and on the role of public intervention; they did not address the means necessary agency to meet what is now a crisis in communications regulation. What was necessary, beyond a better analysis of the regulatory function, were innovations in the practice of the agency extending beyond its rules and procedures and allowing the agency to deal more adequately with different kinds of information, different mandates, different functions and assessment. As well, it may be now that fundamental revisions in the structure of the mandating legislation are required. After all, the distinctions between broadcasting and telecommunications were not a product of the new procedures, although exacerbated by them, but a product of the history of the agency and its mandate.

Possibilities for New Directions in Telecommunications Regulation:

It was noted above that agencies in Canada are remandated and retooled at frequent intervals. The CRTC is no exception; current discussions focus on the proposed legislative changes. Unfortunately, most of these proposed changes are in the nature of fine-tuning. They bring an already politicized process further into the political arena by suggesting new modes of accountability. They attempt to encompass the new technologies and services within the legislation by adding to the current mandates of the agency. They distribute regulatory jurisdiction more widely, in light of the inroads being made by the Provinces with respect to jurisdiction. They are largely defensive in nature.

What is being argued here is that the current problems in telecommunications, and the difficulties being encountered in its regulation, are more serious than can be encompassed by the proposed legislative changes. If regulation in Canada is shaped by the conditions of the day, largely although not entirely in industry, then these conditions now argue for a fundamental reorganization of the mandates of the agency. Distinctions between broadcasting and telecommunications are not viable; the scope of regulatory action needs to be extended if the agency is to cope with the situation it now faces. The usual instruments of telecommunications regulation, largely of an economic nature, need to be expanded by the addition of new forms of assessment allowing the agency greater scope in dealing with the complexity of social and economic questions in every aspect of its jurisdiction. If the changes being encountered by the agency are historically and locationally specific, so too must be the tools of regulation.

It may be that the choice is between adopting fully the American model of regulation (and probably a fair measure of deregulation) or recognizing that the context of regulation in Canada makes redrafting a political/social mandate encompassing both carriers and broadcasters essential. Given the propensity in Canada for using regulation as a tool of public policy and the historical prcedent for regulating broadcasting to protect a cultural, rather than economic interest, it is unlikely, — although not inconceivable — that the first choice would be made. Given the public pressures for social regulation even in the telecommunications sphere and the massive impact of the new communication systems it seems unlikely the first choice would prove politically acceptable, despite current attacks on regulation. Certainly, new thinking about the role of competition and monopoly and about new competitors legislation are prerequisites of adopting the American model.

On a more specific level, certain measures would improve the current situation. First, of course, recognizing the limits of the hearing process without devaluing its contribution seems essential. The hearings are one means of considering an application, but not even the only publicly accessible one. Depending on the hearings to produce the kind of information necessary for a full social assessment ensures an inadequate assessment. Access to studies, ~~perhaps even~~ in conjunction with the heari-- process, would be helpful. Too much of the information and analysis now available to the CRTC and the Department of Communications is never debated in the formal process of regulation.

Finally, it can be argued that inappropriate methodologies are being used for whatever social assessment is now done. These assessments draw heavily on a cost/benefit methodology and use models from economics for the projection and understanding of social demand. Rather than drawing from cost/benefit analysis, social assessment would benefit from some sophisticated marketting study techniques.

As a sidenote, it is somewhat surprising that few serious marketting studies have been done on the new information technologies. Most of the available literature on Telidon, for example, is either descriptive or prescriptive. Empasis in studies available to the public is on what can or may be done with the new technologies - in other words on their technical capabilities. The existing social uses of information, uses which will shape the response to Telidon, have not been studied. If these new information technologies are to take root, a study of what information needs and uses they would replace seems in order.

Marketting studies focus on the elasticity of social demand. At best, they indicate patterns and dynamics of change, the forces that give rise to particular needs and aspirations. They take as given, as few cost/benefit studies do, that needs are changing and can be changed. They emphasize study of what causes dissatisfaction and social problems in order to evaluate potential responses to those problems. They assume that people have a variety of possible responses to problems or dissatisfaction, only some of which may involve a new product or technology. Beginning from an understanding of change and of problems or dissatisfaction, they do not assume that whatever now exists is desirable. Ironically, marketing studies, unlike the social assessment studies using cost/benefit methodologies, are not conservative. Nor do those who conduct marketting studies assume the public is conservative, only anxious to protect the benefits they now have. As such, marketting methodologies are, paradoxically, more tuned to the needs of disadvantaged communities than most social assessment research now done.

The use of marketing studies as a model for how adequate social assessment might be done to bring social and economic data together is both a suggestion and a metaphor. Certainly, better analysis of the effects of new technologies could be made by regulators if proper marketing studies had been done. At the same time, what is being argued here is that cost/benefit analysis is not the only, nor indeed the most appropriate economic methodology to use in assessing social factors in regulation.

Beginning with the assumption that communities affected by new technologies are already in a process of change would enrich the analysis. Focussing on the range of possible responses to new technologies would make the resulting analysis more sophisticated and better able to encompass the fluid realities of social life. Assuming that change may be desired by communities, and assessing alternative means of introducing change, would bring the assessment closer to the aspirations of the public now being expressed in hearings before agencies like the CRTC.

Simply introducing social assessment into the regulatory process is not enough. In fact, in some cases, it has backfired, causing greater backlash against the agency. What is introduced as assessment must correspond to the perceptions held by the public, as expressed in the hearings and in response to decisions made by the agency.

THE ROLE OF ECONOMIC THEORY AND EVIDENCE IN REGULATING TELECOMMUNICATIONS

KEITH ACHESON

Carleton University

I. INTRODUCTION

The regulatory process can be viewed from a number of perspectives. The angle of observation adopted in this paper and elaborated in section II is the following. Regulation is assumed to be a purposeful activity where objectives are rationally pursued subject to constraints.

A rational actor must have an analytical framework linking cause and effect. Orthodox economics provide such a framework. A brief outline of that framework, how it is simplified, and how it is quantified so as to make it operational appears in section III. Criticisms of that framework are then developed and it is argued that by replacing the social welfare function by a voting function the orthodox normative model can be transformed into a suitable positive vehicle of analysis for the regulator. That the characteristics assumed to hold for the voting function substantially modify the orthodox model is also developed in section V. In the next section, the C.R.T.C.'s process for gathering information to delineate the voting function and the effective constraints is explored. The paper concludes with some speculations about the consequences of the argument that has been developed.

II. REGULATION

Although regulation is assumed to be purposeful, the immediate goals being pursued may be difficult to determine. The objectives of regulation are often not comprehensively articulated, and actual objectives may differ from those that have been articulated. General statements of objectives where they exist and are accurate, may have an Ecclestiastean permanence, but the working expression of objectives and the resulting consequences for action can change significantly over time. Frequently, part of the regulatory function is to give precision and form to vaguely prescribed general objectives and to delineate the effective constraints on pursuing these objectives. This delineation is revealed in the activities of the regulator.

Regulation is part of the political process, and the objectives are defined and pursued in that context. Control over the governing process, of which regulation is a part, is gained through a competitive election. The existing regulatory structure is an element in the incumbent government's apparatus for administering and delivering political services now, and of innovating means for attaining a better set of political services in the future. These efforts reflect the government's desire to retain power and to enhance its chances of winning future elections.

In this view, regulatory commissions are part of an extensive managerial complex. Their structure, the domain of their mandate, the instruments at their disposal, their functions, and their relation to other parts of the managerial complex reflect the government's decision as to how to govern. An individual commission is granted a governance structure designed to maintain the government's competitors in the role of opposition parties.

That regulation is purposeful and that its impact is rationally determined does not mean that ex post mistakes are not made or that different structures and programmes may not have been better as judged by a number of philosophical or ethical criteria. What is meant is that government policies at any moment in time are the best choice of those who have gained the right to make such policies. The effective choice set of government depends on the detailed nature of the written and unwritten political constitution. Since the latter is also a creation of the polity, the choice set is ultimately constrained by power distributions in the society and the tastes as to its disposition of those who are able to exercise power.

What determines fundamental power endowments, how they are coordinated and the stability of the resulting social coalitions are fascinating questions that are not addressed here. For our purpose, it is sufficient that the right to govern is won by that party which most accurately identifies those with effective power, and, given that identification, creates and redistributes benefits and costs so as to be their chosen agent.[1]

The efficiency of the government in pursuing its objectives is not questioned in this approach. In fact, it is revealed by the attainment of political power, just as a firm's efficiency is

[1]The causation is not one way. The actions of government over time may alter the distribution of power.

revealed by increasing net worth. Relatedly, the government and its regulators never forego the chance of making Pareto dominant improvements as long as their contribution to attaining these changes is recognized. Making someone else better off without making someone else worse off enhances the popularity of the government. The tough issues for the government concern income distribution.

Parties that contend for the right to govern operate in an uncertain world. There is imperfect information concerning the results of different actions on intermediate targets, and on the valuation by the electorate of whatever is achieved. In the more restricted world of economic regulation, commissions are not simple instruments for carrying out predetermined policy, but are important conduits for information. Information is gathered about competing causal models, the characteristics of market participants, cost relations, regional impacts, profit opportunities and emerging technologies by the regulator acting as a window on the industry. The view is not just one way; the industry and customer groups gain information about contemplated government initiatives, and have a focal point for arguing their case and persuading the government of the efficacy of their paradigms.

The information gathered and disseminated by a regulatory agency will depend on the function it performs for the government. One function of an agency may be to provide a reward structure for party workers and political entrepreneurs through its staffing, where other forms of reward, such as lump sum payments, are legally or ethically proscribed. Little information is required to fulfill this role. In comparison, a considerable information requirement is required to implement existing policy, to alter for political advantage an industry's performance with respect to price, quality and rate of innovation, and an additional one to find new opportunities to do that more effectively in the future. Information, both factual and theoretical, is actively sought in these cases.

This paper examines the extent to which the federal regulator of the telecommunications industry in order to understand the welfare and allocative attributes of its decisions has relied on orthodox economics for its analytical frameworks, for classifications for gathering data and information, and for techniques to quantify those frameworks in order to make them operational.

III. ORTHODOX ECONOMICS PARADIGM

(i) A Skeleton of the Paradigm

In the traditional economic approach, allocation and distributional questions are answered through locating the solution of a properly posed problem. The critical elements in posing the problem are specifying the objective function which provides a measure to compare solutions, the relevant constraints and the way in which a solution is obtained.

The objective or social utility function has as its arguments the utilities of all individuals in the society from now to Kingdom come. The arguments of the individual utility functions typically include partitionable goods and non-partitionable goods. The latter goods include public goods and externality-generating services such as telephone calls. In principle, there are no a priori restrictions on the arguments that may be included, but it is considered a bit of a fudge to include many soft arguments such as nationalism, prestige and the like. A functional form for the social welfare function is assumed which summarizes the contribution of each person's well being to "society's" well being.

On the constraint side, opportunities are limited by the resources, machinery and technology available at the starting time for the problem and the technology and resources expected to be discovered or become available during the period. Technology acts as a bridge between the resources available and the goods or characteristics that individuals can consume.

Uncertainty affects both the objective function and the constraints. Individual's orderings may be dependent on future "states of the world" which are random. Who will be present to be included in future social situations, what the state of health of the living will be, and what plagues, pestilences, resource discoveries and technical innovations will occur are illustrative of the stochastic factors that must be considered.

Othodox economics contain an elaborate theory of individual behaviour under risk which explains how orderings of prospects are related to orderings under certainty. The probability distribution of outcomes for each individual across states, under different allocations of inputs and outputs over time, represents the society's menu, and the properly modified social welfare function attaches a number to each of these. Risk in the orthodox approach may reflect a lack of information or be inherent in the state of things, or both. The lack of information possibility can be accommodated by making probability distributions depend on learning through experience or through allocating resources to gather information.

Solutions to the problem under different rules for solving it are compared. The methods of solution are associated with an institutional framework in the real economy, so that, for example, "competition" can be compared to a "planned" or a "mixed" (competitive-regulated) economy. A method of solution may not be able to find an answer under particular specifications of the objective function or of the constraints. In these cases, a conclusion, such as "competition" will not "work" if the technology is super-additive, will be drawn.

What are the characteristics of the solution method which is given the label "competition" and that which is given the label of a "mixed" economy? In both, firms are introduced as production entities. Within their walls, inputs are technically transformed into outputs with perfect efficiency. A government is also introduced which costlessly enforces those transactions that are allowed to proceed and which redistributes by reassigning rights to "income" from resources or according to some tax and subsidy scheme. Individuals maximize their utility subject to the income they earn on resources they own, net of what the government gives or takes. The metaphor for competitive markets is the activity of an auctioneer who does not permit trading to occur until demands are equal to supply at the prices that he or she calls. No resources are absorbed in the coordinating process by firms, by government or by the market auctioneer. What it is possible to achieve by this method of solving the problem depends on the functions representing tastes and technology, on the fiscal technology, and also on the complexity of prices that the auctioneer can call.

In the competitive economy, the auctioneer rules over the whole domain. In the "mixed" economy, some areas are under the control of the auctioneer and others are regulated. In regulated sectors of the economy, coordination occurs under different rules. For example, where the technology is super-additive, it may be assumed that a single production entity, a monopolist, coordinates resources, and acts so as to maximize wealth subject to a regulatory constraint. The instruments that are available to the regulator must be specified. For example, as a beginning, the regulator could have the power to establish prices for the regulated goods and services, but the specification of powers must be more detailed before a solution can be obtained. Among the additional questions that must be answered are the following: what objective criteria are to guide the regulator in setting prices and under what constraints should these criteria be pursued? Is the regulator limited to linear prices, or can complex prices conditional on personal attributes, events, and behaviour be levied? How do the regulator and the auctioneer interact? Do both call their prices simultaneously or do the calls occur sequentially? Is the government's mandate to redistribute income shared with the regulator or does the government use other instruments to offset the redistributive effects of prices in the regulated sector? Can the government augment the revenue from sales in the regulated sector by subsidies so that resource costs need not be covered by revenue generated within the sector, or must the sector be self-financing?

(ii) <u>Simplification and Quantification</u>

In order to compare different coordinating regimes, the model must be quantified, and that is not possible at the level of generality so far assumed. At that level of generality, the model represents a framework for organizing one's thoughts, a departure point for making simplifications that are operational. The model is simplified either by aggregation, by isolation of a part of the economy for consideration, by dropping entirely some considerations, or, most commonly, by a mixture of the three. Intertemporal considerations may be stressed in one simplification, ignored in others. Similarly, uncertainty may be emphasized or neglected. The level of aggregation may preclude analysing the impact on income distribution or limit the degree to which that may be done. Whether restrictions placed on functional forms in the disaggregated model can be maintained when products, factors and people are grouped and what the conditions are under which pursuit of welfare objectives at a sector level contribute to the efficient attainment of welfare objectives at a global level is the subject of an extensive literature.

If positive predictions are important, the predictive ability of the simplification in the area of concern is the obvious criteria to guide an analyst. In normative areas, choice of a simplification must be guided by whether the ranking of alternative policies would be the same in the simplified model as it is in the complex model. Whether it is so, or is approximately so, depends on the nature of the complex model. If that nature must be known in detail to determine whether the simplification provides meaningful information, then one has come full circle, since it was the impossibility of discovering the former that justified the simplification. However, if a general property must be true of the complex world or a part of it, then it may be possible to test for its existence. An example would be the conditions necessary for real income changes to be measured by consumer's surplus. Conditions can be derived for the measure to be exact and the extent to which it represents a close approximation can also be determined under a range of alternate conditions.

Sometimes, quantification of part of a sector's characteristics may help in making some broad policy decisions. For example, some economists argue that determining whether technology is super-additive or not will, by itself, provide guidance in determining whether and where competition should be encouraged in the telecommunications field. For some problems, a simplified model of the sector is adopted and quantified.

(iii) <u>A Simplified Welfare Model</u>

As an example of a simplified, quantified, orthodox model designed to give policy guidance in the telecommunications sector, consider a study done by Robert D. Willig and Elizabeth E. Bailey[1] on long distance prices. The objective function is a non-weighted sum of consumer's surplus and profit which is maximized subject to a maximum profit constraint. Service definitions are given, and the cross elasticities between services are zero. Demand functions are assumed to be linear in one calculation and have constant elasticity in another. Externalities which are inherent in telephone service are ignored. There are constant marginal costs.

---

[1]Robert D. Willig and Elizabeth E. Bailey, "Ramsey-Optimal Pricing of Long Distance Telephone Services", in John T. Wenders (ed.), <u>Pricing in Regulated Industries Theory and Application</u> (Arizona, 1977).

The telephone sector is isolated from the rest of the economy on the demand side, and no changes in producer's surplus occur outside the sector as a result of pricing policy within it. The regulatory instrument is a set of linear prices. Quality levels are given and are costlessly enforced. There is no explicit time horizon, and presumably we are in a stationary state. There is no consideration of uncertainty. Information is gathered and processed costlessly and regulatory edicts are costlessly enforced.

The optimal prices under these conditions depend on the demand elasticity for each service and its marginal cost. Actual prices diverge considerably from the optimal for the American long distance services considered. In the case of constant elasticity demand curves the loss in consumer's surplus from present pricing policies is estimated to be 249 million dollars. However, because all services have inelastic demands, total consumer's surplus generated within the industry is very large and the welfare loss in percentage terms is considered modest. Realizing that demand conditions are not globally known, Willig and Bailey also outline a method for guiding a limited step towards optimal prices from the present ones. Although analytically ingenious, the guidelines are only valid for infinitessimal changes, and it is not clear for what finite size changes they remain helpful.

This model is illustrative of a family of simple welfare models that have been used to analyse different pricing alternatives in the telecommunication area. We will return to it later when we consider whether such a model is a useful structure for the regulatory to employ in designing prices.

IV. CRITICISMS OF THE ORTHODOX MODEL

Economists are often concerned with the lack of economic sophistication of the public in general and regulators in particular. A quick review of controversies within the discipline concerning the validity of orthodox economics indicates that being sophisticated may lead to mental paralysis. After years of becoming sophisticated, one may become as uncertain of about the causes of economic phenomenon, as Hamlet was of the cause of his father's death. The following list of criticisms is certainly not exhaustive, but is indicative of the universe of problems. Examples are chosen which have particular relevance to the telecommunications sector where possible.

(i) General Criticisms

The first level of criticism concerns the appropriateness of the general structure. The existence of a social welfare function, the assumption of rational and purposeful behaviour, the assumption that tastes are a primitive, a given, that are independent of the institutional framework, and the extent to which the auctioneer process is an appropriate metaphor for any of the observed methods by which resources are coordinated have all been questioned.

The social welfare function is a strange construct. If the utility functions of the individuals embody everything we know about their tastes, which should include ethical judgements, other directed feelings and the like, what is left over to determine how we should aggregate individuals? Where does the information for deriving the social welfare function come from, if all information about tastes has already been incorporated into the model?

From another perspective, the utilities of future generations are included as arguments, but how are their views on social justice knowable? One could truncate the problem by establishing an endowment at some terminal data which must be passed on to those that will continue after the planning period. Nevertheless, the problem will still emerge within the period, and how one settles on a terminal endowment is conveniently ignored. A common out for the within period problem is to assume that present individuals act for themselves and as agents for those who will be affected but are not present; it is assumed that the agents make future related decisions as their principles would have. Granting a weight to an individual's opinion in social decision making then depends on the social worth of the individual and those that are represented by the decision maker. Representation of the interests of future generations is particularly important in telecommunications where very long-lived capital is involved. These comments on the social welfare function do not scratch the surface of the documented problems with the concept.

Happily, criticisms of the logical validity of the social welfare function are not relevant to the attractiveness of the orthodox model to regulators. According to our earlier argument, the regulators are concerned with contributing to the political success of their principles; the social welfare function can be replaced by a political payoff or vote function. The arguments of the two functions would be similar, but there is no mystery about where the political function is derived from or whose opinions it reflects. Future generations, for example, will count to the extent that their interests are thought to motivate present day voters. Therefore, as far as criticisms of the social welfare function are well put, regulators should be more willing to embrace an orthodox model modified to include a vote function than welfare theorists should be to accept its unmodified form.

The second general criticism concerns the association of competition with the auctioneer's call of linear prices. Suppose that instead of linear prices, the auctioneer calls complex price vectors which identify the customer and each unit of the product or service. Firms offer to supply units that would be profitable, and consumers identify how many units they want to purchase. Firms specify whether parts of the offer were conditional on other parts being accepted. Trades are allowed to be consummated when the market for each 'named' and sequentially identified good clears.

If the prices called were complex then super-additivity does not require special attention. Competition for the field in this area would occur and even if a single 'firm' dominated, the process is competitive. This revised auctioneering process deserves the title of competitive as much as the standard auction does. In fact, in the mixed economy, the standard auctioneer could rule in the constant returns to scale part of the economy while the regulator in the super-additive area acted as an auctioneer calling complex prices for that sector. This regulatory programme would parallel "real world" suggestions by Demsetz[1] that monopoly rights be auctioned off.

The view of competition that dominates in orthodox economics is that price taking behaviour accurately describes the phenomenon. That is too narrow a view. Competition should embrace a set of rules where people are rewarded for discovering better ways of doing things. A critical ingredient to effective competition is encouraging those with imagination and ability to find dimensions where improvements can be made. Some of what is competitive behaviour in the actual economy, the competition to find better pricing schemes or better regulatory schemes is confined to the auctioneer, i.e., is ex machina in the standard paradigm.

Would it then be advantageous to have periodic auctions for monopoly rights? Tullock and Posner[2] have both argued that competition for monopoly rights occurs even without a formal auction, but in these cases the gaming for the right absorbs real resources in lobbying and there is a deadweight loss to society. Posner contends that it would be better to have an explicit auction which will avoid this deadweight loss.

---

[1] Harold Demsetz, "Why Regulate Utilities?", The Journal of Law and Economics, (April 1968).

[2] Gordon Tullock, "The Welfare Costs of Tariffs, Monopolies, and Theft", Western Economic Journal, 5(June 1967); Richard A. Posner, "The Social Costs of Monopoly and Regulation", The Journal of Political Economy, 83(August 1975); and Richard A. Posner, "The Appropriate Scope of Regulation in the Cable Television Industry", The Bell Journal of Economics and Management Science, Vol. 3, No. 1, (Spring 1972).

Posner's argument ignores that we already have an auction at a more primitive level. The prior auction for the right to set the terms under which monopoly rights are granted is a political election. If regulation is a purposeful instrument of government, there would be an incentive to scrap present arrangements of control when another instrument promised a higher political payoff. Since auctions of monopoly rights are rare, the political payoff is revealed to be higher from controlling the pricing of the services than through having periodic auctions. In fact, Williamson in an interesting case study documented that the enforcement process, negotiated adjustments and related lobbying associated with an auctioned right to provide cable TV were not significantly different from regulation.[1]

In adapting the economic model to serve a positive role, the regulator will take a broader view of competition than the orthodox economics view. Where a monopoly exists, it will be kept taut, in terms of serving regulatory ends efficiently, through the threat of replacing it in the franchise or modifying the terms of its incumbency. Political competition disciplines the regulator to make the ongoing auction efficient in terms of the vote function.

(ii) Transactions Costs and Governance Structures

The traditional model lacks an explanation of the emergence of different institutions. Even in the competitive sector, if there are cosntant returns to scale, some arbitrary delineation

[1]Oliver E. Williamson, "Franchise Bidding for Natural Monopolies - in General and With Respect to CATV", The Bell Journal of Economics (1975). See also Victor P. Goldbert, "Regulation and Administered Contracts", Bell Journal of Economics (1976).

of firm size is necessary to ensure a sufficient number of firms to make price taking appear plausible. Within their boundaries firms and governments achieve effortlessly the coordination that the auctioneer is struggling to attain between firms and consumers, and firms and resource holders. If governance structures (firms, partnerships, cooperatives, market arrangements, crown corporations, etc.) differ in their relative ability to coordinate resources depending on the nature of the measurement and enforcement problems from which transacting costs emanate, the choice of organizational form should be made endogenous to the system.

Such an extension of theory creates a number of complications. As mentioned in the preceding section, the metaphor for the achievement of equilibrium has to be altered and extended to encompass many all or nothing decisions. In addition, the choice of governance structure may affect the technology matrix for transforming from goods to characteristics as well as the more commonly considered transformation from factors to goods. Since institutions have differing comparative advantages in measuring and enforcing quality along the different characteristic dimensions of a good or a service, the equilibrium nature (i.e., their characteristic composition) of goods and services produced will not be independent of the governance structure, but will be determined simultaneously. The model cannot assume that goods are defined in terms of their characteristics; this definition will emerge from solving the model.[1] For example, the nature of telephone

[1] c.f., J. McManus and K. Acheson, "The Cost of Transacting in Futures Markets", Carleton Economic Papers, #79-22 (Mimeo, 1979).

service, its quality variation, the equilibrium flow of information (informative advertising to opportunistic misrepresentations of quality), default probabilities of suppliers, billing and credit practices, identification of responsibility for maintenance, etc. will all be affected by the choice of governance structure.

Reducing transactions costs may also be an important function of the regulator. For example, in telecommunications, "gaming" costs with respect to the terms of interconnection between different networks may be reduced by the intervention and adjudication of the regulator. Although there may be many interconnections, each interconnection is a bilateral exchange situation where strategic options are large.[1] The regulator may also reduce the costs of negotiating and monitoring a complex contract between the producer of services and various groups in the society.

The inclusion of transaction costs in any meaningful way makes a general equilibrium model unmanageable for theoretical analysis and a fortiori makes quantification of such a model impossible. Most of the literature that discusses the transaction costs problem does so in a more limited context. In the applications to market structure, the comparative advantage of different institutional arrangements have been explored. For example, the X-inefficiency group have argued that the intra-organization efficiency of large firms is less than that of small firms.

[1] Leonard Waverman, "The Regulation of Intercity Telecommunications", in Almarin Phillips (ed.), Promoting Competition in Regulated Markets, (Brookings, 1975).

Others have identified informational situations where monopoly will be a more efficient mode of organization than competition,[1] where quantities of a service available to particular customers must be limited if an equilibrium is to be achieved,[2] and where coercive restriction of entry into professions will be efficient.[3]

Many of these models depend on asymmetric or impacted information for their results, and the quantification of this phenomenon by an outside observer is by its nature difficult, if not impossible. In general, empirical work has been limited to the spinning of instructive stories, to testing by parable. More importantly, the recognition of transacting costs calls into question the relevance of cost or production functions based on the choices experienced under one mode of organization for what would transpire under another. The change in organizational mode will often change the characteristic mix of outputs and the gross of transacting costs terms under which factors of production are combined.

As argued earlier, the regulator has no incentive to encourage sloth or an inefficient production structure, <u>ceteris paribus</u>. The regulator is interested in gathering information and developing techniques to make the regulated companies efficient instruments for pursuing the regulator's objectives. From a positive perspective, the regulator is not only interested in institutional competition, in whether changing governance structures will enhance the popularity of the government or not, but also in playing a monitoring and adjudicatory role in the

[1] Y. Barzel, "Some Fallacies in the Interpretation of Information Costs", Journal of Law and Economics, Vol. 20 (1977).

[2] M. Rothchild and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Competition", QJE Vol. 90 (1976).

[3] Hayne Leland, "Quacks, Lemons and Licensing: A Theory of Minimum Quality Standards", Journal of Political Economy, Vol. 87, No. 6 (Dec. 1979).

process. The regulator is not only an ongoing designer but also a participant in the design. THerefore, a model of relevance to the regulator must consider such options.

(iii) Controversy over the Parts

The theoretical content that is inserted into different parts of the orthodox paradigm can be criticised while accepting the overall structure. For example, how to model time, how to treat capital, how to incorporate uncertainty, and how to reflect technological change have all been subject to dispute. The nature of dispute can be illustrated by considering uncertainty. The traditional approach assumes that a comprehensive outcome space can be delineated with elements that can be ranked under conditions of certainty and to which probabilities can be assigned. Critics claim that such assumptions may be appropriate for analysing behaviour in games of chance, but not for understanding behaviour in the world at large. Keynes,[1] for one, took this position, and derived a typically iconoclastic lesson from it. He argued that although the expected utility approach cannot be applied to most economic decisions, people must behave as if it were applicable if anything is to be done. He thought that people are vaguely aware of the inappropriateness of this paradigm and turn away from it in moments of panic (or lucidity). These periodic losses of faith are the causes of the instability of the private economy, and the stability of the demand for money reflects variations in the "disquietude" with this convention. Subsequent economists have

[1]John Maynard Keynes, "The General Theory of Employment", <u>Quarterly Journal of Economics</u>, Vol. 51 (1937).

shown their own "disquietude" with Keynes' social existentialism, but unease with the standard approach to uncertainty that was the springboard for his speculations on the macroeconomy remains.

The expected utility approach has also been attacked because its predictions or orderings are not consistent with choice in situations that have been constructed to have defined outcome spaces and where the probabilities are known to the decision maker. On its own turf, so to speak, of situations akin to a gambling game, the theory predicts badly. The tests are based on carefully constructed experiments where subjects reveal their choices in a number of hypothetical situations. Kahneman and Tversky[1] have summarized the relevant evidence and propose an alternative account of choice under risk.

Treatment of uncertainty in economics is a relatively non-controversial area, as compared to others such as capital theory where deep divisions exist in the profession concerning appropriate modelling. Where disputes are questions of logic, which is at least in large part the case in capital theory error will eventually be identified and progress made. Where different logical constructs can be developed and choice between them depends on normative judgements, disputes continue forever. These eternal disputes are avoided if the model becomes a positive one. In that case, a pragmatic approach to filling in the parts can be adopted.

[1] Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk", Econometrica, 47, No. 2 (March 1979).

(iv) Quantification

Over the past twenty-five years, the discipline has evolved from a position where there were few texts in econometric methods and where a knowledge of that subject was rare among professionals to a situation where there are many excellent econometric texts, and where courses in relatively advanced econometrics are often mandatory in honours undergraduate programms and are universally required in masters programmes. The discipline of econometrics itself has experienced rapid development and its applicability to problems has been enormously enhanced by the simultaneous development of computer technology.

During this period, there was an optimism engendered that theory would be beneficially disciplined and made more credible by these developments. In my opinion, this promise has not been fulfilled. Despite the elaborate hardware and software, no major dispute has been settled by econometric work. With the many degrees of freedom offered by different 'simplifications' of general theory, different test specifications of these simplifications, different proxies for theoretical concepts, different estimation techniques and the reduced cost of experimenting with many permutations and combinations of these, the sets of results that appear consistent with experience has, if anything, widened.

Even a simple problem such as whether the demand for money is sensitive to the interest rate and to what extent has received a number of different published results; with old chestnuts such as the appropriate theory to explain the term structure of interest rates, the jury is still out and is getting more confused as the stock of empirical studies expands.

Consider the telecommunication regulator who is persuaded that economies of scale is important. What is the empirical evidence concerning this phenomenon? There are a number of studies that have been made, and the results vary from finding no evidence of scale economies to finding evidence of very substantial scale economies. As the production functions and related cost functions become more complex we find the same authors revising their estimates within a short period of time. These studies are all in the neoclassical tradition of assuming well behaved aggregate cost and production relations exist, so that variation in results is not due to "unorthodox" views on production functions. Some studies used cost minimizing conditions or profit maximizing conditions to improve the precision of the estimates. Unfortunately that precision is bought at a cost since a joint hypothesis is now being tested. All the studies employ the same data sources and therefore variability in results from using different series for capital and labour, for example, is minimal. Nevertheless, by disaggregating between one to three service categories, changing specifications, and adding auxillary conditions a wide range of estimates is derived.[1] The regulator's confusion might

[1]J. Carr, "Demand and Cost: An Empirical Study of Bell Telephone of Canada", in H.E. English (ed.), Telecommunications for Canada (Methuen 1973); A.R. Dobell, L. Waverman, T.H. Liu & M.D.G. Copeland, "Telephone Communications in Canada: Demand, Production and Investment Decisions", Bell Journal of Economics and Management Science (Spring 1972); M. Fuss & L. Waverman, "Multi-product, Multi-input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada", NBER Conference Paper (1977); J.B. Smith & V. Corbo, "Economies of Scale and Economies of Scope of Bell Canada", Institute of Applied Economic Research, Concordia University, Final Report to the Department of Communications (March 1979); M. Denny, M. Fuss & C. Everson, "Productivity, Employment and Technical Change in Canadian Telecommunications: The Case for Bell Canada", Final Report to the Department of Communications (March 1979).

be further enhanced when it was brought to his or her attention that with the same techniques, a study of the unregulated manufacturing sector indicates that it is characterised by economies of scale.[2]

Fortunately, the regulator is not limited to formal statistical procedures for quantifying its model of the world. Again, it seeks information by many means and is not concerned with the social scientist's constraint that others be able to duplicate the results. Regulatory processes can only be understood if they are perceived as an information generating and testing process.

---

[1] Ernst R. Berndt and Mohammed S. Khaled, "Productivity Measurement and Choice Among Flexible Forms", Journal of Political Economy Vol. 87, No. 6 (December 1979).

V. THE REGULATOR AND ECONOMIC THEORY

By substituting a vote function for the social objective function, the orthodox normative model can be transformed into a positive model for studying regulatory behaviour. So transformed, the general orthodox economic model would be useful for the regulator as an organizational framework for gathering information and checking interdependencies, and various simplifications would provide operational significance in different areas. Disputes about how to treat capital or uncertainty or any other part of the model would be monitored and different simplified, modified, welfare models would be applied, if they passed the acid test of suggesting policies that were politically rewarded.

In this context, there are two aspects of a vote function which are important for understanding regulatory behaviour. First, the use of a different simplified variant of the general model for analysing different problems such as distance tapers in the toll price structure, peak-load pricing, access pricing, pricing of inputs to value-added carriers, or "sustainable" pricing is appropriate for the regulator viewed as a purposeful political actor. However, if apparent consistency in approach is also valued politically, the regulator cannot openly justify decisions based on different simplifications. A regulator cannot declare that cross elasticities among services can be ignored in pricing long-distance rates, but must be taken into account in deciding sustainability and entry issues in that area; cannot consider uncertainty and optimal rationing in a peak load pricing problem but ignore uncertainty in examining other pricing problems, consider externalities in access pricing but assume that these are taken into account through private arrangements between subscribers as far as services are concerned.[1] Such apparent inconsistencies would be exploited politically. The actual rationale may be it works politically, but to state that would be politically counter-productive. Therefore, if a regulator does apply economics in this manner, disclosure of that fact would not be made, although one might be able to deduce it from examining the decisions made.

Secondly, the political objective function concerns the change in the welfare of voters that those voters attribute to the policy initiatives of the authorities. Modifications of standard representations of objective functions that ignore this recognition factor will not be effective in understanding regulatory decisions. Consider, for example, the positive pursuit of redistributional objectives by adapting the simplified model of Bailey and Willig described earlier. As those authors note, the objective function can be altered to accommodate distributional targets by assigning appropriate weights to the consumers' surplus of different groups. An answer to the optimal pricing problem modified in this manner could be calculated where everyone paid the same price for particular services and where the weights of groups in the demand

---

[1]The problems have been chosen because they have been addressed in the orthodox manner by Bailey and Willig in conjunction with other authors. In addition to the article cited above, the references are: J.C. Panzar and R.D. Willig, "Free Entry and the Sustainability of Natural Monopoly", <u>Bell Journal of Economics</u>, 8(Spring 1977); E.E. Bailey and Eric B. Lindberg, "Peak Load Pricing Principles: Past and Present", in Harry M. Trebing (ed.), <u>New Dimensions in Public Utility Pricing</u> (Michigan State University, 1976); R.D. Willig, "The Theory of Network Access Pricing", (Mimeo, 1979).

for each service in conjunction with elasticities played a role in determining the relation of different rates to their incremental costs.

Unfortunately, this modification of the objective function is not suitable for capturing the essence of the vote function, if those aided are not aware of the causal link between their welfare and the pricing policy of the regulator. The ideal policy politically is one where beneficiaries perceive clearly their gains and losers are not aware of the extent of their losses. The importance of making a gain apparent without substantial intellectual effort is enhanced, if the transfer not only garners votes from those directly involved but from others who are generally sympathetic to the plight of that group. Groups with political power that lack these externalities or with negative externalities of this sort will be helped in a less direct manner. The modified Ramsey approach discussed fails to take these considerations into account. It takes service categories as given, presumably from cost considerations where, in fact, the determination of service categories is an important instrument for political discrimination.

One stock answer of why utilities are regulated is that regulation curbs discrimination by private holders of the monopoly rights.[1] In my opinion, the power is not curbed but transferred from private to political hands. Discriminatory initiatives by the regulator take two forms. At one level, broad statements of intent

---

[1]Roland H. Koller II, "Why Regulate Utilities? To Control Price Discrimination", The Journal of Law and Economics, XVI(1), (April 1973).

are made, and a general hierarchy of rates with different social priorities are declared. Broad statements of intent to discriminate are sprinkled through the CRTC's decisions. An example is

> The equalization of message toll rates between points at similar distances in Canada, at least with respect to public voice traffic, is in the Commission's view clearly in the public interest, and settlement arrangements should be adjusted if necessary to make this possible. (Telecom Decision CRTC 78-7, p. 90)

Discrimination occurs here by treating as homogeneous for pricing purposes services that may have different costs. The CRTC has also indicated that various service categories have different social importance. Listed from most deserving to least, these are basic residential, basic business, message toll, and services to large business.

General statements of priorities and of serving broad national objectives are intended for those who do not have the time or inclination to discover what is actually happening (or more accurately to discover how difficult it is to know what is happening). Even if these broad priorities were actual guides to policy making, the effects of their pursuit on any particular group are difficult to establish. It is, in fact, almost impossible to calculate the ultimate incidence of telephone rates since, for example, consumers buy both local and long distance services as well as goods, the price of which are raised to an unknown extent by various business rates.

Rate structure decisions where a direct recipient is clearly identified define a second level of discrimination. Highly visible concessions are made to groups with positive political externalities. In recent cases, these have included special rates for the handicapped, for northern communities, for hospitals, for rural communities, and for the indigent. Other groups such as large businesses with political clout but negative sympathy externalities will receive concessions by more indirect means. If the power of large business is correlated with a more sophisticated knowledge of cause and effect, it can be helped in more complex ways and recognize it; other less sophisticated groups who would resent aid being granted to large business are less likely to perceive the aid when granted in a complex manner. For example, a structural change such as introducing limited competition for private line services and for data services, services which are predominantly used by large business, may substantially reduce the ability to tax these services. No declaration of lower rates being granted is made but the effect is the same.

With some groups benefitting by visible concessions and others by indirect concessions, who is being taxed? As mentioned, no one has been able to identify the ultimate incidence of the telephone rate structure, but it would appear rationale for the regulator to seeks its tax base where deadweight losses are minimized.

To summarize, the appropriate vote function incorporates means as well as ends in its arguments, or, more accurately, incorporates individuals' causal models linking policy and the resulting state of society as well as each individual's welfare.

The regulator can attempt to change those "models" through disseminating information and educating the public. That is often a long run objective, but, at least in the short run, account must be taken of the public's actual causal models, and degree of knowledge.

A corollary of the two issues discussed, the political value of avoiding inconsistencies and the necessity of taking into account the public's model, is that the commission will not publicly be concerned with, nor take positions in, disputes over the proper way to model various aspects of the world in which it operates. It was noted above, as an example of the state of flux in orthodox economics, that there is some controversy about how to best account for uncertainty. To be rational, the Commission must have some explicit or implicit framework for taking into account uncertainty, but it is difficult to find any articulation of what this is. There are Delphic references to a concern with risk in some decisions. For example, when BC Telephone acquired GTE Automatic, the CRTC decided that the appropriate minimum rate of return on this investment was 15 per cent. In establishing this rate of return, the CRTC stated

> that the appropriate rate of return on the investment in Automatic Electric, should not be related to the return on the equity of B.C. Tel, since the real issue is not relative risk, but rather the inherent risk associated with Automatic Electric itself. (Telecom Decision CRTC 79-17)

Although the statement may sound reasonable to an untrained ear, it would appear not to be consistent with orthodox treatments of risk in economics, nor with any proposed alternative. Perhaps, the CRTC holds the nihilistic views of Keynes, but feels it must say something. The approach to risk on this issue when compared with that taken by the CRTC in considering an appropriate return on new services, or with dealing with the Saudi Arabian contract of Bell Canada, appears inconsistent. Inconsistency in actual approach to problems that appear different, but are not, is unlikely to be noticed, and become politically embarrassing. Inconsistency in approach will be noticed where the situations are, at least, superficially similar, and the Commission would be expected to eschew that. For example, the CRTC has announced that it will deal with Bell's investments in Northern in exactly the same manner as BC Tel's in Automatic.[1] The inherent risk in the two situations is deemed to be the same, despite substantial differences in the subsidiaries.

Once the objective function is appropriately specified, it is my judgement that a structure similar to the orthodox economic model and much more sophisticated than the statements of the regulator reveal is adopted by a purposeful regulaor. To be operational, the bits and pieces of such a model or simplified variants must be quantified. Perhaps the widespread view that the regulator is not economically literate arises from the regulator's considerable dependence on sources of gathering and processing information other than the standard statistical techniques that dominate academic research.

[1]Telecom Decision CRTC 80-14, p. 59.

VI. INFORMATION AND THE REGULATOR

Purposeful behaviour requires searching out the possibilities for policy and assessing them. How does the regulator obtain information necessary to assess the political consequences of different policy options? The hearings of the CRTC provide a forum where the different groups argue for and against different price configurations, and changes in the conditions of entry. In terms of page tons a great deal of information is exchanged in a typical hearing. For example, during the 21 day hearing leading to the most recent decision on Bell Canada's rates (12 August, 1980), 583 interrogatories were filed and answered by the applicant, 3960 pages of transcript were recorded, 77 exhibits were filed by the applicant and 57 by intervenors. In the accompanying regional hearing, there were 1242 interventions. Represented at the hearings were broadly based associations such as the Consumer's Association of Canada, the National Anti-Poverty Association, the Inuit Tapirisat of Canada, and the Wa-Wa-Ta Native Communications Socie More narrowly based associations interested as customers in particular rates also attended. Illustrative examples are the Canadian Industrial Communications Assembly, the Canadian Press, and the Ontario Hotel and Motel Association. Counsel, representing the Radio Common Carrier's Corporation and a number of that association's members individually, provided information from a different perspective, since the radio common carriers are both competitors and customers of Bell. Bell's system, data and private line competitor, CNCP Telecommunications, who also is a customer of Bell, was represented, as were a number of parties, such as Rolm Corporation, with different interests in encouraging more liberal terminal interconnection and in monitoring Bell's pricing response in areas where interconnection has occurred. As well, there were representatives of the Ontario and Quebec Government, the Director of the federal Combines branch and representatives from the Canadian Federation of Communications Workers.

A hearing is, in part, a surrogate political market survey for the Commission, permitting it to design a more politically effective set of prices that discriminate between different customer groups. As well, the participating groups become aware of the existence and strength of competitive interests; in this setting, proclamation of a particular view of a just price is unlikely to go uncontested. The present commission has encouraged wide participation by different groups in a number of ways. For those groups that are expected to be able to fincance their own participation, the Commission has acted to penalize a failure to do so. For example, in agreeing to a 10 per cent rise in business basic service, as compared to only a 5 per cent increase on residential basic service, in the August 10, 1978 Bell rate decision, the CRTC drew attention to the lack of opposing "noise" from business:

> ...in spite of the sizeable increases which were proposed in the application, no major business user or association other than CNCP Telecommunications chose to intervene at the central hearing, as distinct from numerous individuals and groups with far fewer resources, who participated on behalf of residential users throughout the hearings. (CRTC 78-8, p. 95)

The same point was made in approving no increase in the rate for competitive terminal equipment where the Commission stated:

> Although the Company's marketing judgment could have been buttressed by more quantitative data on demand, market share and costs, since no evidence was presented to contradict the Company's claim, the Commission is prepared to accept the evidence in this present case. (CRTC 78-8, p. 87)

In addition, where an intervenor has represented the interests of a number of subscribers, been responsible, has improved the understanding of the Commission and does not have sufficient alternative funding, the Commission will award the applicant costs. The costs are paid by the telephone company. A discretionary "subsidy" is being granted financed by a "tax" of which the ultimate incidence is difficult to assess.

In one of the first applications of this policy, Challenge Communications Inc. was awarded costs of $33,983.11 with respect to a hearing for relief under section 321, the unjust discrimination clause of the Railway Act. Challenge had asked for $121,045.16 in costs. Of the $33,983.11 awarded, twenty-two thousand dollars was to cover the fees of counsel.[1]

Another instrument for mobilizing groups to provide information is a requirement that the applicant notify all subscribers about a proposed rate change as well as providing public notice through newspapers. These notices request the public to write to the CRTC with their views of the proposal.

---

[1]In discussing the appropriate rate to be established for a solicitor's service, the Commission's Taxing Officer referred to the case of Re Solicitors (1972) where one of the four criteria listed is "the monetary value of the matters in issue". In this context, the Commission's counsel confirms a belief in the value of service principle in rates, at least in paying lawyers.

Information with respect to the distributional consequences of different actions must be buttressed by an appreciation of what is possible. What are the costs of providing different configurations of telecommunications services? To answer this question, the regulator could specify a cost function with the latest properties, and estimate the relevant coefficients from the information contained in a few columns of numbers, or it could turn to the operators of the franchise and obtain its information directly from them. Since every business has an incentive to invest rationally, and no business to my knowledge invests guided by the demand curve for capital derived from a statistical cost function, there is revealed a superiority of the knowledge embedded in a company's personnel and systems.

What may be an issue in relying on the regulated company is the ability of the regulator to obtain accurate information, since there is an obvious conflict for the company to provide better information which allows it to be squeezed more effectively.

There are a number of considerations to take into account in assessing this danger. First, this problem is not avoided by using the statistical approach, as the data on inputs, outputs, etc., will come from the regulated company. Secondly, almost all tax and subsidy schemes in Canada depend on self-provided information where the same incentive for opportunism exists. Thirdly, this incentive for opportunism is disciplined by the continuing relationship between the regulator and the regulated company. If the truth becomes evident in the fullness of time, the company will be punished for any shading of the truth it may have undertaken. Fourthly, the discipline of ongoing relations is buttressed by tracking anything that is measurable. For instance, in the August 10, 1978, Bell rate decision, the CRTC expressed some concern with the accuracy of Bell's revenue estimates and the basis on which they were made. The CRTC announced it would track actual revenues against proposed on a monthly basis and would reopen the question of the estimates if significant deviations appeared.[1] Tracking of demand, revenues, resource quantities and costs is also an integral part of the evaluation studies which are required for new service filings. Again, significant variations between actual and forecast require a company to explain.[2]

Fifthly, the quality of information is also disciplined by the intervention of informed intervenors. Although regulated duopoly is a market structure that is difficult to analyse theoretically and compare to alternatives in performance terms, it can improve the information flow to the regulator. The recent participation of Bell and CNCP at each other's rate hearings is an example. Customer groups may also question figures provided in hearings. For example, in the recent Bell rate case, the Wa-Wa-Ta disagreed with the company's figures on blockage rates in remote northern Ontario, and the Consumer's Association of Canada expressed concern over Bell's method of measuring utilization.[3]

[1]Telecom Decision CRTC 78-7, p. 78.

[2]Telecommunications Bulletin, Vol. 3, No. 9 (15 September 1979), p. 16.

[3]Telecom Decision CRTC 80-14, p. 18 and p. 35 respectively.

The regulated companies also provide an enormous flow of information by filing economic evaluation studies, by answering interrogatories, and by presenting their case in hearings. The regulator continues to ask for more detailed information on a variety of subjects. Sometimes an incentive is given for the provision of information additional to that which has emerged in a Hearing. For instance, in Telecom Decision 78-5, 50 per cent of CN Telecommunications request for increasing basic service rates in Newfoundland was withheld until much greater detail was provided on the company's five-year upgrading plan. More frequently, the regulated company is simply instructed to provide the information. A sample of studies and information requested by the Commission of Bell Canada in its last two rate hearings are:

(i) Off Peak Pricing:

> ...the Company is directed to explore further the development of rate structures for off-peak usage which incorporate a lower unit charge than obtains under the message toll rates approved in this decision. In this regard the company is directed to set out a schedule for this undertaking by 30 November 1980 for consideration by the Commission. (Telecom Decision CRTC 80-14, p. 120)

(ii) Digital, DDD and Remote Communities

> The Company is therefore requested to propose to the Commission, by 31 December 1980, a program for the installation of digital technology and Direct Distance Dialing in remote communities, including those located in remote northern Ontario. (Telecom Decision CRTC 80-14, p. 17)

(iii) Optional Calling Plan

> Bell recommended that an exchange should qualify for such treatment if a minimum of 50% of customers call another exchange at least once a month. The Company acknowledged that exchange boundaries do not always conform to patterns of economic and

social activity. The Commission encourages Bell to assess the feasibility of qualifying criteria other than those stated above. As an example, these could include situations where a substantial minority (e.g. more than 20%) of an exchange's customers call a contiguous exchange three or more times a month. These criteria should attempt to capture those cases where exchange boundaries are divisive. The Commission directs Bell to report, within six months of this decision, the results of their assessment of such criteria. (Telecom Decision CRTC 80-14, p. 29)

(iv) Econometric Models

The Commission considers it essential for Bell to further develop the application of its econometric models, not only to estimate revenue curtailment, but to evaluate changes in demand caused by changes in economic conditions and the consequences thereof on the construction program. The Company is accordingly requested to file a report on this matter at least one month prior to the next preliminary meeting of the Construction Program Review Committee, tentatively scheduled for November 1980.
(Telecom Decision CRTC 80-14, p. 41)

(v) Separate Northern Rate Structure

The Company should therefore provide within one year, a report on the merits of a separate rate structure for remote northern areas, based on actual calling patterns. The report should also cover remote Northern Ontario, which, while discussed further below, may have similar characteristics to those of Northern Quebec and the Northwest Territories with respect to long distance dialing. (Telecom Decision CRTC 78-7, p. 19)

(vi) Criteria for extending base-rate and locality-rate areas

The Company is therefore directed to develop and file with the Commission within six months, criteria of a more objective and quantifiable nature to be used in the establishment and extension of base-rate and locality-rate areas. Revised tariffs which reflect these criteria should be filed at that time. (Telecom Decision CRTC 78-7, p. 24)

(vii) Two-party rates as budget service

Bell Canada is invited to elaborate on this suggestion with full supporting information within six months. Details should be provided on how two-party service quality might be effectively measured, on any changes to two-party rates that the Company wished to propose in the light of the rates approved for individual line service in this decision, and on the demand implications of such a proposal.
(Telecom Decision CRTC 78-7, p. 27)

(viii) Usage sensitive prices

Before drawing any further conclusions, the Commission would wish to review the Company's studies related to usage sensitive pricing and it therefore directs the company to file full details of all such studies within one month of this Decision.
(Telecom Decision CRTC 78-7, p. 29)

(ix) Construction and quality linkages

While it recognizes that the relationships are complex, the Commission considers it essential that the Company develop a method for determining and quantifying the links between construction expenditures and related service quality indicators.
(Telecom Decision CRTC 78-7, p. 35)

(x) Precision in monitoring construction

The Company is accordingly directed to file with the Commission within six months, a study on methods for achieving such greater precision, and for monitoring the different aspects of the construction program.
(Telecom Decision CRTC 78-7, p. 40)

(xi) Performance indicators

In addition, the Commission considers that it would be to the benefit of both the Company and the hearing process if certain key indicators were employed which would permit greater comparability of performance among telephone companies...Accordingly, the Company is directed to file with the Commission within six months, a proposal for the selection of indicators which would enable such comparison and assessment.
(Telecom Decision CRTC 78-7, pp. 79-80

(xii) Elasticities and curtailment

> The Commission regards these factors as important to the rate-making process and will require more detailed information on these revenue effects in future rate applications.
> (Telecom Decision 78-7, p. 84)

(xiii) Hospital PBX's

> The Commission requires that a study be carried out by the Company, specifically to determine the appropriateness of the SL-1 and Centrex services for serving the basic telephone requirements of hospitals.
> (Telecom Decision CRTC 78-7, p. 93)

(xiv) Small business vs. large business rates

> In addition, the Commission expects the Company to review various approaches to establishing rate structures for contract primary service which recognize the distinction between small and large businesses and to report to the Commission within a year of this Decision.
> (Telecom Decision 78-7, p. 96)

There is more "studying" occurring at Bell Canada than in most universities.

Information is solicited to quantify the options available, but also to monitor the regulated companies. Fifteen quality indices are now calculated, and performance according to those indices is assessed in rate cases. Bad performance must be justified by the company involved. For example the rationale provided by Bell Canada for poor performance in the repair appointment indicator, poor weather, was not accepted by the Commission. The indicators relieve some of the uncertainty that the operating companies have of what is expected of them; the incentives are clear since the Commission has announced that it "will take appropriate action in the future when determining just and reasonable rates"[1] if quality is not up to snuff. On the other hand, quantification will create a bias where the dimensions of quality that are being measured will receive attention and some which have not, or cannot be objectively measured, will be ignored.

The Commission also disciplines the companies by providing information to customer groups of what their rights may be. For example, in the most recent rate hearing for Bell Canada rates, the Commission noted

> ...it appears that not all communities are aware that rebates may be available for service outages. (Telecom Decision CRTC 80-14, p. 23)

A final source of information for the Canadian regulator is experience in the United States. In the 1976-77 Bell Canada rate hearings, the vice chairman of the CRTC when confronted with the problem of cross-subsidy indicated that the FCC approach to policing cross-subsidization might be adopted in Canada as an interim response by the Commission. More recently, the terminal interconnect standards of the FCC have been adopted as an interim measure.[2]

[1]Telecom Decision 80-14, p. 13.

[2]Telecom Decision CRTC 80-13, p. 28.

VII. CONCLUSION

A summary of the structure of the paper was presented in the introduction and need not be repeated. Instead some interpretive comments are presented.

By substituting a vote function for the social welfare function an operative model was created. The results of choice in the modified model probably deviate significantly from that which would emerge from the normative model assuming a social welfare function can be specified. To acknowledge this is not to recommend a campaign to exhort the regulator to behave differently. Such exhortation if effective would be counter-productive under present arrangements, since a change in behaviour would increase the probability that the present reformed regulator would also be the ex-regulator. To argue that the regulator and its principle, the government, do not really know what is politically attractive is to argue that amateurs know more than professionals - an argument that is only taken seriously in England. If regulation is to be changed, either the constraints of the regulator have to be altered by, for example, the public being better informed, or by constitutional constraints being imposed, such as limits on the scope for political discrimination.

A second lesson from the investigation stems from the extent of the informational demands on the industry and on customer groups. If democracy is to be a competitive process, challengers to the incumbent must not be at a severe informational disadvantage. Through regulatory processes, the government collects and processes a vast amount of information about the economy, at the expense of the taxpayer, and has an enormous advantage over any potential challenger who must invest personal resources to duplicate such knowledge. In my opinion, the informational demands of the regulator far exceed those required for the immediate requirements of regulation and instead, represent an investment in information that is politically valuable. The regulator is part of a political research and development complex which gathers and processes information that allow new policy initiatives to be taken, and provide the basis for showing the inadequacy of the competing policies. The persistence of some regulatory schemes despite their apparent redistributional inefficiency may be explained by the political value of the information they generate. A case in point would be price and incomes controls which generate much detailed information about a variety of businesses. The social loss from the information asymmetry is the reduction in effective competition in the political arena.

COMMENTS ON THE PAPERS IN THE REGULATORY PROCESSES AND SOCIO-ECONOMIC ISSUES SESSION

RICHARD LIPSEY

QUEEN'S UNIVERSITY

The papers in this session raise broad issues and thus inevitably touch on such matters of competition, monopoly and scale that were the subject of papers in previous sessions.

One really key word in everything that we have heard is CHANGE. Viewed as a series of snapshots of say one or two per decade the economy changes dramatically and continually. There is neither permanence nor natural monopoly under this time span of observation.

Viewed under the shorter perspective of, say, one-half of a human life with observation starting at, say, age 20 we tend to see a rather static picture. Great stable giants -- bulwarks of the economy or grasping monopolies according to your perspective -- large, important and stable. Then as time passes and change proceeds, the "stable" giants sooner or later encounter trouble. Our reaction is all too often to shore them up and to try to return to the good days of stability -- which are merely an optical illusion due to the shortness of human life. The U.K. experience of trying to stabilize an industrial structure in the face of continually changing tastes and technology is an experience that it would be unwise to copy (although Canada is in some danger of trying to do so).

The short-sighted human view leads us to study large industries that are candidates for regulation using static models that freeze tastes (demand) and technology (supply). Keith Atcheson has discussed the use of some such models. He criticizes their use of social welfare functions and proposes to substitute some other objective function such as political payoffs or vote-maximization. If we wish to predict the regulators behaviour then the type of substitution suggested by Atcheson makes a lot of sense. When we are interested in evaluating regulatory policy, however, Atcheson's substitution makes little sense. To evaluate policy we need some equivalent of the social welfare function. Here I must confess to not taking the force of Atcheson's criticisms. Although as one of the authors of the "General Theory of Second-Best" I should be as aware as anyone of the limitations, I see no alternative than to using some form of the most sophisticated versions of consumer surplus as our evaluative criterion.[1]

Whatever evaluative criteria we use, these static models give us the concepts of natural monopoly and are the basis of much of the

---

[1] Atcheson also criticizes current econometrics. The achievements of econometrics have certainly fallen below the high expectations held for them fifty years ago. But then most expectations of a new technique are higher than what is ultimately delivered. Econometrics has not had great success in discriminating among such general models as monetarism, neo-Keynesianism and post-Keynesianism. This is partly because such "models" really constitute whole research programmes in Lakatos' sense of the term, which are not subject to direct testing. The inability to discriminate among more precise macroeconomic models is at least partly due to methodological failures (which I have briefly discussed in R.G. Lipsey, "World Inflation", The Economic Record, December 1979). But there can be little question that econometrics has served to narrow the range of possibilities in many cases. Understanding the agricultural problem in terms of price and income elasticities was an early triumph of econometrics. General acceptance of a stable demand curve for money with a significant interest elasticity was a more recent one.

case for regulation. Schumpeter has argued, however, that over the long run there are no natural monopolies. The undeniable facts are that yesterday's industrial giants are often today's declining firms; while yesterday's small competitive firms are often today's industrial giants, oligopolies or monopolies. What Schumpeter suggested was that a policy model that froze the economy at a moment of time and then asked how we could tinker with it to improve its performance was irrelevant to the major forces governing economic welfare. Indeed, he suggested that the use of such a model might actually be harmful because, for example, when profits were regulated this could reduce the incentive for change by reducing the prize of monopoly profits waiting to be usurped. More importantly, by declaring an industry to be a natural monopoly in year X and regulating it, the industry may have a policy-maintained monopoly in year X plus 10 when circumstances have changed and technology would allow new firms to enter and challenge the existing monopoly.

Someone from the floor this morning asked whether technological change inevitably encourages competition. The answer, of course, is "no". Technology has sometimes created the circumstances that give rise to monopoly. Schumpeter, however, turns this question on its head and asserts that monopoly encourages technological innovation as new firms seek to gain a share of the monopoly's profits through the process of creative desctruction. Today we are seeing an example of Schumpeter's creative destruction in the communications industry. As Jurgen Mueller's paper shows, the scope for competition in this industry is large. The idea of telecommunications as a natural communications monopoly will undoubtedly go the same way as the idea of the railway as a natural transport monopoly.

Of course there are examples in the literature of models that take dynamic perspectives, but much of our stock of models, and of our regulation policy, is still based on static concepts. What the theory and practice of regulation needs is a framework that is dynamic and focuses on the life cycle of industries and products. Taking the automobile industry as a typical example we see a series of stages. First, with the introduction of a new product there are many new firms each trying different experiments. This is followed by a period of consolidation when the product catches on and it becomes clear which major lines will be successful; scale economies then dictate the concentration of the industry in a smaller number of firms. This is followed by a period of the mature industry dominated by a few large firms; they are stable, successful and major contributers to investment, employment and output in the economy. Then comes a time of troubles when the industry is challenged with new innovations, new products, new ideas; the old established giants either succumb to the challenge or change dramatically in the face of it.

We need pure thought among theorists on how to model such a process tightly. We also need theoretical work on regulation in such dynamic settings. We need to ask such questions as: Is regulation advantageous when we consider the entire life cycle of such industries and products and not just their situation in the time when they are mature, stable industrial giants? In particular we must wonder about the role of the regulatory process in the period of ferment when new products are arising to challenge the old. If regulation is considered desireable, we need to study how it can be made flexible enough to remain appropriate to the changing circumstances of the industry over the later stages of its life cycle.

Dealing with such questions would undoubtedly lead to a rethinking of many of our fundamental ideas about objectives and methods of the regulatory process. We also need applied studies of industries in such transitional situations as the telecommunications industry at the moment. Here U.S. and Canadian comparisons and contrasts promise an excellent approximation to a controlled laboratory experiment. The two countries are faced with a common technology but a rather different regulatory climate: a movement toward decontrol in the U.S., and a movement toward more control and greater politicization in Canada as Loria Salter explains in her paper. Will, as a Schumpeterian would suspect, Canadian policy reduce the pace of change by partly inhibiting the growth of competition and hence leading to higher prices compared with the U.S.; or will the different national policies have little effort on the paths of economic change, price and profits in the two countries? This is a very important question because if the answer is that the effects are small, we can feel more sanguin about the use of regulatory bodies for social purposes as described by Ms. Salter. If the Schumpeterians are right, however, then we must view current Canadian developments with some concern, possibly even alarm.

Although there is nothing wrong with advocating long-run structural changes in economic policy, we must accept the world in which we live when giving current advise. In Canada today we have a federal government that in European terms would be described as Social Democratic. Also, we have concerns over regional inequalities and national unities that will undoubtedly lead to substantial regulation of the economy into the foreseeable future. So we must live with this and ask how we might improve the regulatory structure whose existence is inevitable. It seemed to me from the papers and comments today that four basic approaches could be distilled.

(1) Rules and regulation change slowly and often become outdated as technology advances rapidly. We need to design our regulatory bodies with built-in flexibility that would allow them to change. It is a major research task to discover what techniques would produce the necessary flexibility. Sunset clauses are a possibility not just for the regulatory body as a whole but for each particular regulation. The fact that others at the conference replied to this suggestion that sunset clauses had been tried in Canada and found ineffective illustrates my main point: the need for much theoretical research, and empirical observation on past behaviour, to discover how such regulations can be made as flexible as possible in the light of changing circumstances.

(2) We must try as much as possible to resist the temptation to meddle with individual markets, firms and industries for distributive goals. I still accept the basic social democratic ideals of trying to build a society that is more just and less harsh than what the pure unaided market would produce. I believe it is possible to cushion some of the worst disasters that the free market inevitably produces

for some individuals. I must admit, however, the total failure of the programme to affect the size distribution of income by intervening in particular markets by measures such as rent control, agricultural marketing boards and energy price controls. Whatever haphazard redistributions of income these measures produce, they are not systematic measures for redistributing national income. (I have discussed this matter in more detail in my Nobel Conference Lecture, "An Economist Looks at the Future of the Price System".)

Here I find myself somewhat out of sympathy with the Lesser and Osberg paper. Theirs is a very moderate and reasonable paper, but the basic philosophy seems to me to be dangerous. Judging <u>every single</u> market development in terms of its impact on regional income differentials and attempting to influence it or to compensate for it by subsidies etc. appears to me dangerous and much in the spirit of the belief that intervention in individual markets by such techniques as rent controls are an efficient means of changing the size distribution of income. Of course we do intervene in particular markets to affect the regional distribution of income by, for example, making communications facilities available to the north much more cheaply than the free market ever could. Such policies for holding the nation together are understandable. What I find worrying in the Lesser/Osberg paper is its reflection of the attitude of looking at every single market development and assessing each one by what it does to particular, below-average-income areas and to argue for compensation, subsidies etc. whenever the impact is unfavourable. Such a defensive attitude is not one that is likely to produce good long-term results across the whole country.

Their paper does suggest, however, a more interesting positive problem. In some ways developments in telecommunications may abolish old spatial differences. For example, it may become no more costly to communicate across the nation as across the town. This may have major effects on the patterns of regional comparative advantage and regional income distribution.

(3) We must learn how to reduce the chances of regulatory bodies pointing the wrong way and how to mitigate the effects when they do. Alan Bauchcan explained this morning that these are serious dangers. In market economies decentralized decision makers simultaneously explore many alternative directions. One individual will almost inevitably make a mistake sooner or later. When he does so in a decentralized economy he suffers temporarily or even permanently, and as a result has less control over the nation's resources, while those who have made successful experiments earn profits and gain a larger share of control over the direction of the nation's resources. In a centralized or regulated system when the centralized decision maker makes a mistake a major part of a nation's efforts may be directed up a blind alley. I only wish that any of the speakers today, or myself, could suggest a mechanism to reduce the potential for loss when such errors occur under a centralized system.

(4) Finally and most importantly, we must learn how to use regulatory bodies to fulfill both economic and social objectives more efficiently than they now do. Ms. Salter has pointed out the rising importance of social objectives in Canadian regulatory mechanisms. Of course the regulatory process has always tended to confuse economic goals of efficiency with social goals including income redistribution. Ms. Salter makes a strong case, however, that social aspects are becoming

much more important in the regulatory process than they were even a few years ago. This increasing confusion of economic and social objectives seems to me to be most undesireable. It mixes the problem of attempting to achieve economic efficiency (which is after all the original idea of regulatory bodies designed to control natural monopolies and to make them act in the economic interests of the consumer) with the perceived need sometimes to violate economic efficiency in order to achieve social gains. Both economic and social objectives are clearly important. But they are conceptually quite distinct.and to mix them in the same regulatory body invites confusion. By showing the increasing significance given to social problems Ms. Salter's paper suggests the need for some thinking on how these two decisions might be separated.

One possibility would of course be to kill the entire economic regulatory aspect as is the American mood at the moment. This would leave our regulatory bodies free to concentrate on the social aspects. This is probably Utopian in Canada today. So urgent work is needed to theorize on regulatory bodies in order to discover how they can become more effective in dealing with the twin goals of economic efficiency and social objectives with which they are charged. Towards the end of their papers both Len Waverman and Ms. Salter give constructive ideas that are food for further thought in this direction. As well as theoretical thought we clearly need constant monitoring of what our regulatory bodies are doing in order that we can become aware of where in practise they appear to succeed and where the major failings develop.

AGENDA FOR THE FUTURE

THE TELECOMMUNICATIONS CHALLENGE IN CANADA: SOME POLICY ISSUES ON THE CURRENT AGENDA

JEAN T. FOURNIER

You have just come through three days of in-depth discussions about the implications of current trends in the telecommunications industry. You have heard many approaches to resolving some of the crucial policy questions which must be addressed quickly if Canada is to take full advantage of the benefits of new technology and maintain its reputation for excellence in telecommunications services. You have heard descriptions of the latest econometric tools for analyzing the industry and for measuring the impact of changes in telecommunications policy.

I believe we all benefit from this type of symposium, and I trust this exchange of ideas will promote a better understanding of each other and the difficult issues which we all face. I commend l'Ecole des Hautes Etudes Commerciales and the University of Victoria on the program which was organized in conjunction with the Department of Communications.

The knowledge gained from this conference will help those of us in government to draft the kinds of policies that will best meet the needs of the industry and public in the 80's. I hope these proceedings will be of equal benefit to industry representatives and guide you in the planning and vital investment decisions required in this dynamic sector of our economy.

I will not try in these few minutes to go into as much detail as your other participants have about any single aspect of the telecommunications industry. Instead, I would like to try to fit some of the pieces together and draw attention to a number of the major public policy questions currently facing my colleagues and me in the Department of Communications. As you know, the department regularly seeks public input in formulating policy. I would like to invite your help in finding answers to the questions before us.

New technology and attitudes toward competition have raised several areas of immediate concern to us. For example, how far should we go in encouraging interconnection between competing networks? Can we license new microwave systems to meet the demand for communication services in densely populated areas without hampering existing and planned services in less-developed communities? Will a less restrictive attitude towards terminal attachment help to encourage a thriving domestic electronics industry, or will foreign manufacturers dominate the market, leaving our carriers without the needed funds to maintain current levels of service, and at a price Canadians can afford? Should we allow satellites to be used for transborder telecommunications with the United States? Who should be allowed to operate satellite earth stations in Canada?

Inextricably linked with these questions is the topic that is foremost in our minds today: Who will pay for new communications services and equipment? Will increased competition drive costs down for all users of the system, or will it benefit only a few? What will be the effect of such changes on equipment costs and local and long distance rates? Will Usage Sensitive Pricing prove to be a workable solution to our dilemma over tariff structures? Is the current regulatory process for setting long distance rates adequate for the needs of the 1980's?

I propose to deal with some of these questions, starting with network interconnection.

NETWORK INTERCONNECTION

The federal government has strongly supported the principle of healthy competition between telecommunications carriers since 1953 when CNCP and the TransCanada Telephone System were first permitted to set up competing microwave relay systems. In 1962, this competition was intensified when CNCP was authorized to establish a nationwide microwave system in competition with TCTS. This relationship had, until recently, proved to be a happy compromise between a single, regulated monopoly and unlimited competition. It led to the development of two first-class nation-wide telecommunications networks which were built with a large proportion of Canadian hardware and technology.

During the past decade, however, a new dimension has developed in the regulatory fabric with the introduction in the United States of network interconnection between competing carriers. It became increasingly obvious to us in Canada that the growing demand for specialized business telecommunications services would require a change in the relationship between our own carriers.

These pressures for change were brought to a head on May 17, 1979 when, after lengthy hearings, the Canadian Radio-television and Telecommunications Commission announced its decision to approve an application by CNCP to interconnect its network facilities with those of Bell Canada so that CNCP could use Bell lines to offer specialized, dedicated voice and data services to its customers in Ontario and Quebec. The CRTC ruling was opposed by members of the TCTS and some provinces who petitioned the Governor in Council, arguing that interconnection would lead to a significant loss of revenue for all TCTS members and deterioration in regular service to Bell customers. After giving careful consideration to all aspects of this matter, however, the federal government upheld the CRTC decision. The Honourable David MacDonald, Minister of Communications at the time, said the government had concluded that "opening up the telephone system to greater competition ... would create significant benefits for the economy in general". The government supported the view that interconnection would lead to more innovation in Canadian industry and a wider choice of services for system users.

The introduction of network interconnection should help to restore balance in a market where the telephone companies had an inherent advantage because of their control over the local distribution system. In an era when new uses for computer communications are being discovered every day and the demand for existing services threatens to outstrip supply, it is vital that we encourage innovative ideas and new investment to meet the needs of business and industry.

In its recently released report, the Alberta Public Utilities Board has also favoured interconnection agreements to enhance competition in the delivery of non-basic services. The Board's report, compiled after two years of study and public hearings, suggests a liberalized approach should be adopted in establishing fair and reasonable terms and conditions for the use of monopoly carriers' networks by competing services.

CNCP can be expected to press for interconnection with other telephone companies and has already requested such privileges from the other federally-regulated member of TCTS, B.C. Tel. The remaining TCTS companies are under provincial jurisdiction and we must wonder how provincial regulators will deal with interconnection applications and the thorny issues which they will raise. In any event, the national interest would be well served, as would the interest of business and consumers, if interconnection privileges were extended across the country.

TERMINAL ATTACHMENT

Like network interconnection, terminal attachment is a relatively new phenomenon which is raising questions for policy-makers in many countries. The Department of Communications has long supported the idea of orderly liberalization of the existing rules which restrict use of a competitor's equipment on carrier networks. Our smaller market in Canada has not allowed us to throw the field open to competition as our counterparts in the U.S. have done, but we have been moving in a controlled fashion towards more competition in this area. Should we move to full scale deregulation in this field, or should we be content just with less regulation? Can we develop more flexible regulations which allow industry to respond quickly to changing market conditions? How do we enforce the regulations to prevent unacceptable harm to the integrity of the networks?

The Department of Communications took a lead role in developing standards and certification procedures in this field in 1974 when the Terminal Attachment Program was established on a voluntary, co-operative basis involving federally-regulated carriers, some provincial agencies, equipment manufacturers and system users.

Under the Terminal Attachment Program, DOC certified telephone answering machines and similar devices. Later, we turned our attention to data handling devices. Our first set of standards for network addressing devices, specifically for single line telephones, was published in the Canada Gazette on January 31 this year and we expect the next set of standards for push button systems and PBX's to be issued this spring. The government believes in Canadian standards, on a national basis, to ensure the integrity of our telecommunications systems from coast to coast and to give our manufacturers the national marketing base which they must have if they are to thrive in the world market.

Understandably, terminal attachment has been a matter of concern to the carriers for a number of years. The issue was brought before the CRTC in November 1979 by Bell Canada in an application which asked the Commission to determine whether liberalized terminal attachment would be in the public interest. Last August, the CRTC issued an interim decision outlining provisional rules for attachment of all types of customer-provided terminals to the Bell network. The Governor in Council has received a number of petitions and submissions on the decision, and these are currently being reviewed. A public hearing is expected next fall to consider the full impact of liberalization and its implications for public interest issues. In the meantime, the other major federally-regulated telephone company, B.C. Tel, has requested CRTC authorization of terminal attachment rules identical to those specified in the Bell decision.

More recently, Alberta Government Telephones filed tariffs on February 16 this year to allow the attachment of subscriber-owned terminals. AGT plans to issue connection standards modelled on those developed by the federal Terminal Attachment Program.

Finally, the Restrictive Trade Practices Commission has launched a special inquiry into the question of terminal attachment and its findings should be available in the near future.

CANADA-U.S. TRANSBORDER SATELLITE COMMUNICATIONS

Since the day that Alexander Graham Bell first invented the telephone (it happened in Canada, or the United States, depending on which phone company you believe), our two countries have enjoyed a uniquely harmonious relationship in many areas of telecommunications policy. We have, in effect, treated each other simply as extensions of the other's domestic networks and Canadian transborder telecommunications have been provided by member carriers of TCTS and by CNCP. Through interconnection agreements with U.S. carriers, revenue sharing and routing are being carried out via a small number of border crossing points.

Today, there are many suggestions that this relationship should be extended to include telecommunications via satellite. Neither Canada nor the U.S. has used domestic satellites for Canada-U.S. transborder traffic to date except on an experimental basis, but the possibility has been discussed informally between governments and industries in our two countries.

We recognize that if such an agreement were to be made, it would be of the utmost importance that Canadian manufacturers and telecommunications carriers be given a chance to share in the industrial and carriage revenues that would result. We must be prepared to move quickly to carve out our share of this exciting new industry or we will be overwhelmed by the burgeoning satellite-telecommunications giants south of the border.

Careful consideration is being given to the impact such a move would have on our domestic industry and business users. Canada-U.S. traffic is a significant source of revenue for Canadian telecommunications carriers and they are, quite naturally, concerned about the possible erosion of their revenue base and the effect this would have on the quality of regular services. Our figures show that in 1978, for example, Canada-U.S. traffic represented 30.2 percent of TCTS-originated revenues, while purely domestic traffic accounted for 51.3 percent and Canada-overseas traffic represented 18.6 percent.

On the other hand, satellite technology is particularly attractive for private business circuits and some multinational companies have expressed a desire to interconnect their facilities via satellite. We are also witnessing a steady increase in demand for satellite delivery of broadcast industry traffic, both for regular programming and for temporary hookups for special events. In addition, we foresee a demand for satellite transmission of teletext traffic. Later this year, for example, a full transponder on a U.S. satellite will be devoted to a multi-channel teletext service operated by Time Incorporated of New York, using the Telidon system developed by Department of Communications researchers.

In considering our policy options, we are trying to strike a balance which will reflect the national interest. Can we minimize the risk to our carriers and maximize the opportunities for all Canadians to profit from satellite carriage revenues and the supply of related technology such as sophisticated ground equipment and ensure that all users have full access to the most up-to-date services?

Can we find a revenue-sharing formula that will protect existing carriers from economic hardship and ensure that our industry has an equitable share in profits? Who should represent Canadian interests in negotiations with U.S. carriers? These are just a few of the questions which must be answered before such a service is launched.

EARTH STATION LICENSING

One of the more contentious questions relating to satellite policy today is the issue of earth station licensing, which has figured so prominently in the press in recent months.

This is an area which has been gradually opened to competition in recent years, engendering a new range of social and economic policy issues.

Until 1979, only Telesat Canada was permitted to own and operate earth stations. The government moved in February of that year to allow broadcasters and telecommunications carriers to apply for radio licences for such installations, and in November of 1979, the Minister of Cómmunications, the Honourable Francis Fox, announced that the field was widened to include provincial educational agencies.

Today we are studying new ways to meet the urgent need for a wider range of radio and television programming for remote and rural communities. Last November 24, the Minister of Communications announced a simplified application process for those wishing to own and operate satellite television receive-only (TVRO) earth stations for the reception of signals from Canadian satellites. At the same time, the Minister launched a review of policy and procedures for applications for earth station ownership and we are currently studying the possibility of extending earth station ownership privileges to new categories, including private individuals, remote small communities and business undertakings, resource exploration teams and other groups.

The Minister of Communications has made it clear that the government is not prepared at this time to license earth stations for the direct reception of TV signals from foreign satellites. Such action is prohibited by international agreements which bind both Canada and the U.S. The Minister has also expressed concern that unrestricted reception and distribution in Canada of U.S. satellite signals would pose an unacceptable threat to Canada's broadcasting system and the thousands of Canadians employed in that industry.

MICROWAVE LICENSING POLICY REVIEW

Throughout Canada, new pressures are being placed on our domestic telecommunications networks by the steadily increasing need for local and national microwave relay services, particularly in the delivery of broadcast signals for the television and cable industries. Is our current microwave policy, which was established by former Communications Minister Eric Kierans in 1969, still adequate to meet the needs of the 1980's?

On December 3, 1980 Mr. Fox announced that the Department would review certain aspects of the federal government's microwave system licensing policy in view of increasing requirements for the intercity delivery of TV program signals.

A number of major areas have been identified by the Department and interested parties have been asked to comment on these and other issues. Our preliminary review has raised several areas of concern. For example, it has been noted that current microwave licensing and spectrum allocation policies only allow intercity trunking of a limited number of video channels. We must also ask what impact the licensing of new microwave networks would have on existing services. Would new networks in high density, contiguous markets inhibit the extension of new programming and service to residents of remote areas that would be better served by satellite? What licensing criteria are most appropriate in defining the public interest aspects of microwave applications? How can we make the best use of spectrum allocation? These are the questions that can only be solved with the help of industry and public interest groups.

LOCAL AND LONG DISTANCE TELEPHONE PRICING

All of these complex issues which I have outlined will have an inevitable impact on the one subject which affects all telephone users in Canada - tariffs and rate structures.

As you are well aware, Canadian tariff structures evolved in a monopoly environment and were strongly influenced by federal and provincial requirements that carriers meet the social equity goal of universal access. This has meant that rates for basic local residential service have been set as low as possible and the service has been cross-subsidized with revenues from other telecommunications services.[1]

Traditionally, residential rates have been kept to a minimum, local business charges have been set much higher and the price of long distance message toll service has been set considerably higher than the cost of providing the services. (We have had estimates from one carrier that it spends $1.30 for every dollar of revenue to provide local residential service, while message toll service costs the carrier only 30 cents for every dollar of revenue generated.)

With increased competition in the lucrative terminal equipment and long distance markets, there are growing pressures to change the degree of cross-subsidization.

Regulators at the federal and provincial levels must be prepared to analyze carefully the impact of any such change so that tariffs can be developed which are economically efficient but do not jeopardize the essential goal of universal access or undermine the viability or quality of the basic switched telephone network.

In this regard, it is interesting to compare Canadian rate structures, which have developed in a protected market, to those in the U.S. where increased competition has been permitted in the past decade. While direct comparisons are complicated by the Canadian preference for Extended Area Service (which results in higher basic charges but fewer toll calls), it is fair to conclude that for similar user groups, basic residential service is cheaper in Canada while basic local business charges are higher. We also find that Canadian charges for long distance message toll services have gone down over the years in relation to the U.S. but still remain significantly higher than in the U.S., particularly for long-haul service.(2) This situation, of course, adds to the cost of doing business in Canada and undermines the competitiveness of Canadian industry. In the U.S., specialized

---

(2) See Appendix 2

carriers such as MCI, which competes with AT&T in long distance services, usually offer significantly lower rates than those of AT&T for all except low-volume users. For instance, a New York Times survey in July 1980 showed that for similar users groups, MCI's charge for a two-minute call from New York to San Francisco was 30 cents while the Bell system charged 63 cents for the same call.

What can we learn from our Southern neighbours about future tariff policies in Canada?

We know that the current structure of local and long distance tariffs in Canada has been based on a flat rate charge at the local level, with the cost of equipment, network access and the cost of the call bundled together.

I think it is fair to predict that increased competition will produce more pressure for cost-of-service pricing and unbundling of local service charges into their components. Are Canadians ready for this type of billing? What benefits will they be offered in exchange? If, as seems likely, local service rates start to move towards cost-of-service pricing, how do we ensure that we do not stray from the goal of universal access? Can we find a new way to cross-subsidize these services to maintain a high quality of service at a cost that is fair to all users of the system?

Canada and the U.S. are almost alone in the world in maintaining the bundled charge for local telephone service. Usage Sensitive Pricing (USP), in which the user is charged for the number and/or duration of outgoing calls, is the accepted norm in many countries. In Canada, it is offered on a limited basis only to business users in some communities and is primarily aimed at small firms that make a few hundred calls each month.

In 1979 Bell Canada applied to the CRTC for permission to conduct an experiment in Usage Sensitive Pricing, but the application was withdrawn, partly as a result of negative reaction from consumer groups. I do not think the issue will sit on the shelf for long, however.

Increasing demands are being placed on the phone system, particularly for data transmission and other services that were unheard of two decades ago. I think we can look forward to even greater demand within the next few years as the requirement for data service extends from the world of business to the consumer's home with the introduction of interactive videotex services.

As policy makers, we recognize that USP could be an economically efficient option for pricing services which generate such high traffic volumes. Could it be possible to create a tariff structure in which all local business services and all data services, whether business or residential, are subject to USP, while residential voice circuits are billed at a flat rate? How could such a rate structure be administered in an environment of unrestricted terminal interconnection?

These are questions which challenge all of us in the field of communications, and they are made more complex because they cannot be answered by any single agency. Unlike the U.S., which has central control over all long distance interstate tolls through the Federal Communications Commission, Canada has a number of regulatory bodies at the federal and provincial levels, and there is little or no cooperation among them in setting long distance rates, although one should not overlook the CRTC's efforts in establishing an inter-regulatory Committee in the late 70's. As a result, even the "trans-Canada" rates of the TCTS may vary from province to province. We have quirks in our rate structure that are as baffling to the experts as they are to ordinary citizens.

Why, for instance, should a cheap-time call from Ottawa to Winnipeg cost as much as a similar call from Ottawa to Vancouver, almost twice the distance? And why is the price of an east-bound call sometimes different than the price of a west-bound call?

If we are to deal effectively with these and other "national" regulatory issues, do we not need to establish a Joint Board with federal and provincial involvement, as proposed by the federal government during last summer's marathon constitutional negotiations? Or is the indirect influence which the CRTC currently exerts over interprovincial rates, by virtue of regulating Bell and B.C. Tel, sufficient to safeguard the national interest? What happens if jurisdiction over these two carriers is transferred to the provinces?(3)

The current regulatory patchwork leaves much to be desired. Yet it has served us very well over the years, giving us one of the best communications systems in the world. Why then the obsession to tinker with something which works reasonably well?

Questions such as these must be addressed if our nation is to keep its leading edge in the field of communications. For the sake of our carriers, our manufacturers, our businessmen, and above all our system users, they must be answered. They cannot be answered, however, by dismantling the current national system of communications and replacing it by 10 loosely-knit provincial communications system "with or without federal participation", as proposed by the provinces during the latest round of constitutional negotiations. Those proposals would have extended provincial concurrency to include jurisdiction over Teleglobe, Telesat and spectrum management, assuredly paralyzing our communications industry, causing havoc to the regulatory process, opening up new areas of intergovernmental conflict and generally undermining Canadian leadership in communications, at home and abroad.

---

(3) For an elaboration of the Joint Board concept and a brief analysis of federal and provincial constitutional positions, see my [illegible] 18th Annual Conference of the Canadian Industrial

One feels dismayed and saddened at times by the lack of understanding shown by some interested parties in the benefits of maintaining a truly national system of telecommunications. Is it that we are taking communications too much for granted and have become complacent? If as a country we are to meet both the technological and content challenges facing us, it follows that the national dimension of communications in the 80's has to be affirmed and strengthened, while giving full effect to regional interests wherever it is practicable to do so.

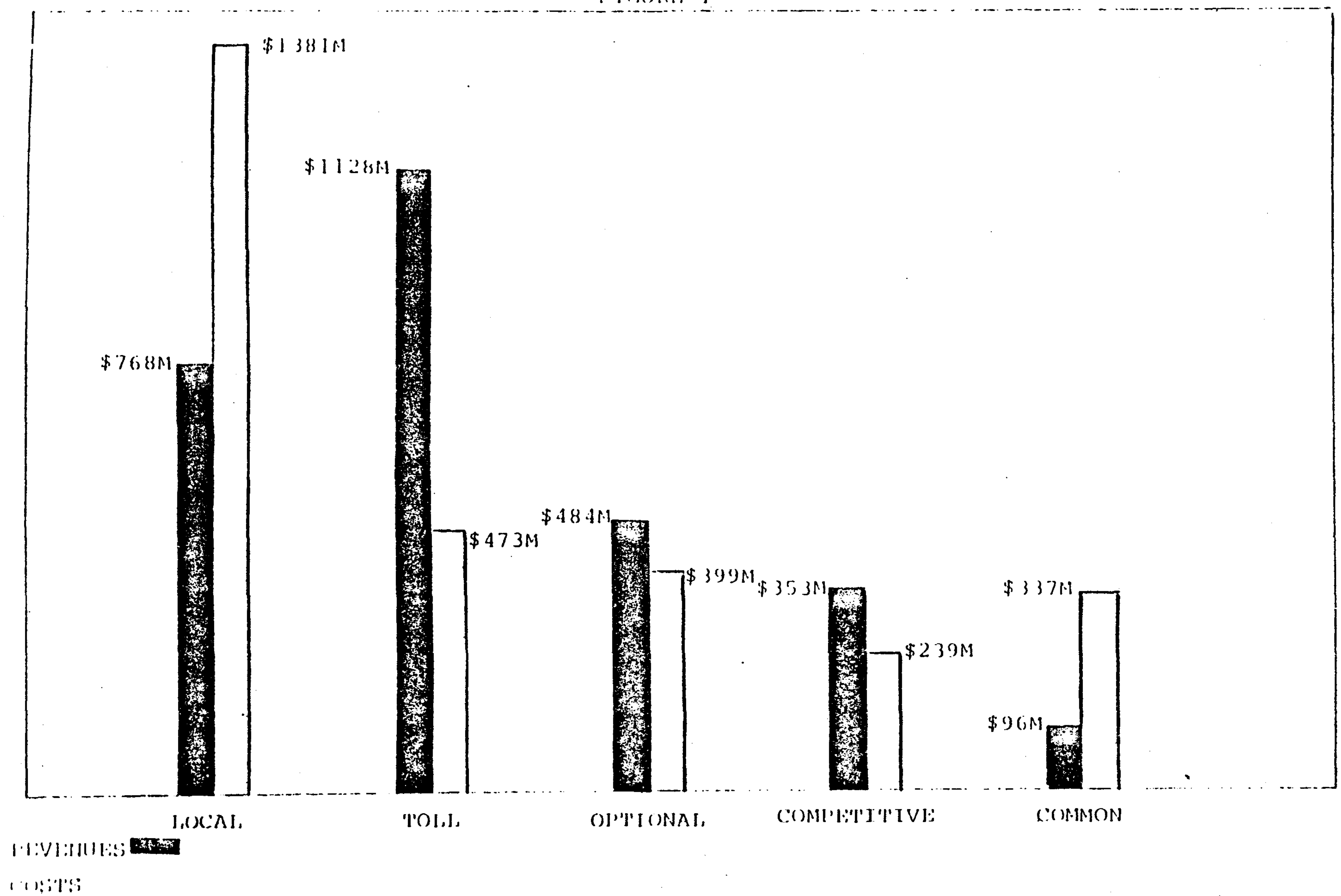
The stakes are high in communications. The issues are many and complex, and decisivness is essential. We must approach the task rationally, with open discussion between government and industry and with the help of the academic community. We must approach the task from a national perspective, with the goal of international excellence in mind. We need an open dialogue at all levels, and we must have your participation if we are to succeed.

Local, Toll, Optional and Competitive

Service Categories

YEAR: 1979

FIGURE 1



COMPARISON OF LONG DISTANCE TELEPHONE RATES IN CANADA AND THE UNITED STATES

Table 1

Selected Long-Distance Telephone Rates
(day, 3 min., station to station, DDD, 1979)

Approx. Distance (miles)	Canada	United States
	Can. $	U.S.$
25	.73 - .78	.57 - .74
75	1.09 - 1.32	.80 - 1.04
200	1.39 - 1.92	.94 - 1.47
400	1.59	1.47
700	2.28	1.18
1000	2.70	1.18

Source: Various Telephone Directories

Table 2

Maximum Long Distance Rates
(day, 3 min., station to station)

	1959	1960	1963	1966	1970	1976	1980
Canada	$5.00	3.95	3.35	3.00	3.00	3.70	3.70
Canada (DDD)	-	-	-	-	-	3.15	2.97
United States	2.50	2.25	-	2.00	1.70	2.25	2.25
United States (DDD)	-	-	-	-	1.35	1.30	1.30

Source: Instant World, TCTS, AT&T

Discussion: TABLE 1

From an examination of the table, there are three points which appear to be significant.

First, and most obvious, the Canadian rates are generally higher than the U.S. rates at all distances. Where there is a range, the lower end of the U.S. range is always lower than the corresponding Canadian rate. On the longer distances (1000 miles or over), the Canadian rate is more than 128% higher than the U.S. rate.

Second, the U.S. rate reaches a plateau at about 700 miles ($1.18). This happens in Canada as well, but only after 1000 miles.

Third, the distance range over which the Canadian rates are closest to the U.S. rates, and in some cases lower, is up to 200 miles. This distance range would, in most cases, constitute intra-provincial traffic. This adds weight to the federal government's concern that inter provincial rates might rise disproportionately in the absence of federal involvement, resulting in burdensome costs for Canadian communications users throughout the country.

Discussion: TABLE 2

Table 2, which attempts to present the maximum LD rates in the two countries, presents somewhat of a problem, because as shown in Table 1, the U.S. rate for long-haul calls is actually less than for some medium-haul calls. Manhattan-Syracuse, for example is $1.47 for 3 minutes, while Manhatten-San Francisco is $1.30. The New York intra-state rates seem to be an aberration, and accordingly we have used the Manhattan-San Francisco rate as the maximum in Table 2, as that rate was felt to be more directly comparable with the maximum Canadian rate (Vancouver-Halifax).

Table 2 shows that the maximum LD rates for operator-handled calls declined both in Canada and in the U.S. between 1959 and 1970, and have increased in the past 10 years. The U.S. rate has been consistently lower than the Canadian rate, although the differential has decreased (in 1959 the Canadian rate was 100% higher than the U.S. rate, in 1980 it was 64% higher).

The direct-dialed call has, of course, replaced the operator-handled call as the most important type of long distance call. The Table shows that in 1980 the maximum Canadian DDD rate is 128% higher than, or more than 2½ times, the maximum U.S. rate. Even allowing for the difference in the exchange rate, in real terms the Canadian rate is more than twice the American.

Conclusions

It is difficult to draw definitive conclusions from a brief analysis of rates between two different telecommunications systems. One thing is absolutely clear. American LD rates have been and continue to be lower than corresponding Canadian rates. Perhaps more significantly, in the important area of DDD calling, the percentage rate differential is greater than it was in operator-handled calls twenty years ago.

One could speculate without too much risk that a major reason for the U.S. rates being lower is the degree of competition which exists in the marketplace. The difference in geography is undoubtedly also a factor. The U.S. companies do not have to serve large, virtually uninhabited (and very costly) areas.

It is interesting to note also that, at least within New York State, the cost of a call within the State is often higher than an inter-state call. This may reflect in part at least, the high cost of re-vitalizing the telephone system in New York. Also, the FCC regulates inter-state rates at the federal level and may in fact be exerting a downward pressure on them. Here again, though, one would suspect that competition plays the major role, for it is on the medium to longer hauls that the specialized carriers offer customers the greatest savings.

AGENDA FOR THE FUTURE

JOHN LAWRENCE

ONE OF THE DIFFICULTIES ABOUT PLANNING FOR THE FUTURE - OR EVEN TALKING ABOUT THE FUTURE - IN THE TELECOMMUNICATIONS INDUSTRY, IS THAT HARDLY BEFORE ONE'S ATTENTION CAN BE TURNED TO IT, THAT "FUTURE" HAS BECOME "THE PRESENT" OR EVEN "THE PAST". THIS IS NOT, OF COURSE, A PHENOMENON FACED ONLY BY REGULATORS, ACADEMICS AND FUTURISTS - IT IS ALSO FACED, PERHAPS MOST IMMEDIATELY OF ALL, BY COMPANIES ACTIVE IN THE INDUSTRY. SIEMENS CORPORATION IN WEST GERMANY, FOR EXAMPLE, RECENTLY WROTE DOWN SOME 230 MILLION DOLLARS OF DEVELOPMENT COSTS FOR A SWITCH THAT WOULD HAVE BEEN OBSOLETE BY THE TIME OF PRODUCTION.

TECHNOLOGICAL DEVELOPMENTS ARE PARTICULARLY IMPORTANT FACTORS LEADING TO CHANGE IN THE TELECOMMUNICATIONS INDUSTRY - BUT THEY ARE NOT THE ONLY FORCES TO BE RECKONED WITH. IT SEEMS THAT WE ARE WITNESSES TO, AND, TO A GREATER OR LESSER DEGREE, PARTICIPANTS IN, A CONFLUENCE OF FORCES FOR CHANGE - A CONFLUENCE THAT MAKES THIS PERIOD PARTICULARLY UNIQUE IN THE HISTORY OF THE INDUSTRY IN TERMS OF THE PERVASIVENESS AND RAPIDITY OF CHANGE. THE RELATIVELY RECENT PHENOMENON OF CONSISTENTLY HIGH LEVELS OF INFLATION HAS HAD AN

...2.

ENORMOUS EFFECT ON THE INDUSTRY, AS IT HAS ON US ALL. AS WELL, THE HEIGHTENING OF CONSUMER CONSCIOUSNESS AND ORGANIZATION, EVIDENCED BY DEMANDS FOR INCREASED CHOICE, BETTER QUALITY PRODUCTS AND SERVICES, MORE EQUITABLE COSTS, AND FULLER INVOLVEMENT IN DECISION-MAKING, HAS BEEN AN IMPORTANT FORCE FOR CHANGE. BUSINESS, IN THE AGE OF THE MULTINATIONAL CORPORATION, HAS ENGENDERED AN UNPRECEDENTED DEMAND FOR HIGH SPEED AND VOLUME DATA TRANSMISSION SERVICES, LOCALLY, REGIONALLY, NATIONALLY AND INTERNATIONALLY. FINALLY, INFORMATION EXCHANGE IS NOW INCREASINGLY RECOGNIZED AS AN IMPORTANT INDUSTRY IN ITSELF RATHER THAN BEING MERELY AN ADJUNCT TO THE CONDUCT OF OTHER BUSINESS.

ONE WRITER HAS OBSERVED THAT "THE GENIE" OF TECHNOLOGY IS NOW "OUT OF THE BOTTLE". WHATEVER THE GENIE GOT OUT OF, I THINK IT IS ACCURATE TO SAY, MIXING THE METAPHOR, THAT WHAT IT HAS LEFT BEHIND IS A "PANDORA'S BOX". THIS SITUATION, BROUGHT ABOUT BY THESE MANY FORCES OF CHANGE, DENIES TO US ALL THE LUXURY OF MAKING DECISIONS AS THOUGH THE GENIE WAS STILL BOTTLED - OR AS THOUGH THE SPILLED CONTENTS OF THE PANDORA'S BOX CAN EASILY BE ROUNDED UP AND STUFFED BACK INTO THEIR FORMER CONTAINER AS IF ALL WAS WELL IN THE BEST OF ALL BEAUTIFUL WORLDS.

...3.

IN THESE CIRCUMSTANCES, IT SEEMS TO ME THAT WE ARE IN THE POSITION WHERE WE MUST BE PREPARED TO THINK AND ACT BOLDLY AND I SUGGEST THAT WE SHOULD BEGIN THAT PROCESS BY RE-EXAMINING THE FUNDAMENTAL PREMISES UPON WHICH TELECOMMUNICATIONS REGULATION HAS BEEN BASED UNTIL NOW IN ORDER TO DECIDE WHETHER THEY HAVE CONTINUING MERIT IN THE NEW AGE IN WHICH WE ARE BEGINNING TO LIVE. IN THAT EXAMINATION IT HAS TO BE REMEMBERED THAT NEITHER TECHNOLOGY NOR PARTICULAR INDUSTRY STRUCTURES ARE ENDS IN THEMSELVES TO BE SERVED AND PROTECTED BY PUBLIC POLICY. IT IS PEOPLE - INDIVIDUAL HUMAN BEINGS - THAT HAVE TO BE SERVED. THIS OBJECTIVE PUTS A HEAVY BURDEN ON US ALL. IT CALLS ON US TO APPROACH OUR RESPECTIVE RESPONSIBILITIES AS PARTICIPANTS IN THE TELECOMMUNICATIONS INDUSTRY WITH RESPECT, CARE, OPENNESS AND SENSITIVITY. IT IS IN THIS SPIRIT THAT I AM PARTICULARLY GRATEFUL FOR THE OPPORTUNITY TO BE HERE AT THIS CONFERENCE, THE FIRST SUCH CONFERENCE IN WHICH I HAVE PARTICIPATED SINCE MY APPOINTMENT TO THE CRTC. IT IS A GREAT COMFORT TO ME TO KNOW THAT YOUR SKILLS, EXPERTISE AND EXPERIENCE ARE BEING DEVOTED TO THE THORNY PROBLEMS OF TELECOMMUNICATIONS POLICY. APART FROM THE VALUE TO THE PUBLIC, I HAVE A SELFISH MOTIVE: I EXPECT TO BE THE BENEFICIARY OF YOUR EFFORTS!

I TURN NOW TO MY SPECIFIC MANDATE TODAY WHICH IS TO LOOK TO THE FUTURE AND TO SAY SOMETHING ABOUT AN AGENDA FROM THE PERSPECTIVE OF THE CRTC. FOR US, THAT FUTURE AGENDA RELATES TO A RE-EXAMINATION OF THE FUNDAMENTAL PREMISES OF REGULATION.

A RE-EXAMINATION OF PREMISES

WHAT, THEN, ARE SOME OF THE FUNDAMENTAL PREMISES OF REGULATORY POLICY THAT NOW APPEAR TO REQUIRE RE-EXAMINATION?

(1) PERHAPS THE MOST BASIC PREMISE IS THAT MONOPOLY IS, IN THE MAIN, THE MOST APPROPRIATE INDUSTRY STRUCTURE FOR THE PROVISION OF TELEPHONE SERVICE. TO WHAT EXTENT IS THIS STILL SUPPORTABLE AND, IF SO, OVER WHAT RANGE OF SERVICES AND FACILITIES? ARE THERE EFFICIENCIES TO BE GAINED FROM VERTICAL INTEGRATION THAT JUSTIFY ITS CONTINUANCE? ARE THERE ECONOMIES OF SCALE IN THE PROVISION OF TELEPHONE SERVICE THAT CONTINUE TO JUSTIFY A MONOPOLY APPROACH TO THE PROVISION OF THAT SERVICE? TECHNOLOGICAL DEVELOPMENTS, AND THE ATTENDANT PHENOMENON OF INCREASING ENTRY OF NEW ACTORS INTO THE TELECOMMUNICATIONS MARKET, ARE FACTS OF LIFE THAT MAKE THE CONSIDERATION OF THESE ISSUES A PRESSING CONCERN. AT THE SAME TIME THE RAPIDITY OF CHANGE MAKES IT DIFFICULT TO IDENTIFY ALL THE CONSTITUENT FACTORS IN PLAY, MUCH LESS TO FIND ANSWERS. NOW MORE THAN EVER, USEFUL RESEARCH ON THESE ISSUES IS CALLED UPON TO TAKE A DYNAMIC AND PRAGMATIC APPROACH. WHILE THIS MAY INVOLVE THE SACRIFICE OF SOME PRECISION AND ELEGANCE IT HOLDS OUT THE EXCITEMENT OF MOVING BEYOND THE INHERENT LIMITS OF STATIC ASSUMPTIONS AND MODELS TO NEW APPROACHES.

THE RE-EXAMINATION OF THIS BASIC PREMISE OF TELECOMMUNICATIONS REGULATION WILL INVOLVE, AMONG OTHER THINGS, A CONSIDERATION OF THE ROLE OF COMPETITION IN THE CARRIER INDUSTRY. THE ALTERNATIVE OF A COMPLETELY COMPETITIVE INDUSTRY - THE OPPOSITE END OF THE SPECTRUM OF POSSIBILITIES FROM MONOPOLY - IS A TANTALIZINGLY SIMPLE SOLUTION. HOWEVER, AS I HAVE SAID, WE MUST LOOK UPON INDUSTRY STRUCTURES AS THE MEANS TO PARTICULAR ENDS AND NOT ENDS IN THEMSELVES. IT IS UNDERSTANDABLE THAT, IN THE WESTERN TRADITION OF SOCIO-POLITICAL PHILOSOPHY (AT LEAST SINCE THE 17TH CENTURY), COMPETITION WOULD HAVE TAKEN ON A POSITIVE, ALMOST FUNDAMENTALIST VALUE - EVEN THOUGH IT WOULD ONLY BE REALISTIC TO ADD THAT THE HISTORY OF PRACTICAL AFFAIRS IN THE WEST DOES NOT PARALLEL OUR PHILOSOPHICAL MUSINGS ON THIS SCORE.

NEVERTHELESS, IF WE ARE TO BE DILIGENT IN SEEKING TO SERVE THE NEEDS OF HUMAN BEINGS IN AN INCREASINGLY INTERDEPENDENT AND COMPLEX TELECOMMUNICATIONS WORLD, WE MUST APPROACH SUCH ARTICLES OF FAITH CAUTIOUSLY. THERE IS NO DIMINUTION, AS FAR AS I AM AWARE, IN THE COMMITMENT TO THE DESIRABILITY OF UNIVERSAL ACCESS TO REASONABLY PRICED BASIC TELEPHONE SERVICE. THERE IS NO DIMINUTION OF COMMITMENT TO THE NEED FOR A HIGH QUALITY OF TELECOMMUNICATIONS SERVICES. THESE COMMITMENTS <u>MAY</u> REQUIRE A MEASURE OF CROSS-SUBSIDIZATION

...6.

OF CERTAIN SERVICES BY OTHERS. TO SOME EXTENT, THEREFORE, SOLUTIONS WHICH SEEK A SHIFT COMPLETELY TO THE COMPETITIVE END OF THE SPECTRUM MAY CONFLICT WITH THOSE COMMITMENTS.

THIS IS NOT BY ANY MEANS TO SUGGEST THAT THERE IS NO PLACE FOR COMPETITION IN THE INDUSTRY. INDEED, THE RE-EXAMINATION OF THE MONOPOLY PREMISE NECESSARILY INVOLVES CAREFUL CONSIDERATION OF ALL OTHER OPTIONS. IN CANADA, FOR EXAMPLE, THERE HAS BEEN COMPETITION IN DATA COMMUNICATION SERVICES FOR OVER A DECADE AND I WOULD VENTURE THE OPINION THAT IT HAS NOT PROVED CONTRARY TO THE PUBLIC INTEREST.

(2) A SECOND PREMISE OF REGULATORY POLICY IS THAT TELECOMMUNICATIONS CARRIERS CAN AND SHOULD BE REGULATED AS PUBLIC UTILITY MONOPOLIES, EMPLOYING THE PRINCIPLE OF RATE BASE RATE OF RETURN IN DETERMINING THE JUSTNESS AND REASONABLENESS OF RATES. I BELIEVE THAT THIS APPROACH, PERMITTING CARRIERS TO EARN REVENUES TO COVER ALLOWABLE EXPENSES AND TO ENABLE A REASONABLE RETURN ON NET INVESTED CAPITAL, SHOULD ALSO BE RE-EXAMINED.

...7.

IN THIS AGE OF RAPID TECHNOLOGICAL CHANGE, INVOLVING SHORTER PRODUCT LIFE CYCLES AND INCREASED RATES OF PRODUCT OBSOLESCENCE, IT MAY NO LONGER BE APPROPRIATE TO RELY SOLELY ON THAT STANDARD, WITHOUT SOME MODIFICATION AND AUGMENTATION. UNDER MONOPOLY, RATE BASE RATE OF RETURN CONDITIONS, THE IMPACT OF INFLATION ALONE GENERATES INCREASINGLY FREQUENT APPLICATIONS FOR RATE INCREASES. ATTEMPTS BY THE CARRIERS TO MITIGATE THIS, BY CONTROLLING THE RATE OF PRODUCT OBSOLESCENCE, CAN HAVE DETRIMENTAL EFFECTS IF THEY RESULT IN THE SLOWER IMPLEMENTATION OF PRODUCT INNOVATIONS THAT WOULD BETTER SERVE SUBSCRIBERS.

SURELY IT IS DESIRABLE BOTH TO SEEK WAYS TO IMPROVE THE EFFICIENCY OF CARRIERS IN THE PROVISION OF SERVICES AND FACILITIES AND TO EDUCATE SUBSCRIBERS TO THE NEED FOR PRUDENT UTILIZATION OF, AND DEMAND FOR, SERVICES. IN THIS WAY PROGRESS MAY BE MADE IN COUNTERACTING THE FORCES BEHIND THE UPWARD COST SPIRAL.

(2) A THIRD PREMISE OF TRADITIONAL REGULATION, FLOWING FROM THE FIRST TWO, IS THAT <u>PRICING STRUCTURES BASED ON A HIGH DEGREE OF CROSS-SUBSIDY AND RATE AVERAGING ARE ACCEPTABLE</u>. THIS PREMISE, TOO, IT SEEMS TO ME, NEEDS RE-APPRAISAL. CENTRAL TO THIS RE-EXAMINATION MUST BE A CONSIDERATION OF THE EXTENT TO WHICH "VALUE OF SERVICE" PRICING SHOULD BE MODIFIED OR REPLACED BY PRICING BASED ON THE COST OF SPECIFIC

SERVICES. OF COURSE, CONSIDERATION MUST BE GIVEN TO WHICH SPECIFIC SERVICES ARE AMENABLE TO SUCH CHANGE. BUT HERE, AS WELL, THERE ARE DANGERS LURKING IN THE WINGS. WE NEED ONLY LOOK TO THE AMERICAN EXPERIENCE IN REGARD TO SERVICE COSTING TO REALIZE THAT, IN SEEKING NEW REGULATORY APPROACHES, WE MUST AVOID BITING OFF MORE THAN WE CAN CHEW. AN APPROACH EMPHASIZING THE DISAGGREGATION OF INDIVIDUAL RATES OR PARTICULAR RATE STRCUTURES, INVOLVING A REQUIREMENT TO CHOOSE AND EMPLOY COST ALLOCATION TECHNIQUES, CAN QUICKLY RESULT IN A RECOGNITION OF HOW LIMITED ARE THE RESOURCES AVAILABLE FOR THAT TASK.

THE RE-EXAMINATION OF THIS PREMISE WILL NECESSARILY INVOLVE A CONSIDERATION OF IMPORTANT SOCIAL FACTORS SUCH AS: THE TREATMENT THAT SHOULD BE ACCORDED TO SERVICES PROVIDED TO SUBSCRIBERS IN RURAL AND REMOTE AREAS AS COMPARED TO SUBSCRIBERS IN URBAN AREAS; THE TREATMENT THAT SHOULD BE ACCORDED TO SERVICES PROVIDED TO RESIDENTIAL SUBSCRIBERS AS COMPARED TO BUSINESS SUBSCRIBERS AND THE TREATMENT THAT SHOULD BE ACCORDED TO SERVICES PROVIDED TO A VARIETY OF SPECIAL INTEREST AND MINORITY GROUPS. THESE FACTORS ARE IN ADDITION TO THE ECONOMIC AND ECONOMIC WELFARE CONSIDERATIONS THAT ARE ALSO NECESSARY TO THE RE-EXAMINATION OF THIS PREMISE. HERE, PREEMINENTLY, THE DISCIPLINE OF ECONOMICS

...9.

MUST, IF IT IS TO BE OF PRACTICAL ASSISTANCE, BE INFORMED BY A BROAD SOCIO-POLITICAL UNDERSTANDING. I AM SURE THERE ARE AMONG YOU THOSE WHO WILL WELCOME THIS CHALLENGE - AND MEET IT.

(4) A FINAL IMPORTANT PREMISE OF TRADITIONAL TELECOMMUNICATIONS CARRIER REGULATION THAT I WANT TO SPEAK ABOUT IS ONE THAT HAS BEEN ADOPTED TO DETERMINE THE APPROPRIATE BOUNDARIES OF REGULATION. THIS PREMISE IS <u>THAT REGULATION SHOULD BE CONCERNED WITH CARRIAGE BUT NOT WITH CONTENT</u>. THIS PREMISE, AS WELL, IS BOUND UP WITH THE FIRST TWO AND IS ONE THAT TECHNOLOGICAL DEVELOPMENTS HAVE MADE EXTREMELY DIFFICULT TO CONTINUE TO APPLY. FOR EXAMPLE, SINCE 1966, THE FCC HAS BEEN STRUGGLING WITH DEFINITIONS AND SEMANTICS IN ORDER TO DELINEATE MARKETS IN THE FACE OF TECNOLOGICAL ADVANCES. SOME OF THE DEFINITIONS THEY HAVE ELABORATED INCLUDE:

- DATA PROCESSING
- COMMUNICATIONS
- COMMUNICATION PROCESSING
- HYBRID DATA PROCESSING
- HYBRID COMMUNICATIONS
- BASIC VOICE SERVICE

- NON-VOICE SERVICE
- ENHANCED NON-VOICE SERVICE
- SMART TERMINALS
- DUMB TERMINALS

EVENTUALLY, THE LAW OF DIMINISHING RETURNS BEGINS TO SUBVERT THE ENTIRE EFFORT AND ONE ENDS UP AS PERHAPS ONE SHOULD HAVE BEGUN - MUTE!

THE AMERICAN EXPERIENCE ILLUSTRATES THE DIFFICULTY OF THE TASK BUT THAT IS OF LITTLE COMFORT. THE NEW AGE IN TELECOMMUNICATIONS COMPELS US TO FACE THE PROBLEM. THE NEW TECHNOLOGY IS BLURRING THE DISTINCTION BETWEEN CARRIAGE AND CONTENT. IN A SINGLE UNIT, WORD PROCESSING, DATA PROCESSING, ELECTRONIC MESSAGE SERVICES, AND DATA AND VOICE COMMUNICATION SERVICES CAN NOW BE COMBINED. WHERE, NOW, ARE THOSE NICE BOUNDARIES BETWEEN COMMUNICATIONS, DATA PROCESSING, MAIL AND OFFICE EQUIPMENT? WHAT DO WE DO IN THEIR ABSENCE?

AGENDA FOR THE FUTURE

THE RESEARCH NECESSARY TO INFORM THE RE-EXAMINATION OF THE TRADITIONAL REGULATORY PREMISES THAT I HAVE BEEN TALKING ABOUT SHOULD BE ABLE TO KEEP US ALL GLORIOUSLY OVERWORKED, IF PERENNIALLY UNDER-PAID, FOR SOME CONSIDERABLE TIME TO COME. IN THIS CONTEXT

...11.

IT MIGHT NOW BE USEFUL FOR ME TO BE SOMEWHAT MORE SPECIFIC CONCERNING PARTICULAR ISSUES THAT WILL NEED TO BE DEALT WITH BY THE COMMISSION IN THE NEAR FUTURE. MANY OF THESE WILL ARISE IN THE COURSE OF THE FORMAL PROCEEDINGS THAT WE WILL BE ENGAGED IN RELATING TO TERMINAL ATTACHMENT AND SUBSEQUENT PHASES OF THE COST INQUIRY. BEFORE I TURN TO SOME OF THE MORE SPECIFIC ISSUES, I THINK IT MIGHT BE OF SOME INTEREST IF I OUTLINE ONE, AT LEAST, OF THE PRACTICAL CONSTRAINTS UNDER WHICH WE WORK AT THE CRTC. THIS OPERATIONAL CONTEXT IS A MOST IMPORTANT FACTOR IN DETERMINING OUR PRIORITIES AND INITIATIVES IN REGULATION.

I CANNOT EMPHASIZE STRONGLY ENOUGH THE SUBSTANTIAL CONSTRAINTS THAT ARE PLACED ON OUR TIME AT THE CRTC. IN THE YEAR 1980, FOR EXAMPLE, THE COMMISSION CONSIDERED 1,407 APPLICATIONS CONCERNING BROADCASTING MATTERS IN PUBLIC HEARINGS HELD IN ALL PARTS OF CANADA AND ISSUED 1,122 BROADCASTING DECISIONS. IN THE SAME YEAR, THE COMMISSION ISSUED 21 TELECOMMUNICATIONS DECISIONS - A DECEPTIVELY SMALL NUMBER SINCE MOST OF THOSE INVOLVED SIGNIFICANT AND DIFFICULT ISSUES THAT WERE THE SUBJECT OF LENGTHY PUBLIC HEARINGS. INDEED, WHILE WE ARE AHEAD IN THE NUMBERS GAME IN OUR TELECOMMUNICATIONS WORK AT THE COMMISSION THIS IS MORE THAN

...12.

BALANCED BY THE COMPLEXITY AND INTRICACY OF THE ISSUES INVOLVED. MOREOVER, THE COMMISSION MUST REGULARLY CONSIDER TARIFF FILINGS AND OTHER MATTERS, ITEMS WHICH, IN 1980, ACCOUNTED FOR IN EXCESS OF 600 ORDERS. (IF ANY OF YOU HAVE A PENCHANT FOR PUSHING PAPER, YOU SHOULD THINK OF JOINING US AT THE COMMISSION). WHILE I AM PRINCIPALLY INVOLVED IN TELECOMMUNICATIONS MATTERS I, AND THE OTHER VICE-CHAIRMAN OF THE CRTC ARE, TOGETHER WITH THE CHAIRMAN AND OUR OTHER COLLEAGUES ON THE NINE-MEMBER EXECUTIVE COMMITTEE, INVOLVED IN ALL AREAS OF THE COMMISSION'S WORK. WE SIT ON HEARINGS CONCERNING BOTH BROADCASTING AND TELECOMMUNICATIONS MATTERS. IN RELATION TO BOTH BROADCASTING AND TELECOMMUNICATIONS MATTERS IT IS ONLY THE EXECUTIVE COMMITTEE OF THE COMMISSION, OR A DESIGNATED PANEL, THAT HAS THE FINAL STATUTORY DECISION-MAKING AUTHORITY. SO YOU CAN SEE THAT PRIORITY SETTING MUST - OF NECESSITY - BE VERY MUCH INFLUENCED BY THIS "TIME PRESSURE" FACTOR.

SOME OF YOU MAY FEEL THAT I SHOULD BE EMBARRASSED TO MAKE SUCH AN ADMISSION. ONE COULD ARGUE THAT RATIONAL, RESPONSIBLE POLICY AND ACTION CAN ONLY BE DETERMINED IN A CALM AND SOMEWHAT DETACHED ATMOSPHERE. I WOULD CERTAINLY CONTEND THAT IT IS PART OF THE JOB OF THE CRTC TO ENSURE THAT THE REGULATORY PROCESS PROCEEDS

...13.

IN AN ORDERLY FASHION IN THE PUBLIC INTEREST. BUT THOSE WHO KNOW THAT PROCESS ALSO KNOW THAT "DETACHED" IT IS NOT AND "CALM" IT IS NOT. AT THE SAME TIME, I WOULD ALSO HAVE TO SAY THAT THE CONSTRAINTS OF TIME, HOWEVER ABHORRENT TO THE PURIST, CAN BE AN IMPORTANT FORCE IN LEADING THE COMMISSION TO SET PRIORITIES AND EVEN TO MAKE DECISIONS FROM WHICH IT MIGHT OTHERWISE SHY AWAY. OF COURSE, OPINIONS WILL DIFFER AS TO THE VALUE OF THE RESULTS THAT FLOW FROM THIS!

I ALSO WOULD LIKE TO REMIND YOU OF ONE OTHER IMPORTANT FEATURE OF THE CONTEXT IN WHICH WE WORK AT THE CRTC - A FEATURE I DO NOT THINK OF AS A CONSTRAINT. THERE IS A MIX OF EXPERTISE AMONG THE STAFF OF THE COMMISSION. THE COMMISSIONERS THEMSELVES HAVE VARIED BACKGROUNDS, EXPERIENCE AND EXPERTISE. IT IS ESSENTIAL THAT PROFESSIONAL INFORMATION AND ADVICE BE CLEAR AND BE CLEARLY UNDERSTANDABLE. FOR EXAMPLE, ECONOMISTS MUST BOTH SPEAK TO AND BE UNDERSTOOD BY LAWYERS. AND VICE VERSA. (TALK ABOUT THE MILLENNIUM!) AND BOTH MUST SPEAK TO AND BE UNDERSTOOD BY THE COMMISSIONERS. IN THIS SITUATION, NO ONE CAN BE CONTENT TO ADDRESS THEIR WORK ONLY TO THE INITIATED IN THEIR OWN DISCIPLINE. IT IS IMPORTANT THAT THOSE ENGAGED IN RESEARCH IN TELECOMMUNICATIONS MATTERS KEEP THIS IN MIND. THERE HAS BEEN A PRETTY CONSISTENT BIAS IN FAVOUR OF THE GENERALIST AS PUBLIC POLICY DECISION-MAKER. AND WHILE I TEND TO

...14.

SHARE THIS BIAS, AND TO BELIEVE THAT INFORMED GENERALISTS ARE LIKELY TO BE BETTER ABLE TO TAKE A BROAD VIEW OF THE PUBLIC INTEREST, IT IS OBVIOUSLY OF THE MOST PROFOUND IMPORTANCE THAT THEY BE INFORMED. AS YOU WELL KNOW, THE TASK OF INFORMING THE GENERALIST IS NOT SIMPLE, AND IT IS YOU, WITH YOUR VITALLY IMPORTANT SPECIALIZED EXPERTISE, ON WHOM WE RELY TO CARR OUT THIS TASK.

SPECIFIC AREAS OF RESEARCH

NOW I WOULD LIKE TO GIVE A BRIEF SHOPPING LIST OF SOME OF THE MORE SPECIFIC AREAS WHERE RESEARCH WOULD APPEAR TO BE DESIRABLE. I DON'T INTEND TO DISCUSS EACH ITEM ALTHOUGH, IF SO DESIRED, WE CAN PERHAPS COME BACK TO SOME OF THEM IN OUR DISCUSSION.

(1) ONE AREA OF POLICY RESEARCH RELATES TO THE GOALS OF REGULATION. WHAT SHOULD BE THE SHORT-TERM AND LONG-TERM OBJECTIVES OF REGULATION? IN PARTICULAR, WHAT ARE THE IMPORTANT SOCIAL AND ECONOMIC OBJECTIVES INVOLVED IN THIS PROCESS AND WHAT TRADE-OFFS AMONG THEM ARE MOST APPROPRIATE IN THE PUBLIC INTEREST?

(2) A RELATED AREA OF POLICY RESEARCH RELATES TO THE QUESTION OF DEREGULATION. WHICH TELECOMMUNICATIONS SERVICES AND FACILITIES, IF ANY, MIGHT BE DE-REGULATED AND UNDER WHAT TERMS AND CONDITIONS?

...15.

(3) A THIRD AREA OF RESEARCH RELATES TO RATE STRUCTURES. WHAT RATE STRUCTURE ALTERNATIVES ARE FEASIBLE AND DESIRABLE? WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF EMPLOYING A FACILITIES ORIENTED APPROACH TO PRICING? IS IT APPROPRIATE TO CONTINUE TO USE THE DISTINCTION BETWEEN "LOCAL" AND "TOLL" ? WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF ADOPTING A LOCAL MEASURED RATE, PARTICULARLY IN VIEW OF THE PROBLEM OF COSTING OF THE FACILITIES INVOLVED? TO WHAT EXTENT IS FURTHER EXPANSION OF EXTENDED AREA SERVICE (EAS) APPROPRIATE?

(4) A FOURTH AREA OF RESEARCH RELATES TO COMPETITION AND LIBERALIZED ENTRY. TO WHAT EXTENT IS IT FEASIBLE TO QUANTIFY THE BENEFITS ASSOCIATED WITH COMPETITION AND LIBERALIZED ENTRY, AND WHAT ARE THOSE BENEFITS? WHAT ARE THE STRENGTHS AND WEAKNESSES - BOTH IN TERMS OF SOCIAL AS WELL AS ECONOMIC EFFECTS - OF VARIOUS REGULATORY RESPONSES TO LIBERALIZED ENTRY? WHAT, FOR EXAMPLE, ARE THE PROS AND CONS OF DEPENDING ON SERVICE COSTING AS OPPOSED TO A REQUIREMENT THAT SERVICES BE PROVIDED THROUGH SEPARATE, ARMS-LENGTH SUBSIDIARIES?

...16.

(5) A FINAL AREA OF RESEARCH IS THAT OF THE REGULATORY PROCESS ITSELF. IN WHAT WAYS CAN THE REGULATORY PROCESS BE IMPROVED BOTH AS TO ITS FLEXIBILITY AND FAIRNESS? FOR EXAMPLE, THE CRTC IS THE ONLY FEDERAL REGULATORY COMMISSION THAT AWARDS COSTS TO PUBLIC INTEREST INTERVENERS; ARE THE BENEFITS ACHIEVED WORTH THE COSTS ASSOCIATED WITH IT? ARE THERE OTHER WORKABLE METHODS FOR FACILITATING RESPONSIBLE PUBLIC PARTICIPATION IN THE REGULATORY PROCESS?

CONCLUSION

WELL, THAT SCRATCHES THE SURFACE! IT IS LATE IN THE DAY AND MANY OF YOU HAVE FAR TO GO, SO I WILL LEAVE IT TO YOU TO FILL IN THE CREVICES.

I HOPE I HAVE SUCCEEDED IN INDICATING MY PERCEPTIONS OF THE RANGE OF ISSUES FACING THE CRTC. THE TASK BEFORE US IS A FORMIDABLE ONE WHICH, ALONE, WE WOULD HAVE DIFFICULTY IN MEETING. WE MUST HAVE YOUR HELP. THAT REALITY MAKES ME ESPECIALLY GRATEFUL THAT AN IMPORTANT CONFERENCE OF THIS NATURE HAS BEEN UNDERTAKEN IN CANADA AND I WANT TO TAKE THIS OPPORTUNITY TO THANK THE ORGANIZERS FOR HAVING MADE IT POSSIBLE. I HOPE THIS CONFERENCE WILL BE BUT THE FIRST OF MANY. I ALSO SINCERELY HOPE THAT ALL OF YOU WILL CONTINUE TO TURN YOUR IMPRESSIVE TALENTS TO A CONSIDERATION OF THE DIFFICULT PROBLEMS IN WHICH WE WHO ARE INVOLVED IN THE TELCOMMUNICATIONS SECTOR ARE ENGAGED AT THIS HISTORIC TIME.

I WANT TO TAKE A MOMENT, IN CONCLUDING MY REMARKS, TO RETURN TO THE GENERAL THEME WITH WHICH I BEGAN. ALL OF THE RESEARCH AND PLANNING AND DECISION-MAKING ON ALL THESE TECHNICAL AND COMPLEX ISSUES HAS AS ITS ULTIMATE GOAL THE PROVISION OF SERVICE TO THE PUBLIC. AS ALFRED KAHN HAS POINTED OUT, REGULATION IS NOT IN THE BUSINESS OF IMPOSING ITS MORAL OR AESTHETIC STANDARDS OF WHAT IS A GOOD LIFE ON THE PUBLIC. BUT WE ALL MUST RECOGNIZE THAT PLANNING AND DECISION-MAKING FOR THE TELECOMMUNICATIONS INDUSTRY NECESSARILY DOES AFFECT THOSE STANDARDS AND DOES PROVIDE LIMITS ON THE POSSIBLE DIRECTIONS IN WHICH THE SEARCH FOR THE GOOD LIFE CAN PROCEED. THAT AWARENESS DEMANDS FROM ALL OF US THAT WE ATTEND TO OUR OWN GROWTH AS HUMAN BEINGS AND THAT OUR ACTIONS RESPECT DEEPLY THAT IMPORTANT PROCESS IN ALL OTHER PERSONS.

BIBLIOGRAPHY

Abramovitz, M., "Resource and Output Trends in the United States Since 1870," American Economic Review, 46, No. 2 (May, 1956), pp. 5-23.

Adams, W., "The Role of Competition in the Regulated Industries," American Economic Review (May, 1958).

Adams, William J., "Firm Size and Research Activity: France and the U.S.," QJE, Vol. LXXXIV, No. 3.

Acheson, Keith, "The Role of Economic Theory and Evidence in Regulating Telecommunications," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Acton, J.P., B.M. Mitchell and R.S. Mowill (1976), "Residential Demand for Electricity in Los Angeles: An Econometric Study of Disaggregated Data," Rand Corporation (R-1899-NSF), Santa Monica, California, (September).

Allais, M., "La Theorie General des Surplus et l'apport Fondamental de Vilfredo Pareto," Revue d'Economie Politique, 83 (1973), pp. 1044-1097.

Allais, M., A la Recherche d'une Discipline Economique, Tome I, Ateliers Industria, Paris (1943).

Alleman, J.H. (1977), "The Pricing of Local Telephone Service," U.S. Department of Commerce, Office of Telecommunications, OT 77-14 (April).

Alleman, J.H., R.E. Park, B.M. Mitchell and B.M. Wetzel, "Charging for Local Telephone Calls: How Household Characteristics Affect the Number of Calls in the GTE Illinois Experiment," The Rand Corporation, R-2535-NSF, forthcoming.

Appelbaum, E. and R. Harris, "Capital, Accumulation and Investment in the Regulated Firm" (1978).

Arditti, F.D. and Y.C. Peles, "The Regulatory Process and the Firm's Capital Structure," The Financial Review, Vol. 15, No. 1 (1980), pp. 1-8.

Arnott, Richard J. and Joseph E. Stiglitz, "Aggregate Land Rents, Expenditure on Public Goods and Optimal City Size," Quarterly Journal of Economics, XCIII, No. 4 (November, 1979), pp. 471-500.

Arrow, K.J., Essays in the Theory of Risk-Bearing, Markham, Chicago (1971).

Arrow, K.J., "Vertical Integration and Communication," 6, Bell Journal of Economics and Management Science, #1 (1975), pp. 173-183.

Artle, R. and C. Averous, "The Telephone System as a Public Good: Static and Dynamic Aspects," Bell Journal of Economics, 4 (1973).

Atkinson, A.B. and J.E. Stiglitz (1972), "Indirect Taxation and Economic Efficiency," Journal of Public Economics, 1, pp. 97-120.

Auerbach, A.J., "Share Valuation and Corporate Equity Policy," Journal of Public Economics (1979).

Autin, C. and G. LeBlanc, "A National Telecommunications Planning and Policy Simulation Model," in Models and Decisions in National Economies, J.M.L. Janssen, L.F. Pau, A. Straszak (eds.), North-Holland Publishing Company (1979).

Autin, C. and G. LeBlanc, "Empirical Evaluation of Cross-Subsidy Tests for Canadian Interregional Telecommunications Network," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Autin, C. and G. St-Cyr, Expansion sur plusieurs periodes, Note technique, projet N.P.P.S., troisieme phase (avril, 1975).

Averch, H. and L.L. Johnson, "Behavior of the Firm Under Regulatory Constraint," American Economic Review, 52, pp. 1052-1069.

Averch, H. and L. Johnson, "Behavior of the Regulated Firm," American Economic Review, Vol. 52 (December, 1962), pp. 1052-69.

Baer, W.S., Testimony Prepared for Presentation to the Sub-Committee on Communications, Senate Commerce Committee, March 22, 1977, The Rank Corp. (March, 1977).

Baer, W.S. and B.M. Mitchell, "Impact of Competition on an Independent Telephone Company," Public Utilities Fortnightly (October 23, 1975).

Bailey, E.E., Economic Theory of Regulatory Constraint, D.C. Heath, Lexington, Mass (1973).

Baron, D.P. and R.A. Taggart, "A Model of Regulation Under Uncertainty and a Test of Regulatory Bias," The Bell Journal of Economics, 8, 1 (Spring, 1977), pp. 151-67.

Baron, D.P. and R.A. Taggart, "Regulation and the Investor-Owned Firm Under Uncertainty," Working Paper, Northwestern University (November, 1976).

Baron, D.P., "Investment Policy, Optimality and the Mean-Variance Model," Journal of Finance, 34, 1 (March, 1979), pp. 207-32.

Baude, J.A., G. Cohen, L. Garfinkel, M. Krehmeyer and J. Ogy, Perspectives on Local Measured Service, Proceedings of the Telecommunications Industry Workshop, Kansas City, Missouri (1979).

Baughcum, Alan, "On the General Impossibility of Communications Monopoly," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Baumol, W.J. (1979), "Quasi-optimality: The price we must pay for a price system," Journal of Political Economy, 87, pp. 578-599.

Baumol, William J., "On the Proper Cost Test for Natural Monopoly in a Multiproduct Industry," American Economic Review (AER) (December, 1977), pp. 809-822.

Baumol, William J., Welfare Economics and the Theory of the State, Harvard University Press (1965), 212 pages.

Baumol, William J., E.E. Bailey and R.D. Willig, "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly," AER (June, 1977), ...

Baumol, W.J. and D.F. Bradford (1970), "Optimal departures from marginal cost pricing," American Economic Review, 60, pp. 265-283.

Baumol, W.J. and Y.M. Braunstein, "Empirical Study of Scale Economies and Production Complementarity: The Case of Journal Publication," Journal of Political Economy (October, 1977), pp. 1037-48.

Baumol, W.J. and A.K. Klevorick, "Input Choices and Rate of Return Regulation: An Overview of the Discussion," The Bell Journal of Economics and Management Science, Vol. 1, No. 2 (1970), pp. 162-90.

Baumol, W.J., D. Fischer and M.I. Nadiri, "Forms for Empirical Cost Functions to Evaluate Efficiency of Industry Structure," Paper No. 30, Centre for the Study of Business Regulation, Graduate School of Business Administration, Duke University, Durham, N.C. (1978).

Baumol, W.J., E.E. Bailey and R.D. Willig, "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly," American Economic Review (June, 1977), pp. 350-65.

Beauvais, E.C. (1977), "The Demand for Residential Telephone Service Under Non-Metered Tariffs: Implications for Alternative Pricing Policies," paper presented at the Western Economic Association Meetings, Anaheim, California (June).

Becker, Gary S. (1965), "A Theory of the Allocation of Time," Economic Journal (September).

Beigie, Carl, "An Economic Framework for Policy Action in Canadian Telecommunications," in English, Edward H. (ed.), Telecommunications for Canada, Methuen Publications, 1973 study 2, p. 37.

Bell Canada General Increase in Rates, 1980, Part B: Memoranda of Support, February 19, 1980.

Berkowitz, M.K. and E.G. Cosgrove, "Financing and Investment Behavior of the Regulated Firm Under Uncertainty," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Berman, Paul J. Computer or Communications. Allocation of Functions and the Role of the Federal Communications Commission. Harvard University, Cambridge, Mass. Program on Information Technologies and Public Policy (May 28, 1974).

Berndt, E.R., "Comment on U.S. Productivity Growth by Industry, 1947 - 73," by F.M. Gollop and D.W. Jorgenson. In J.W. Kendrick and B.N. Vaccara, ed., New Developments in Productivity Measurement and Analysis, Chicago and London: National Bureau of Economic Research (1980), pp. 124-36.

Berndt, E. and M. Khaled, "Parametric Productivity Measurement and Choice Among Flexible Functional Forms," Journal of Political Economy, 87 (1979), pp. 1220-45.

Berndt, E.R. and D.O. Wood, "Engineering and Econometric Approaches to Energy Conservation and Capital Formation: A Reconciliation," MIT Energy Laboratory Working Paper (1977).

P.4

Berndt, E.R. and N.E. Savin, "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances," Econometrica, 43 (September - November, 1975), pp. 937-57.

Bernstein, J.I., "Corporate Investment and Financing with Taxes and Accounting Depreciation" (1980).

Bernstein, Jeffrey I., "Taxes, Financing and Investment for a Regulated Firm," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Berthon, D., J.L. Chabrol, F. Leroy and M. Werkoff, "Le Panel des abonnes au telephone," 3me Congres international sur l'Analyse, la Prevision et la Planification dans les Services Publics (Juin, 1980).

Berthon, D. and N. Curien, "Le plan de sondage utilise pour le Panel des abonnes au telephone," Note Direction Generale des Telecommunications, (Octobre, 1980).

Black, F. and M. Scholes, "The Pricing of Options and Corporate Liabilities," Journal of Political Economy, 81, 3 (May - June, 1973), pp. 637-54.

Blau, P.M., "Interdependence and Hierarchy in Organizations," Social Science Research, 1 (1972).

Boadway, R.W., "Corporate Taxation and Investment: A synthesis of the neo-classical theory," Canadian Journal of Economics, XIII, No. 2 (May, 1980), pp. 250-67.

Boadway, R.W. and N. Bruce, "Depreciation and Interest Deductions and the Effect of the Corporation Income Tax on Investment," Journal of Public Economics, 11 (1979), pp. 93-105.

Boiteux, M., "Le 'Revenu Distruable' et les Pertes Economiques," Econometrica, 19 (1951), pp. 112-133.

Boiteux, M., "Sur la Gestion des Monopoles Publics Asteints a l'Equilibre Bugetaire," Econometrica, 24 (1956), pp. 22-40.

Booth, A., "Telesat," 1 Canadian Bar Journal (N.S.) 25 (1970).

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, 26, pp. 211-243.

Box, C.E.P. and G.M. Jenkins (1976), Time Series Analysis Forecasting and Control, revised edition, Holden-Day, Inc., San Francisco.

Brander, J. and B. Spencer, "Optimal Discriminating Two-Part Tariffs," mimeo (1980).

Brander, James A. and Barbara J. Spencer, "Local Telephone Pricing: Two-Part Tariffs and Price Discrimination," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Brandon, B.B. ed. (1981), "The Effect of the Demographics of Individual Households on Their Telephone Usage in Chicago," Ballinger Publishing Co., Cambridge.

P.5

Breeden, D.T. and R.H. Litzenberger, "Prices of State-Contingent Claims Implicit in Option Prices," Journal of Business, 51, 4 (October, 1978), pp. 621-52.

Brennan, M., "Valuation and the Cost of Capital for Regulated Utilities: Comment," Journal of Finance, 27, 5 (December, 1972), pp. 1147-49.

Breslaw, Jon and J. Barry Smith, "Efficiency, Equity and Regulation: A Model of Bell Canada," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Breslaw, J., V. Corbo and J.B. Smith (1979a), "A direct test of the A-J effect: The case of Bell Canada," Working Paper #1979-22, Department of Economics and Institute of Applied Economic Research, Concordia University.

Breslaw, J., V. Corbo and J.B. Smith (1979b), "A micro test of the neoclassical production theory," Working Paper #1979-21, Department of Economics and Institute of Applied Economic Research, Concordia University.

Breslaw, J., V. Corbo and J.B. Smith (1979c), "More pitfalls in the testing of duality theory," Working Paper #1979-23, Department of Economics and Institute of Applied Economic Research, Concordia University.

Breslaw, J., V. Corbo and J.B. Smith (1979d), "The restrictiveness of flexible functional forms in the modelling of regulatory constraint," Working Paper #1979-20, Department of Economics and Institute of Applied Economic Research, Concordia University.

British Columbia Telephone Company, "General Decrease in Rate, 1980," Burnaby, B.C. (1980).

Brigham, E.F. and M.J. Gordon, "Leverage, Dividend Policy and the Cost of Capital," Journal of Finance, 23, 1 (March, 1968), pp. 85-103.

Brown, R.S. and L.R. Christensen, "Estimating Substitution Possibilities in a Model of Partial Static Equilibrium," Discussion Paper No. 8007, Social Systems Research Institute, University of Wisconsin-Madison (1980).

Burness, H.S., W.D. Montgomery and J.P. Quirk, "Capital Contracting and the Regulated Firm," American Economic Review, 70, 3 (June, 1980), pp. 342-54.

Burrows, E.M., "Office Employment and the Regional Problem," Regional Studies, 7 (1973).

Canada, The Department of Communications, Proposals for a Communications Policy for Canada: A Position Information Canada, Ottawa (1973).

Canada, The Department of Communications. Corporate Ownership and Integration in the Telecommunciations Industry. Study 2(f) Information Canada, Ottawa (1971).

Capron, William M. Technological Change in Regulated Industries. Brookings Institutions, Washington, D.C. (1971).

P. 6

Carr, J.L. and P.J. Halpern, "Vertical Integration and Regulation with Emphasis on Bell Canada and Northern Telecom," mimeo (June, 1977).

Caves, D.W., L.R. Christensen and J.A. Swanson, "Productivity in U.S. Railroads, 1951 - 1974," Bell Journal of Economics (Spring, 1980)(a).

Caves, D.W., L.R. Christensen and J.A. Swanson, "Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955 - 1974," Discussion Paper No. 8002, Social Systems Research Institute, University of Wisconsin-Madison (1980)(b).

Caves, D.W., L.R. Christensen and M.W. Tretheway, "Flexible Cost Functions for Multiproduct Firms," Review of Economics and Statistics (August, 1980), pp. 477-81.

Champness, B.G., "Attitudes Toward Person-Person Communications Media," Human Factors, 15 (1973).

Charyk, J.V., "Future of Satellite Telecommunication Systems," Telecommunication Journal (May, 1971).

Chaudry, M.A. and M. Burnside, "Net Income and Productivity Analysis (NIPA) as a Planning Model," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Chaudry, M.A., M. Burnside and D. Eldor, "NIPA: A Model for Net Income and Productivity Analysis," forthcoming.

Choudry, M.A., "Net Income Productivity Analysis," paper given at the American Productivity Centre Conference, Houston (1980).

Christensen, L.R., "U.S. Real Product and Real Factor Input, 1929 - 1967," Review of Income and Wealth, Series 16, No. 1 (1970), pp. 19-50.

Christensen, L.R. and W.H. Greene, "Economies of Scale in U.S. Electric Power Generation," Journal of Political Economy, 84 (August, 1976), pp. 655-76.

Christensen, L.R. and D.W. Jorgenson, "Measurement of U.S. Real Capital Input, 1929 - 1967," Review of Income and Wealth, Series 15, No. 4 (December, 1969), pp. 293-320.

Christensen, L.R. and D.W. Jorgenson, "Measuring the Performance of the Private Sector of the U.S. Economy, 1929 - 1969," in Moss (ed.), Measuring Economic and Social Performance, New York, National Bureau of Economic Research (1973), pp. 233-338.

Christensen, L.R. and D.W. Jorgenson, "The Measurement of U.S. Real Capital Input, 1929 - 1967," Review of Income and Wealth, 15 (December, 1969), pp. 293-320.

Christensen, L.R., D. Cummings and P.E. Schoech, "Productivity in the Bell System, 1947 - 1977," paper presented at the Eighth Annual Telecommunications Policy Research Conference, April 27 - 30, 1980.

Christensen, L.R., D. Cummings and P.E. Schoech, "Econometric Estimation of Scale Economies in Telecommunications," paper presented at the Conference "Telecommunications in Canada: Economic Analysis of the Industry" (1980).

Christensen, L.R., D. Cummings and P.E. Schoech, "Productivity in the Bell System, 1947 - 1977," paper presented at the Eighth Annual Telecommunications Policy Research Conference, Annapolis, Maryland (1980).

Christensen, L.R., D. Cummings, P.E. Schoech, "Econometric Estimation of Scale Economies in Telecommunications," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Christie, B. and S. Holloway, "Factors Affecting the Use of Telecommunications by Management," Journal of Occupational Psychology, 48 (1975).

Clark, D., "Communications and the Urban Future: A Study of Trunk Telephone Call Patterns in Wales," Regional Studies, 7 (1973).

Clarke, R.G., "The Effect of Fuel Adjustment Clauses on the Systematic Risk and Market Values of Electric Utilities," Journal of Finance, 35, 2 (May, 1980), pp. 347-58.

CNCP Telecommunications, "Here is What University of Toronto Professors Wilson, Waverman and Fuss Say About Interconnection," presentation to the CRTC by T.A. Wilson, L. Waverman and M.A. Fuss on CNCP Application for Interconnection (1977).

Congress, Subcommittee on Anti-Trust and Monopoly, Industrial Reorganization Act, Part 2: The Communications Industry (July 30 - 31, August 1 - 2, 1973).

Congress, Subcommittee on Anti-Trust and Monopoly, Industrial Reorganization Act, Part 5: The Communications Industry (June 20, 21, 25, 27, 1974).

Congress, Subcommittee on Communications, Amendments to the Communications Act of 1934, Part 3 (May 10, 11, 16 and June 5 - 7, 1979).

Congress, Subcommittee on Communications, Competition in the Telecommunications Industry (September 28 - 30, 1976).

Congress, Subcommittee on Communications, Domestic Common Carrier Regulation (November 10, 11, 13, 17, 18, 1975).

Congress, Subcommittee on Communications, Domestic Telecommunications Common Carrier Policies, Part 1 (March 21 - 22, 1977).

Congress, Subcommittee on Communications, Domestic Telecommunications Common Carrier Policies, Part 2 (March 23, 28, 1977).

Corbo, V., J. Breslaw, J.M. Dufour and J.M. Vrljicak, "A Simulation Model for Bell Canada: Phase II," Institute of Applied Economic Research Concordia University (March, 1979).

Corbo, V., J. Breslaw, J.M. Dufour and J.M. Vrejicak, "A Simulation Model of Bell Canada, Phase II," Economic Analysis Division Working Paper, Department of Communications, Ottawa (1979).

Cosgrove, J.G. and P.B. Linhart (1979), "Customer Choices Under Local Measured Telephone Service," Public Utilities Fortnightly, Vol. 17, (August).

Cowan, Donald D. and Leonard Waverman, "The Interdependence of Communications and Data Processing: Issues in Economies of Integration and Public Policy," Bell Journal of Economics and Management Science (Autumn, 1971).

Craig, C.E. and R.C. Harris, "Total Productivity Measurement at the Firm Level," Sloan Management Review, (Spring, 1973), pp. 13-28.

Criner, James C., History of Resale in Common Carrier Telecommunications, Office of Telecommunications, Washington, D.C. Policy Support Div. (December, 1975).

Criner, James C., "Telecommunications Resale: A Policy Analysis," Telecommunications Policy (September, 1977).

Curien, N., "La consommation telephonique: une approche microeconomique," 3me Congres international sur l'Analyse, la Prevision et la Planification dans les Services Publics (Juin, 1980).

Curien, N., "Le telephone et les biens a acces: modeles de demande et de consommation," Note Direction Generale des Telecommunication (Septembre, 1980).

Curien N. and Nguyen G. Dang, "Prevision du produit du trafic telephonique," 3me Congres international sur l'Analyse, la Prevision et la Planification dans les Services Publics (Juin, 1980).

Dalfen, Charles M., "The Telesat Canada Domestic Communications Satellite System," 1, Canadian Communications Law Review, 182 (1969).

Danielian, N.R., AT & T: The Story of Industrial Conquest, New York, Vanguard Press (1939).

Dansby, R.E. and J.C. Panzar (1981), "An Economic Analysis of Measured Service Options," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Davies, J.B. (1980), "Uncertain lifetime, consumption and dissaving in retirement," Journal of Political Economy, forthcoming.

Davis, B.E., C.J. Caccappolo and M.A. Chaudry (1973), "An Econometric Planning Model for American Telephone and Telegraph Company," Bell Journal of Economics and Management Science, 4 (Spring), pp. 29-55.

Day, George S. and William R. Allison, "Strategic Analysis of Emerging Telecommunications Markets," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

De La Brunetiere, Jean, "Rentabilite economique du trafic des nouveaux services sur les reseaux de telecommunication," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

de Fontenay, A., "Indexation for Telecommunications Services," in O. Anderson (ed.) Public Utilities Forecasting, North Holland (1980).

Debreu, G., "A Classical Tax-Subsidy Problem," Econometrica, 22 (1954), pp. 14-22.

Debreu, G., "The Coefficient of Resource Utilization," Econometrica, 19 (1951), pp. 273-292.

Demsetz, Harold, "Why Regulate Utilities," Journal of Law and Economics (April, 1968).

Denison, E.F., "Sources of Economic Growth in the United States and the Alternatives Before Us," Supplementary Paper 13, Committee for Economic Development (1962).

Denison, E.F., "Theoretical Aspects of Quality Change, Capital Consumption, and Net Capital Formation," in Capital Formation, Studies in Income and Wealth, Vol. 19, New York: National Bureau of Economic Research (1957).

Denny, M. and M. Fuss, "The Effects of Factor Prices and Technological Change on the Occupational Demand for Labour: Evidence from Canadian Telecommunications," Institute for Policy Analysis Working Paper No. 8014, University of Toronto (July, 1980).

Denny, M. and M. Fuss, "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates," American Economic Review (June, 1977), pp. 404-18.

Denny, M., M. Fuss and C. Everson, "Productivity, Employment and Technical Change in Canadian Telecommunications: The Case for Bell Canada," Final Report to the Department of Communications (March, 1979).

Denny, M., M. Fuss and J.D. May, "Inter-temporal Changes in Regional Productivity in Canadian Manufacturing," Institute for Policy Analysis, University of Toronto, Working Paper No. 8017 (1980).

Denny, M., C. Everson, M. Fuss and L. Waverman, "Estimating the Effects of Diffusion of Technological Innovations in Telecommunications: The Production Structure of Bell Canada," paper presented at the Seventh Annual Telecommunications Policy Research Conference April 29 - May 1, 1979, published in the Canadian Journal of Economics, February, 1981 (1979), pp. 24-43.

Denny, M., M. Fuss and L. Waverman, "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications," Working Paper No 7911, Institute for Policy Analysis, University of Toronto (1979).

Denny, M., M. Fuss and L. Waverman, "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications" (1979), paper presented at the National Science Foundation Conference on Productivity Measurement in Regulated Industries, Madison, Wisconsin, April 30 - May 1, 1979; published as Chapter 8 in T. Cowing and R. Stevenson (eds.) Productivity Measurement in Regulated Industries, New York: Academic Press, 1981, pp. 179-218.

Denny, M., A. de Fontenay and M. Werner, "Comparative Efficiency in Canadian Telecommunications: An Analysis of Methods and Uses, Phase II: Productivity Employment and Technical Change in Canadian Telecommunications," Report to the Canadian Department of Communications (May, 1980).

Denny, M., A. de Fontenay and M. Werner, "Comparing the Efficiency of Firms: Canadian Telecommunications Companies," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Denny, M., A. de Fontenay and M. Werner, "Total Factor Productivity for Management: The Post-Mortem and Planning Frameworks," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Deschamps, P.J. (1974), "The Demand for Telephone Calls in Belgium, 1961-1969," paper presented at the Birmingham International Conference in Telecommunications Economics, Birmingham, England (May).

Diamond, P.A., "A Many Person Ramsey Tax Rule," Journal of Public Economics, Vol. 4, No. 4 (November, 1975), pp. 335-342.

Diamond, P.A. and J. Mirrlees, "Optimal Taxation and Public Production," American Economic Review, Vol. 61 (March and June, 1971), pp. 8-27 and 261-278.

Diewert, W.E., "Aggregation Problems in the Measurement of Capital," Discussion paper 77-09, Department of Economics, University of British Columbia (1977).

Diewert, W.E., "Applications of Duality Theory," pp. 106-171 in M.D. Intriligator and D.A. Kendrick (eds.), Frontiers of Quantitative Economics, Vol. II, North-Holland, Amsterdam (1974).

Diewert, E., "Exact and Superlative Index Numbers," Journal of Econometrics, 4 (1976), pp. 115-45.

Diewert, W.E., "Hicks' Aggregation Theorem and the Existence of a Real Value-Added Function," pp. 17 - 51 in M. Fuss and D. McFadden (eds.), Production Economics: A Dual Approach to Theory and Applications, Vol. 2, North-Holland, Amsterdam (1978)..

Diewert, W.E., "The Measurement of Deadweight Loss Revisited," Econometrica, (1981b), forthcoming.

Diewert, W.E., "The Theory of Total Factor Productivity Measurement in Regulated Industries," forthcoming in T. Cowing and R. Stevenson (eds.), Productivity Measurement in Regulated Industries, Academic Press, New York (1981a).

Diewert, W.E., "The Welfare Effects of a Regulatory Constraint: A Productive Efficiency Approach," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Dittberner Associates, Interconnection, An Economic Impact Analysis, Office of Telecommunications Policy, Washington, D.C. (1973).

Dobell, A.R., L.D. Taylor, L. Waverman, T. Liu and M.D.G. Copeland, "Telephone Communications in Canada: Demand, Production and Investment Decisions," The Bell Journal of Economics and Management Science, 5 (1972), pp. 175-219

Doern, "The Regulatory Process in Canada," in The Regulatory Process in Canada (1978), pp. 13-18.

Dunn, D.A., "Policy Issues by the Interdependence of Computer and Communications Services," Law & Contemporary Problems (Spring, 1969).

Economic Council of Canada, Living Together, A Study of Regional Disparities, Ottawa (1977).

Economic Council of Canada, Responsible Regulation, An Interim Report by the Economic Council of Canada (November, 1979).

Eger, John M., "AT & T Must Accept Competition, Not Monopoly if it is to Survive," Communications News (1977).

Eger, John M., "The Bell Bill: The Effort of the Bell System to Stiffle Competition," Computers and People (October, 1977).

Ekern, S. and R. Wilson, "On the Theory of the Firm is an Economy with Incomplete Markets," The Bell Journal of Economics and Management Science, 5, 1 (Spring, 1974), pp. 171-80.

Elton, E.J. and M. Gruber, "Optimal Investment and Financing Patterns Under Alternative Methods of Regulation," in Financial Decision Making Under Uncertainty, edited by H. Levy and M. Sarnat, New York: Academic Press (1977), pp. 265-91.

Elton, E.J. and M. Gruber, "Valuation and the Cost of Capital for Regulated Industries," Journal of Finance, 27, 5 (December, 1972), pp. 1150-55.

Elton, E.J. and M. Gruber, "Valuation and the Cost of Capital For Regulated Industries," Journal of Finance, Vol. 26, No. 3 (1971), pp. 661-70.

Elton, E.J. and M. Gruber, "Valuation, Optimum Investment and Financing for the Firm Subject to Regulation," Journal of Finance (1975).

Elton, E.J., M. Gruber, and Z. Lieber, "Valuation, Optimum Investment and Financing for the Firm Subject to Regulation," Journal of Finance, Vol. 30, No. 2 (May, 1975), pp. 401-25.

English, Edward, "Canadian Telecommunications: Problems & Policies," in Study 1, p. 3.

English, H. (ed.), Telecommunications for Canada, Toronto: Methuen (1977).

Fabricant, S., "Basic Facts on Productivity Change," Occasional Paper 63, New York: National Bureau of Economic Research (1959).

Fama, E. and M.H. Miller, The Theory of Finance, New York: Holt, Rinehart and Winston (1972).

Faraday, J.E., The Management of Productivity, Management Publications Limited, Tonbridge, Kent, Great Britain (1971).

Fask, A. and P.B. Robinson (1977), "The Analysis of Telephone Demand by Dynamic Regression," unpublished paper, AT & T (September).

Faulhaber, G.R., "Cross-subsidization: Pricing in Public Enterprises," Bell Lab. Econ., Discussion paper (May, 1975).

Faulhaber, Gerald R., "Cross-Subsidization: Pricing in Public Enterprises," AER (December, 1975), pp. 966-977.

Faulhaber, G.R., On Subsidization: Some Observations and Tentative Conclusions, Conf. on Communication Policy Res., Office of Telecommunication Policy, Washington (1972).

Faulhaber and Panzar, "Optimal Two-Part Tariffs with Self-Selection," Bell Laboratories Economic Discussion Paper #74 (1977).

Faulhaber, Gerald R. and E.E. Zajac, "Some Thoughts on Cross-Subsidization in Regulated Industries," Bell Laboratories Economic Discussion Paper No. 48.

FCC, Economic Implications and Interrelationships Arising from Policies and Practices Relating to Customer Interconnection, Jurisdictional Separations and Rate Structures, Docket #20003 (September 27, 1976).

FCC, Economic Implications and Interrelationships Arising from Policies and Practices Relating to Customer Interconnection, Jurisdictional Separations and Rate Structures, Docket #20003, Second Report (January, 1980).

FCC, In the Application of Microwave Communication Inc. (specialized common carrier services from Chicago - St. Louis), 18 FCC 2d 953 (1969).

FCC, In the Matter of Allocation of Frequencies in the Bands Above 890, 27 FCC 359 (1959).

FCC, In the Matter of American Telephone and Telegraph Company, Long Lines Department, Private Line Services, Docket #18128, Memorandum Opinion and Order, 61 FCC 2d 587.

FCC, In the Matter of AT & T, Docket #19129, 64 FCC 2d 1 (1977).

FCC, In the Matter of Charges in the Corporate Structure and Operations of the Communications Satellite Corporation, Docket #80-634, Notice of Proposed Rule-Making (October, 1980).

FCC, In the Matter of Establishment of Policies and Procedures for Consideration of Application to Provide Specialized Common Carrier Services in the Domestic Public Point-to-Point Microwave Radio Service, 29 FCC 2d 870 (1971).

FCC, In the Matter of MCI Telecommunications Corporation (Execunet), 60 FCC 2d 25 (1976).

FCC, In the Matter of MTS and WATS Market Structure, Docket #78-72, Report and Third Supplemental Notice (August, 1980).

FCC, In the Matter of Regulatory and Policy Problems Presented by the Interdependence of Computer and Communication Services and Facilities, "Second Computer Inquiry," Docket #20828 (tentative decision July, 1979) final decision April, 1980; modification October, 1980.

FCC, In the Matter of Regulatory and Policy Problems Presented by the Interdependence of Computer and Communication Services and Facilities, 28 FCC 2d (1970) (tentative decision) final decision [illegible]

FCC, In the Matter of Regulatory Policies Concerning Resale and Shared Use of Common Carriers, Docket #20097, 60 FCC 2d 261 (1976).

FCC, In the Matter of the Establishment of Domestic Communications Satellite Facilities by Non-Government Entities, Docket #16495 - Domsat Decision, First Order and Report, 22 FCC 2d 86 (1970); 2nd Order and Report, 35 FCC 2d 844 (1972).

FCC, In the Matter of Use of the Carterphone Device in Message Toll Telephone Service, 13 FCC 2d 420 (1968) recon. denied, 14 FCC 2d 571-573 (1968).

FCC, Proposals for New or Revised Classes of Interstate and Foreign Message Toll Service (MTS) and Wide Area Telephone Service (WATS) First Report and Order, 56 FCC 2d 593 (1975) Second Report and Order, 58 FCC 2d 739 (1976), Third Report and Order, FCC 2d (1978), Docket #19528.

FCC, Report of the FCC on Domestic Telecommunications Policy, submitted to the Subcommittee on Communications, Committee on Interstate and Foreign Commerce, September 30, 1976. (Also found in Hearings of the Subcommittee, September 30, 1976, Serial No. 94-129).

Federal-Provincial Conference of Communications Ministers (1979), "Report of the federal-provincial working group on competition/industry structure in the telecommunications industry," Toronto.

Feldman, J. (1976), "A Preliminary Cross Sectional Analysis of Services," unpublished paper, AT & T (February).

Feldstein, M.S. (1972a), "Distributional equity and the optimal structure of public prices," American Economic Review, pp. 32-36.

Feldstein, M.S. (1972b), "Equity and efficiency in public sector pricing: The optimal two-part tariff," Quarterly Journal of Economics, pp. 175-187.

Feldstein, M.S. (1972c), "Pricing of public intermediate goods," Journal of Public Economics, 1, pp. 45-72.

Fellner, W. (1967), "Operational utility: The theoretical background and a measurement," in Ten Economic Studies in the Tradition of Irving Fisher, Wiley.

Ferguson, C.E., The Neoclassical Theory of Production and Distribution, Cambridge University Press (1969).

Fisher, I., The Making of Index Numbers, Boston, Houghton Mifflin (1922).

Frish, R. (1959), "A complete scheme for computing all direct and cross demand elasticities in a model with many sectors," Econometrica, 27, pp. 177-196.

Froggatt, A.M., "Incremental Costing in Practice," in Essays on Public Utility Pricing and Regulation by Harry M. Trebing (ed.), Lansing: Michigan State University Press (1971).

Fuss, M.A., "A Survey of Recent Results in the Analysis of Production Conditions in Telecommunications," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Fuss, M.A., "The Demand for Energy in Canadian Manufacturing: An Example of the Estimation of Production Structures with Many Inputs," Journal of Econometrics, January (1977), pp. 89-116.

Fuss, M., "The Structure of Technology Over Time: A Model for Testing the 'Putty-Clay' Hypothesis," Econometrica, 45, 8 (November, 1977), pp. 1797-1821.

Fuss M. and L. Waverman, "Regulation and the Multiproduct Firm: The Case of Telecommunications in Canada," paper presented at the NBER Conference on Public Regulation, Washington, December (1977), forthcoming as Chapter 6 in G. Fromm (ed.) Studies in Public Regulation, Cambridge: M.I.T. Press, 1981. This paper also appeared as "Multi-product, Multi-input Cost Functions for a Regulated Utility: the Case of Telecommunications in Canada," Institute for Policy Analysis, Working Paper No. 7810.

Fuss M. and L. Waverman, "The Regulation of Telecommunications in Canada," (1980), Draft of the Final Report to the Economic Council of Canada Regulation Reference, June; Final Draft, February, 1981.

Fuss, M. and L. Waverman, "Multi-Product, Multi-Input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada," paper presented at the N.B.E.R. Conference on Public Regulation, Washington, December, 1977.

Fuss, M. and L. Waverman, "Multi-Product, Multi-Input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada," Draft, Institute for Policy Analysis, June, 1978, revision of Fuss and Waverman (1977).

Fuss, M. and L. Waverman, "Multi-product Multi-input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada," Institute for Policy Analysis Working Paper No. 7810, University of Toronto, Toronto (1978).

Gabel, R., "The Early Competitive Era in Telephone Communication, 1893 - 1920," Law and Contemporary Problems, Vol. 34 (Spring, 1969).

Gable, R., Development of Separations Principles in the Telephone Industry, Lansing: Michigan State University Press (1967).

Gabor, A., "A Note on Block Tariffs," Review of Economic Studies, Vol. 23 (1955), pp. 32-41.

Gale, W.A. (1971), "Duration of Interstate Calls, March 1969," unpublished Bell Laboratories memorandum (December).

Gamble, R.B., "Van Services in the U.S.," Telecommunications (July, 1978).

Gamot, G., "Theorie microeconomique," Cours professe a l'Ecole Nationale Superieure des Telecommunications.

Garfinkel L. and P.B. Linhart (1979), "The Transition to Local Measured Telephone Service, Public Utilities Fortnightly, Vol. 17 (August).

Gensollen, M. and E. Vilmin, "Prevision de la demande telephonique," 3me Congres international sur l"Analyse, la Prevision et la Planification dans les Services Publics (Juin, 1980).

Goddard, J.B., "Office Linkages and Location," Progress in Planning, 1 (1973).

Goddard, J.B., "Office Location and Urban and Regional Development in Britain," in P.W. Daniels (ed.), Spatial Patterns of Office Growth and Location, London: Wiley (1978).

Goddard, J.B., "Organizational Information Flows and the Urban System," Economie Appliquee, 1 (1975).

Goddard, J.B., "Technology Forecasting in a Spatial Context," Futures (April, 1980).

Goddard, J.B. and D. Morris, "The Communications Factor in Office Decentralization," Progress in Planning, 6 (1976).

Goddard, J.B. and R. Pye, "Telecommunications and Office Location," Regional Studies, 11 (1977).

Goldman, M.B., H. Leland and D. Sibley, "Optimal Nonuniform Prices," Bell Laboratories Economic Discussion Paper #100 (1977).

Gollop, F.M. and D.W. Jorgenson, "U.S. Productivity Growth by Industry, 1947 - 73," In J.W. Kendrick and B.N. Vaccara, ed., New Developments in Productivity Measurement and Analysis, Chicago and London: National Bureau of Economic Research (1980), pp. 17-124.

Goodier, Ray J., "Rate Adjustment Formula Procedure (RAFP): A Simple Formula Solution to the Problem of Regulation in the Face of Rising Costs," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Government of Canada, Department of Communications, Telecommission Study 2(d): Telecommunications and Regional Development, Ottawa: Information Canada (1971).

Gorman, W.M., "Measuring the Quantities of Fixed Factors," pp. 141-172 in in J.N. Wolfe (ed.), Value, Capital and Growth: Papers in Honour of Sir John Hicks, Aldine Publishing Co., Chicago (1968).

Gordon, M.J., "Some Estimates of the Cost of Capital to the Electric Utility Industry, 1954 - 1957: A Comment," American Economic Review, 56, 5 (December, 1967), pp. 1267-77.

Gordon, M.J. and McCallum, J., "Valuation and the Cost of Capital for Regulated Utilities: Comment," Journal of Finance, 27, 5 (December, 1972), pp. 1141-46.

Greenwald, B.C., "Admissible Rate Bases, Fair Rates of Return and the Structure of Regulation," Journal of Finance, 35, 2 (May, 1980), pp. 359-68.

Griliches, Z., "Production Functions in Manufacturing: Some Preliminary Results;" in M. Brown (ed.), The Theory and Empirical Analysis of Production, New York: National Bureau of Economic Research (1967), pp. 275-322.

Griliches, Z., "Research Expenditures, Education and the Aggregate Agricultural Production Function," American Economic Review (December, 1964), pp. 961-74.

P.16

Griliches, Z., "The Sources of Measured Productivity Growth: U.S. Agriculture, 1940 - 1960," Journal of Political Economy (August, 1963), pp. 331-46.

Griliches, Z. and D.W. Jorgenson, "Sources of Measured Productivity Change: Capital Input," American Economic Review, 56, No. 2 (May, 1966), pp. 50-61.

Gronau, R. (1970), "The Value of Time in Passenger Transportation: The Demand for Air Travel," National Bureau of Economics Research, Inc., New York.

Guilkey, D.K. and C.A.K. Lovell, "On the Flexibility of the Translog Approximation," International Economic Review, February (1980), pp. 137-148.

Hall, R.E. and D.W. Jorgenson (1967), "Application of the theory of optimal capital accumulation," in Tax Incentives and Capital Spending, G. Fromm (ed.), The Brookings Institution.

Hall, R.E. and D.W. Jorgenson, "Tax Policy and Investment Behaviour," American Review, 57, No. 3 (1967), pp. 391-414.

Hamada, R.S., "Portfolio Analysis, Market Equilibrium and Corporate Finance," Journal of Finance, 24(1) (1969).

Hanoch, G., "The Elasticity of Scale and the Shape of the Average Costs," The American Economic Review (June, 1975), pp. 492-96.

Harberger, A.C., "The Measurement of Waste," The American Economic Review, 54 (1964), pp. 58-76.

Harberger, A.C., "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," The Journal of Economic Literature, 9 (1971), pp. 785-797.

Harris, M. and A. Raviv, "Some Results on Incentive Contracts with Application to Education and Employment, Health Insurance and Law Enforcement," American Economic Review, Vol. 68 (1979), pp. 20-30.

Hartwick, J.M., "Optimal Price Discrimination," Journal of Public Economics, Vol. 9, No. 1 (February, 1978), pp. 83-89.

Hayek, F., Der Wettbewerb als Entdeckungsverfahren, Kieler Vortrage, No. F. 56, Kiel (1968).

Hazelwood, A., "Optimum Pricing as Applied to Telephone Service," Rev. Econ. Studies, Vol. 18, No. 2 (1950-51), pp. 67-78.

Henle, P. and P. Ryscavage, "The Distribution of Earned Income Among Men and Women, 1958 - 77," Monthly Labour Review, Vol. 103, No. 4 (April, 1980).

Hicks, J.R., Value and Capital, second edition, Clarendon Press, Oxford (1946).

Hirshleifer, J., "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, Vol. 72 (1958).

Holmstrom, B., "Moral Hazard and Observability," Bell Journal of Economics, Vol. 10 (1979), pp. 55-73.

Hoover, E., An Introduction to Regional Economics, New York: Knopf (1975).

Hotelling, H., "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Econometrica, 6 (1938), pp. 242-269.

Houthakker, H.S. (1960), "Additive preferences," Econometrica, pp. 224-257.

Hulten, C.R., "Divisia Index Numbers," Econometrica, 41, No. 6 (November, 1973), pp. 1017-26.

Infosino, W.J. (1976), "Estimating Flat Rate Residence Average Local Calling Rate from Demographic Variables," unpublished Bell Laboratories memorandum (August).

International Telegraph and Telephone Consultative Committee, Economic Studies at the National Level in the Field of Telecommunications, Geneva: ITU (1973).

Irish, W.F. (1974), "A Market Analysis of the Demand for Telephone Sets by Class of Service," North Carolina Utilities Commission (December)..

Irwin, M.R., The Telecommunications Industry: Integration vs. Competition, New York: Praeger (1971).

Irwin, Manley R., "Vertical Integration and the Communications Industry: Separation of Western Electric and AT & T?" Anti-Trust Law and Economics Review (Fall, 1960).

Irwin, Manley R., The Telecommunications Industry, Integration & Competition, Praeger Publishers, New York (1971).

Irwin, Manley and Robert McKee, "Vertical Integration and the Communications Equipment Industry: Alternatives for Public Policy," 53, Cornell L. Rev. 446 (1968).

Isard, W., Introduction to Regional Science, Englewood Cliffs: Prentice Hall (1975).

Jaffe, J.F. and G. Mandelker, "The Value of the Firm Under Regulation," Journal of Finance, Vol. 31, No. 2 (1976), pp. 701-14.

Jaffe, J.F. and G. Mandelker, "The Value of the Firm Under Regulation," Journal of Finance, Vol. 31, No. 2 (May, 1976), pp. 801-13.

James, V., J. Marshall and N. Waters, Telecommunications and Office Location, Final Report, Centre for Urban and Regional Development Studies, Newcastle (1979).

Janisch, Hudson, "Policy Making in Regulation: Towards a New Definition of the Status of Independent Regulatory Agencies in Canada," 17, Osgoode Hall L.J., p. 46.

Jensen, M.C. and W.H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership of Structure," Journal of Financial Economics, Vol. 3 (1976), pp. 304-60.

Johnson, L.L., Problems of Regulating Specialized Telecommunications Common Carriers, The Rand Corp. (May, 1976).

Johnson, N., "Harnessing Revolution: The Role of Regulation and Competition for the Communications Industries of Tomorrow," Anti-Trust Bulletin

Jorgenson, D.W. and L.J. Lau, "Duality and Technology," Amsterdam (1977).

Jorgenson, D.W. and Z. Griliches, "The Explanation of Productivity Change," Review of Economic Studies, 34, No. 99 (1967), pp. 249-83.

Jorgenson, D.W., "Capital Theory and Investment Behaviour," American Economic Review Papers and Proceedings, 53, No. 2 (1963), pp. 247-59.

Jorgenson, D.W., "The Theory of Investment Behaviour," in R. Ferber (ed.), Determinants of Investment Behaviour, New York: National Bureau of Economic Research (1967).

Joskow, P.L., "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Regulation," Journal of Law and Economics, 17 (October, 1974), pp. 291-327.

Joskow, P.L., "Pricing Decisions of Regulated Firms: A Behavioral Approach," The Bell Journal of Economics and Management Science, 4, 1 (Spring, 1973), pp. 118-40.

Kahn, Alfred E., The Economics of Regulation: Principles and Institutions, Vols. 1 & 2, Wiley & Sons, Inc., N.Y. (1970).

Kamien, M.I. and N.L. Schwartz, "Market Structure and Innovation: A Survey," Journal of Economic Literature (March, 1975).

Kaysen, C. and D. Turner, Antitrust Policy, An Economic and Legal Analysis, Harvard University Press, Cambridge, Mass. (1965).

Kim, E.H., "A Mean-Variance Theory of Optimal Capital Structure and Corporate Debt Capacity," Journal of Finance, Vol. 33, No. 1 (1978), pp. 45-63.

King, M.A., "Taxation and the Cost of Capital," Review of Economic Studies (1974).

Kearns, T.J. (1978), "Modeling the Demand for Residence Extensions," unpublished Bell Laboratories memorandum (February).

Kendrick, J.W., "Postwar Productivity Trends in the United States, 1948 - 1969," New York: National Bureau of Economic Research (1973).

Kendrick, J.W., "Productivity Trends in the United States," Princeton (1961).

Khaled, M.S., "Productivity Analysis and Functional Specification: A Parametric Approach," Ph.D. dissertation, University of British Columbia, Department of Economics (April, 1978).

Kiss, F., "The Bell Canada Productivity Study," Corporate Analysis Division, Bell Canada (February, 1981).

Kiss, F., "The Bell Canada Productivity Study," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Kiss, F. and J.R. Joseph, "Bell Canada Productivity: How Productivity Gains Increase Income and Absorb Inflation," proceedings of the 6th International Cost Engineering Congress, Mexico City, Vol. II (October 20 - 22, 1980).

P. 19

Kiss, F., S. Karabadjian and B. Lefebvre, "Economies of Scale and Scope in Bell Canada," Corporate Analysis Division, Bell Canada (1981).

Kiss, F., S. Karabadjian and B. Lefebvre, "Economies of Scale and Scope in Bell Canada: Some Econometric Evidence," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Klevorick, A.K., "The Behavior of a Firm Subject to Stochastic Regulatory Review," Bell Journal of Economics (1973).

Kloek, T., Indexcijfers: enige methodologisch aspecten, The Hague, Pasmans (1966).

Knieps, G., J. Muller and C.C. von Weizsacker, Die Rolle des Wettbewerbs im Fernmeldebereich, unpublished report submitted to the W. German Monopolies Commission, Bonn (1980).

Kochen, M and K. Deutsch, "Toward a Rational Theory of Decentralization: Some Implications of a Mathematical Approach," American Political Science Review, 63 (1969).

Koopmans, Tjalling, C., Three Essays on the State of Economic Science, McGraw-Hill Book Company (1957), 231 pages.

Kraepelien, H.Y. (1958), "The Influence of Telephone Rates on Local Traffic," Ericsson Technics, No. 2, published by Telefonaktiebolaget, Stockholm: L. M. Ericsson.

Kraus, A. and R. Litzenberger, "A State-Preference Model of Optimal Financial Leverage," Journal of Finance, Vol. 28, No. 4 (1973), pp. 911-22.

Kraus, J., "Productivity and Profit Models of the Firm," Business Economics, (September, 1978), pp. 10-14.

Krouse, C.G., "Optimal Financing and Capital Structure Programs for the Firm," Journal of Finance, Vol. 27, No. 5 (December, 1972), pp. 1057-71.

Krutilla, J.V. (1967), "Observation Reconsidered," American Economic Review, 57 (September): pp. 777-786.

Kuehn, R.A., "Competition May Control Telephone Costs," Management Review, Vol. 64, No. 9 (September, 1975).

Kwok, P.K., P.C. Lee and J.C. Pearce (1975), "Econometric Models of the Demand for Bell Originated Message Toll," Bell Canada, CCNS Network Market Planning, Ottawa, Ontario (January).

Laboratoire d'econometrie; Sores Inc., Development and Empirical Evaluation of Cross-subsidy Tests, Unpublished DOC (March 31, 1977).

Larsen, W.A. and S.J. McCleary (1970), "Exploratory Attempts to Model Point-to-Point Cross-Sectional Interstate Telephone Demand, unphblished Bell Laboratories memorandum (July).

Layard, P.R.G. and A.A. Walters, Microeconomic Theory, McGraw-Hill Book Company (1978), 498 pages.

LeBlanc, G. (1979), "Recent trends in the theory and practice of telecommunications regulation in the U.S. and public policy implications," Contract #OST 78000073, Canadian Department of Communications.

Lee, Marshall and Alain De Fontenay (1981), "B.C. - Alberta Long Distance Calling," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Leland, H., "Production Theory and the Stockmarket," The Bell Journal of Economics and Management Science, 5, 1 (Spring, 1974), pp. 125-44.

Leland, H., "Regulation of Natural Monopolies and the Fair Rate of Return," The Bell Journal of Economics and Management Science, 5, 1 (Spring, 1974), pp. 3-15.

Leland, H. and R. Meyer, "Monopoly Pricing Structures with Imperfect Discrimination," The Bell Journal of Economics, Vol. 7, No. 2 (Autumn, 1976), pp. 449-462.

Lesser, B. and L. Osberg, The Importance of Telecommunications to Regional Economic Development, Phase I Report, Government Studies Programme, Dalhousie University (1978).

Lesser, Barry and Osberg Lars, "Telecommunications and the Location of Employment: Implications for Regional Development Policy in Canada," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Leunbach, G. (1958), "Factors Influencing the Demand for Telephone Service," paper presented at the Second International Tele-traffic Congress, The Hague (July 7 - 11).

Linder, S.B. (1970), The Harried Leisure Class, Columbia University Press, New York.

Lintner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," Review of Economics and Statistics, Vol. 47, No. 1 (1965), pp. 13-37.

Little, A.D., Report on Interconnect Competition, Cambridge, U.S.A. (1980).

Littlechild, S.C., "A Game-theoretic Approach to Public Utility Pricing," Western Econ. Journal, Vol. 8, No. 2 (1970), pp. 162-166.

Littlechild, S.C., "Marginal-cost Pricing with Joint Costs," Econ. Journal, Vol. 80 (1970), pp. 323-335.

Littlechild, S.C., "Peak-Load Pricing of Telephone Calls," Bell Journal of Economics, Vol. 1 (Autumn, 1970).

Littlechild, S.C., "Two-Part Tariffs and Consumption Externalities," The Bell Journal of Economics, Vol. 6, No. 2 (Autumn, 1975), pp. 661-670.

Littlechild, S.C. and J.J. Rousseau, "Pricing Policy of a U.S. Telephone Company," Journal of Public Economics, Vol. 4 (February, 1975).

Loehman, E. and A. Whinston, "A New Theory of Pricing and Decision-making for Public Investment," The Bell Journal of Economics, Vol. 1, No. 2 (1971), pp. 606-625.

Loehman, E. and A. Whinston, "An Axiomatic Approach to Cost Allocation for Public Investment," Public Finance Quarterly, Vol. 2, No. 2 (1974), pp. 236-251.

Lucas, R.E., "Adjustment Costs and the Theory of Supply," Journal of Political Economy (1967).

Lydall, H., "The Distribution of Employment Incomes," Econometrica, Vol. 27 (1959).

Machlup, F. and M. Taber, "Bilateral Monopoly and Vertical Integration," 27, Economica, new series, #106, 101 (1960).

Mahan, G.P. (1979), The Demand for Residential Telephone Service, East Lansing, Michigan, Michigan State University Public Utilities Papers.

Maital, S. (1973), "Public goods and income distribution: some further results," Econometrica, 41, pp. 561-568.

Mantell, L.H., "An Econometric Study of Returns to Scale in the Bell System," Staff Research Paper, Office of Telecommunications Policy, Executive Office of the President, Washington, D.C. (February, 1974).

Martin, F., Regional Aspects of the Evolution of Canadian Employment, Economic Council of Canada, Ottawa (1976).

May, J.D. and M. Denny, "Post-War Productivity in Canadian Manufacturing," Canadian Journal of Economics, XII (1979), pp. 29-41.

McFadden, D., "Cost, Revenue and Profit Functions," pp. 3-109 in M. Fuss and D. McFadden (eds.), Production Economics: A Dual Approach to Theory and Applications, North-Holland, Amsterdam (1978).

McFadden, D. and K. Train (1979), "Specification Manual for Business Telecommunications Demand Model," unpublished paper, AT & T (March).

McKay, D.J., "Two Essays on the Economics of Electricity Supply," PhD Thesis, California Institute of Technology, 1978.

Mehta, D.R., "Impact of Financial Policies on the Valuation of the Firm: A Control Theoretic Approach," paper presented at the XXII International Meeting, The Institute of Management Sciences, Kyoto, Japan, in the session on Applied Optimal Control (1975).

Meier, R., A Communications Theory of Urban Growth, Cambridge: M.I.T. Press (1962).

Mera, K. (1969), "Experimental determination of relative marginal utilities," Quarterly Journal of Economics, 83, pp. 464-477.

Merton, R.C., "The Theory of Rational Option Pricing," The Bell Journal of Economics and Management Science, 4, 1 (Spring, 1973), pp. 141-83.

Meyer, C.W., "Marginal Cost Pricing of Local Telephone Service," Land Economics, Vol. XLII (1966).

Meyer, R.A., "Capital Structure and the Behavior of the Regulated Firm Under Uncertainty," Southern Economic Journal, Vol. 42, No. 4 (1976), pp. 600-09.

Meyer, R.A., "Regulated Monopoly Under Uncertainty," Southern Economic Journal, 45, 4 (April, 1979), pp. 1121-29.

Meyer, R.A., "Risk-Efficient Monopoly Pricing for the Multiproduct Firm," Quarterly Journal of Economics, 90, 3 (August, 1976), pp. 461-74.

Meyer, J.R., R.W. Wilson, M.A. Baughcum, E. Burton and L. Caouette, The Economics of Competition in the Telecommunications Industry, Charles River Associates Incorporated, Boston, Mass. (August, 1979).

Miller, M. and F. Modigliani, "Some Estimates of the Cost of Capital to the Electric Utility Industry, 1954 - 57," American Economic Review, 56, 3 (June, 1966), pp. 333-91.

Miller, M.H. and M. Scholes, "Dividends and Taxes," Journal of Financial Economics (1978).

Miller, M.H. and Modigliani, F., "Dividend Policy, Growth and the Valuation of Shares," Journal of Business, Vol. 34, No. 4 (October, 1961), pp. 411-33.

Mirman, L.J. and D. Sibley, "Optimal Nonlinear Prices for Multiproduct Monopolies," The Bell Journal of Economics, Vol. 11, No. 2 (Autumn, 1980), pp. 659-670.

Mitchell, B. (1981), "Local Telephone Costs and the Design of Rate Structures," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Mitchell, B.M., Economic Issues in Usage-Sensitive Pricing, The Rand Corporation, P-6530-NSF (September, 1980).

Mitchell, B.M., "Optimal Pricing of Local Telephone Service," American Economic Review, Vol. 68, No. 4 (September, 1978), pp. 517-537.

Modigliani, F. and M. Miller, "Corporate Income Taxes and the Cost of Capital - A Correction," American Economic Review, 53, 3 (June, 1963), pp. 433-43.

Modigliani, F. and M. Miller, "The Cost of Capital, Corporate Finance and the Theory of Investment," American Economic Review, 48, 3 (June, 1958), pp. 261-97.

Mohring, H., "The Peak Load Problem with Increasing Returns and Pricing Constraints," American Economic Review, Vol. 60, No. 4 (September, 1970), pp. 693-705.

Mossin, J., "Equilibrium in a Capital Asset Market," Econometrica, Vol. 34, No. 4 (1966), pp. 768-83.

Müller, Jürgen, "Cost Based Tariffs, Integrated Network Use and Network Competition in the Telecommunication Sector," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Myers, S.C., "A Simple Model of Firm Behavior Under Regulation and Uncertainty," The Bell Journal of Economics and Management Science, 4, 1 (Spring, 1973), pp. 304-15.

Myers, S.C. and S. Turbull, "Capital Budgeting and the Capital Asset Pricing Model - Good News and Bad News," Journal of Finance, 32, 2 (May, 1977), pp. 321-32.

Nadiri, M.I. and M. Shankerman, "The Structure of Production, Technological Change and the Rate of Growth of Total Factor Productivity in the Bell System" (1979), paper presented at the N.S.F. Conference on Productivity in Regulated Industries, University of Wisconsin, April 30 - May 1, 1979; published in T. Cowing and R. Stevenson (eds.) Productivity Measurement in Regulated Industries, New York: Academic Press, 1981.

Nadiri, M.I. and M.A. Schankerman, "The Structure of Production, Technological Change and the Rate of Growth of Total Factor Productivity in the Bell System," New York University, National Bureau of Economic Research (1979).

Nadiri, M.I. and Mark A. Schankerman, "The Structure of Production, Technological Change and the Rate of Growth of Total Factor Productivity in the Bell System," forthcoming in Productivity Measurement in Regulated Industries, T. Cowing and R. Stevenson, eds., Academic Press (1980).

Nelson, James R., "The Role of Competition in the Regulated Industries," Anti-Trust Bulletin, January - April, 1966, reprinted in The Journal of Reprints for Anti-Trust Law and Economics (Winter-Spring, 1973-74).

Nelson, Richard W., "Domestic Satellites, the FCC, and Competition in Domestic Telecommunications," Land Economics, WI 3 (August, 1975).

Nelson, Richard, "Domestic Satellites, The FCC, and Competition in Domestic Telecommunications," 3, Land Economics, 235 (1975).

Ng, Y. and M. Weisser, "Optimal Pricing with a Budget Constraint - The Case of the Two-Part Tariff," Review of Economic Studies, Vol. 41 (July, 1974), pp. 337-345.

Nilles, J., et al., The Telecommunications-Transportation Trade-Off, Baltimore: John Wiley (1976).

Norwood, Frank W., Telecommunications Programs Affecting Network Development, American Library Association, Chicago, Ill. (September, 1970).

Ohta, M., "A Note on the Duality Between Production and Cost Functions: Rate of Returns to Scale and Rate of Technical Progress," Economic Studies Quarterly, 25 (1974), pp. 63-65.

Oi, W.Y., "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," Quarterly Journal of Economics, Vol. 85 (February, 1971), pp. 77-96.

Ordover, J. and J. Panzar, "On the Nonexistence of Pareto Superior Outlay Schedules," Bell Journal of Economics, Vol. 11, No. 1 (Spring, 1980), pp. 351-354.

Owen, B.M. and R. Braeutigam, Regulation Game: Strategic Use of Administrative Process, Cambridge, Mass. (1978).

Panzar, C.C., "Economies of Scale and Economies of Scope in Multi-Output Production," Bell Laboratories Economic Discussion Paper No. 33 (August, 1975).

Panzar, C.C. and R.D. Willig (1977), "Economies of scale in multi-output production," Quarterly Journal of Economics, 91, pp. 481-494.

Panzar, J.C. and E. Bailey, A Theory of Contestable Markets, paper prepared for the Conference on Regulatory Problems, Duke University (1980).

Panzar, J.C. and R.D. Willig, "Economies of Scope, Product Specific Economies of Scale, and the Multiproduct Competitive Firm," Bell Laboratories Economics Discussion Paper #152 (August, 1979).

Panzar, J.C. and R.D. Willig, "Economies of Scope, Product Specific Economies of Scale, and the Multi-Product Competitive Firm," Bell Laboratories, 1978. Revised in 1979.

Panzar, J.C. and R.D. Willig, "Free Entry and the Sustainability of Natural Monopoly," The Bell Journal of Economics, Vol. 8, No. 1 (1977), pp. 1-22.

Pavarini, C. (1975), "Identifying Normal and Price-Stimulated Usage Variations of Groups of Customers Part II: Individual Customer Comparison," unpublished Bell Laboratories memorandum (October).

Pavarini, C. (1976a), "Residence and Business Flat Rate Local Telephone Usage," unpublished Bell Laboratories memorandum (July).

Pavarini, C. (1976b), "The Effect of Flat-to-Measured Rate Conversions on the Demand for Local Telephone Usage," unpublished Bell Laboratories memorandum (September).

Pavarini, C. (1979), "The Effect of Flat-to-Measured Rate Conversions on Local Telephone Usage," paper prepared for the Mountain Bell Economics Seminar, Keystone, Colorado, August 27-30, 1978, published in Pricing in Regulated Industries, Vol. II, J.T. Wenders, ed. Denver: Mountain States Telephone and Telegraph Co.

Pavarini, C. and W.J. Infosino, "Class of Service Choice Among Residence Telephone Customers," forthcoming.

Peles, Y. and J.L. Stein, "The Effect of Rate-of-Return Regulation is Highly Sensitive to the Nature of Uncertainty," American Economic Review, 66, 3 (June, 1976), pp. 278-89.

Perl, L.J. (1978), "Economic and Demographic Determinants of Residential Demand for Basic Telephone Service," National Economic Research Associates, Inc. (March).

Perrakis, S. and J. Zerbinis, "An Empirical Analysis of Monopoly Regulation Under Uncertainty," Applied Economics, forthcoming.

Perrakis, S., "Factor Price Uncertainty with Variable Proportions: A Note," American Economic Review, forthcoming (December, 1980).

Perrakis, S., "On the Regulated Price-Setting Monopoly Firm with a Random Demand Curve," American Economic Review, 66, 3 (June, 1976), pp. 410-16.

Perrakis, S., "On the Technological Implications of the Spanning Theorem," Canadian Journal of Economics, 12, 3 (August, 1979), pp. 501-11.

Perrakis, S., "Rate-of-Return Regulation of a Monopoly Firm with Random Demand," International Economic Review, 17, 1 (February, 1976), pp. 149-62.

Perrakis, Stylianos, "The Value of the Firm Under Regulation and the Theory of the Firm Under Uncertainty: An Integrated Approach," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Pestieau, P. (1975), "Regressiveness of efficiency rates in public economics," Canadian Journal of Economics, 8, pp. 269-275.

Phillips, Charles F., "Domestic Telecommunications Policy: An Overview," Washington & Lee Law Review, Vol. XXIX (Fall, 1972).

Pool, I. de Sola (ed.), The Social Impact of the Telephone, Cambridge: M.I.T. Press (1977).

Porat, M.U., "Global Implications of the Information Society," Journal of Communications, 28 (1978).

Posner, Richard A., "Theories of Economic Regulation," Bell Journal of Economics & Management Science (Autumn, 1974).

Posner, Taxation by Regulation, Bell Journal of Law and Economics (1976).

Pousette, T. (1976), "The Demand for Telephones and Telephone Services in Sweden," presented at the European Meetings of the Econometric Society, Helsinki, Finland (August).

Powell, A.A., T. Van Hoe and R.H. Wilson, "A multi-sectoral analysis of consumer demand in the post-war period," Southern Economic Journal, 35, (1968), pp. 109-120.

Pye, R., "Office Location and the Cost of Maintaining Contact," Environment and Planning A., 9 (1977).

Quirk, J. and R. Saposnik, Introduction to General Equilibrium Theory and Welfare Economics, New York (1968).

Ramsey, F.P., "A Contribution to the Theory of Taxation," Economic Journal, (March, 1927), pp. 47-61.

Rand, G.K., "Methodological Choices in Depot Location Studies," Operations Research Quarterly, 27 (1976).

Rappaport, R.N., Information for Decision-Making: Quantitative and Behavioral Dimensions, New York: Prentice-Hall (1975).

Rash, I.M. (1972), "Residence Sector Demand for Message Toll Service," Bell Canada Working Paper, Ottawa, Ontario (April).

Rault, C., "Etude econometrique de la possession d'un ensemble de biens durables de consommation," Annales de l'INSEE n° 1.

Raynor, Richard A., "Competition Within a Regulated Environment," Public Utilities Fortnightly (July 4, 1974).

Rea, J.D. and G.M. Lage (1978), "Estimates of Demand Elasticities for International Telecommunications Services," Journal of Industrial Economics, 26 (June): pp. 363-381.

Reich, R., "Warring Critizues of Regulation," Regulation (Jan./Feb., 1979).

Reimeringer, J.N., "Global Factor Productivity (GFP) and EDF's Management," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Reimeringer, J.N., "La Productivite Globale des Facteurs a Electricite de France, Principe, Resultats et Utilization," Electricite de France, Etudes Economiques Generales (November, 1980).

Reitman, C.F. (1977), "The Measurement and Analysis of Demand Elasticities of Residential Verical Telephone Service," unpublished Ph.D. Dissertation, New School for Social Research (May).

Rheaume, Gilles C., "Welfare Optimal Subsidy-Free Prices Under a Regulated Monopoly," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Robichek, A., "Regulation and Modern Finance Theory," Journal of Finance, 33, 3 (June, 1978), pp. 693-705.

Rohlfs, J. (1974), "A Theory of Independent Demand for a Communications Service," Bell Journal of Economics and Management Science, 5 (Spring): pp. 16-37.

Rohlfs, J., "Economically efficient Bell system pricing," Economic Discussion Paper #138, Bell Laboratories (1979).

Rohlfs, J.H., A.R. Goldstein and E.P. Marfisi, "Economic Evaluation of U.S. Telecommunications Policy Proposals," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Rohliffs, J., "A Theory of Interdependent Demand for a Communication Service," Bell Journal of Economics, 6 (1975).

Ross, S.A., "A Simple Approach to the Valuation of Risky Streams," Journal of Business, 51, 3 (July, 1978), pp. 453-76.

Ross, S.A., The Arbitrage Theory of Capital Asset Pricing," Journal of Economic Theory, 13 (December, 1976), pp. 341-60.

Ross, S.A., "The Economic Theory of Agency: The Principal's Problem," American Economic Review, Vol. 63 (1973), pp. 134-39.

Ross, S.A., "Options and Efficiency," Quarterly Journal of Economics, 90, 1 (February, 1976), pp. 75-89.

Ross, T.L. and R.J. Bullock, "Integrating Measurement of Productivity into a Standard Cost System," Financial Executive (October, 1980), pp. 34-40.

Rosse, James N., "The Sustainability of Natural Monopoly," Draft, Stanford University.

Rostow, Eugene V., "The Great Telephone Debate," Telephony, Part I, June 6, 1977), Part II (June 20, 1977).

Rothenberg, Jerome, The Measurement of Social Welfare, Prentice-Hall Inc. (1961), 357 pages.

Roy, R., "La Distribution du revenue entre les divers biens," Econometrica, 15 (1947), pp. 205-225.

Royal Commission on Financial Management and Accountability, Report, Ottawa (1979).

Rubinstein, M.E., "A Mean-Variance Synthesis of Corporate Financial Theory," Journal of Finance, Vol. 28, No. 1 (1973), pp. 167-81.

Salter, Liora, "Limitations of Conventional Approaches to Regulation: The Problem of Social Regulation in Rate and Service Based Determination," paper prepared for a DOC Conference; Montreal (March 4 - 6, 1981).

Samuelson, P.A., "Prices of Factors and Goods in General Equilibrium," Review of Economic Studies, 21 (1953), pp. 1-20.

Samuelson, P.A. and S. Swamy, "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," American Economic Review, 64 (1974), pp. 566-93.

Sappington, D., "Strategic Firm Behaviour Under a Dynamic Regulatory Adjustment Process," The Bell Journal of Economics, 11(1) (1980), pp. 360-372.

Sato, K., "Additive utility functions with double-log consumer demand functions," Journal of Political Economy, 80 (1972), pp. 102-124.

Sato, R. and T. Koizumi, "Substitutability, Complementarity and the Theory of Derived Demand," The Review of Economic Studies, 37 (1970), pp. 107-118.

Sauve, Jeanne, Satellites are for People, Notes for an Address to the Royal Society of Canada, Department of Communications (November, 1977).

Scala, B., "Telecommunications: Issues Abound, Arguments Mount Up," Administrative Management, Vol. 37, No. 5 (May, 1977).

Schnee, V. and W.J. Gorkiewicz (eds.), "The Communications Act of 1978: The Larger Issues," Probe: Telecommunications Journal (Summer, 1978).

Schnee, V. and W.J. Gorkiewicz, The Future of AT & T, Probe Research Inc., Millburn, N.J. (1976).

Schnee, V. and W.J. Gorkiewicz (eds.), "Probe Special Report: The Execunet Blockbuster," Probe: Telecommunications Journal (August - October, 1977).

Schwartz, Herman, "Comsat, The Carriers, and the Earth Stations: Some Problems with 'Melding Variegated Interests'," 76, Yale Law Journal, 441 (1976).

Selwyn, Lee L., "Pricing Telephone Terminal Equipment Under Competition," Public Utilities Fortnightly (December 8, 1977).

Shapley, L.S., "Cores of Convex Games," International Review of Game Theory (1971), pp. 11-26.

Sharpe, W.F., "Capital Asset Prices: A Theory of Equilibrium Under Conditions of Risk," Journal of Finance, Vol. 19, No. 3 (1964), pp. 425-42.

Shavell, S., "Risk Sharing and Incentives in the Principal Agent Relationship," Bell Journal of Economics, Vol. 10 (1979), pp. 74-91.

Shephard, D., "Cost and production functions," Princeton, Princeton University Press (1953).

Shephard, R.W., Cost and Production Functions, Princeton University Press, Princeton, N.J. (1953).

Shephard, R.W., Theory of Cost and Production Functions, Princeton University Press, Princeton, N.J. (1970).

Shepherd, William G., "Entry As a Substitute for Regulation," AER (May, 1973).

Short, J., E. Williams and B. Christie, The Social Psychology of Telecommunications, London: Wiley (1976).

Shrieves, R.F., "Market Structure and Innovation: A New Perspective," The Journal of Industrial Economics (June, 1978).

Smith, C.W., Jr., "Option Pricing: A Review," Journal of Financial Economics, 3, 1/2 (January - March, 1976), pp. 3-52.

Smith, J.B. and V. Corbo, "Economies of Scale and Economies of Scope of Bell Canada," Institute of Applied Economic Research, Concordia University, Final Report to the Department of Communications (March, 1979).

Solow, R.M., "Technical Change and the Aggregate Production Function," Review of Economics and Statistics, 39, No. 3 (August, 1957), pp. 312-20.

Spence, A.M., "Multi-Product Quantity Dependent Prices and Profitability Constraints," Review of Economic Studies, Vol. 47 (October, 1980), pp. 821-842.

Spence, A.M., "Nonlinear Prices and Welfare," Journal of Public Economics, Vol. 8, No. 1 (1977), pp. 1-18.

Spence, M. and R. Zeckhauser, "Insurance Information and Individual Action," American Economic Review, Vol. 61 (1971), pp. 350-87.

Squire, L. (1973), "Some Aspects of Optimal Pricing for Telecommunications," Bell Journal of Economics and Management Science, 4 (Autumn): pp. 515-525.

Stanford Research Institute, Analysis of Issues and Findings in FCC Docket 20003. Study prepared for AT & T (1976). (Summary and conclusions also found in Congress, Subcommittee on Communications, Domestic Telecommunications Common Carrier Policies, Part 1, March 21 - 22, 1977, Committee Serial No. 95-42.)

Steffy, W., T. Zearley and J. Strunk, "Financial Ratio Analysis, An Effective Management Tool," Industrial Development Division, Institute of Science and Technology, The University of Michigan, Ann Arbor, Michigan (1974).

Steiner, Peter O., "Peak Loads and Efficient Pricing," Quarterly Journal of Economics, Vol. 71 (November, 1957).

Stigler, G.F. and M.F. Cohen, Can Regulatory Agencies Protect Consumers? Am. Enterprise Institute for Public Policy Research (1971).

Stiglitz, J.E., "A Re-Examination of the Modigliani and Miller Theorem," Vol. 49, No. 5 (1969), pp. 784-93.

Stiglitz, J.E., "Taxation, Corporate Financial Policy and the Cost of Capital," Journal of Political Economics (1973).

Stone, R.F., et al., Final Report: Further Analysis of Docket 20003, Technology and Economics Inc. (December, 1977).

Stone, R.F., et al., Selective Competition in the Telephone Industry - An Independent Appraisal Based on Responses to FCC Docket 20003, Technology and Economics Inc. (T & E, Inc.) (1976).

Strick, John, "An Overview of the Analysis of Competition vs. Monopoly in Telecommunications Services," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Stuntebeck, S. (1976), "The Relation of Day/Night Price Differentials and Message Volume Distributions for Intrastate Toll Calls," unpublished Bell Laboratories memorandum (October).

Subisatti, E. (1973), "An Econometric Model for WATS and Business Toll Services, Computer Communications and Network Services - Bell Canada," Ottawa, Ontario (August).

Swan, Peter, "Market Structure and Technological Progress. The Influence of Monopoly on Product Innovations," QJE (November, 1970).

Taylor, L.D. (1975), "The Demand for Electricity: A Survey," Bell Journal of Economics, 6 (Spring): pp. 74-110.

Taylor, L. D. (1981), "Problems and Issues in Modeling Telecommunications Demand," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Taylor, Lester D., Telecommunications Demand: A Survey and Critique, Ballinger Pub. Co., Cambridge, Mass. (1980).

Taylor, L.D., G. Blattenberger and R. Rennhack (1981), "The Residential Demand for Energy in the United States," Report to Electric Power Research Institute on RP1098, Data Resources, Inc., Lexington, Massachusetts.

Telesat Canada, Proposed Agreement with Trans-Canada Telephone System, 3 C.R.T. 265 (1977). Also referred to as Telecom Decision CRTC 77-10.

Theil, H., "Economics and Information Theory," Amsterdam: North Holland (1967).

Theil, H., "The Information Approach to Demand Analysis," Econometrica, Vol. 33, No. 1 (January, 1965), pp. 67-87.

Thorngren, B., "Communication Studies for Government Office Dispersal in Sweden," in M. Bannon (ed.), Office Location and Regional Development, Dublin: An Foras Forbatha (1973).

Thorngren, B., "How Do Contact Systems Affect Regional Development?" Environment and Planning, 2 (1970).

Tilanus, C.B. (ed.), Quantitative Methods in Budgeting, Martinus Nijhoff, Leiden (1976).

Tornqvist, L., "The Bank of Finland's Consumption Price Index," Bank of Finland Monthly Bulletin, No. 10 (1936), pp. 1-8.

Tornqvist, G.E., Contact Systems and Regional Development, Lund Studies in Geography, No. 35 (1970).

Treadway, A.B., "Adjustment Costs and Variable Inputs in the Theory of the Competitive Firm," Journal of Economic Theory (1970).

Trebilcock, M.J. and J. Robert S. Prichard, "Crown Corporations: The Calculus of Instrument Choice".

Trebilcock, Waverman and Prichard, "Markets for Regulation," in Ontario Economic Council, Government Regulation: Issues and Alternatives (1978), p. 11.

Tupper, Allan, "The Nation's Business: Canadian Concepts of Public Enterprise," Doolorate Thesis submitted to Department of Political Studies, Queen's University, Kingston (1977).

U.S. Court of Appeals, Hush-a-Phone v U.S., 238 F 2d 266 (1956).

Usher, D., "The Suitability of the Divisia Index for the Measurement of Economic Aggregates," Review of Income and Wealth, 20 (1974), pp. 273-88.

Vickrey, William, "The City as a Firm," in The Economics of Public Services, M.S. Feldstein and R.P. Inman, (eds.), London: Macmillan (1977).

Vickrey, William (1981), "Local Telephone Costs and the Design of Rate Structures: An Innovative View".

Vickrey, William, "Responsive Pricing of Public Utility Services," Bell Journal of Economics and Management Science, Vol. 2, No. 1 (Spring, 1971), pp. 337-346.

Vilmin, E. and N. Curien, "La consommation telephonique: previsions a l'horizon 85," Note Direction Generale des Télecommunications (Juin, 1979).

Vilmin, E. and N. Curien (1981), "Demande et Consommation Telephoniques: Un Modele Residentiel Global," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Vinod, H.D., "Application of New Ridge Regression Methods to a Study of Bell System Scale Economies," Journal of the American Statistical Association (December, 1976).

Von Rabenau, B. and K. Stahl, "Dynamic Aspects of Public Goods: A Further Analysis of the Telephone System," Bell Journal of Economics, 5 (1974).

Vogelsang, I. and J. Finsinger, "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," The Bell Journal of Economics, 10(1) (1979), pp. 157-171.

Wang, R.C. (1976), "Demand for Yellow Pages Advertising Space," unpublished paper, Mountain Bell (October).

Warskett, G. and Alain De Fontenay (1981), "The Regulatory Process Under Partial Information," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Waverman, L. (1974), "Demand for Telephone Services in Great Britain, Canada and Sweden," paper presented at the Birmingham International Conference in Telecommunications Economics, Birmingham, England (May).

Waverman, L., "The Regulation of Intercity Telecommunications," Chapter 7 in A. Phillips (ed.) Promoting Competition in Regulated Markets, Brookings Institution, Washington, D.C. (1976).

Weiss, L.W., "Antitrust in the Electric Power Industry," in Promoting Competition in Regulated Markets, A. Phillips, ed., Washington, D.C.: Brookings Institution (1975).

Welch, J.B., "Workable Alternatives to Regulation," Public Utilities Fortnightly (October 23, 1975).

Wenders, John T., "Direct Testimony Before the Arizona Corporation Commission," Docket No. 9981-E-1051.

Werner, M., "Productivity Based Planning Model for Teleglobe Canada," paper given at the Ninth International Teletraffic Congress, Spain (October, 1979).

Werner, M., "Productivity Based Planning Model for Teleglobe Canada," Proceedings of the International Telecommunications Conference (1979).

P.32

Werner, M., "Teleglobe Canada Productivity Study," Teleglobe Canada, Internal Report (1977).

Wert, G.M. (1976), "The Effects of Day/Night Price Ratios on Residence Weekday Interstate MTS Traffic," unpublished Bell Laboratories memorandum (June).

Westfield, F., "Methodology of Evaluating Economic Regulation, "American Economic Review (May, 1971).

Westfield, F., "Regulation and Conspiracy," American Economic Review (June, 1965).

Westfield, F., "Vertical Integration: Does Product Price Rise or Fall?" American Economic Review (June, 1981), forthcoming.

Westfield, Fred M., "The Regulation of Telecommunication and Vertical Industry Structure," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Wiley, Richard F., "The Consumer Communications Reform Act: The FCC Position on the Proposed Legislation," Telecommunications Policy, Vol. 1, #2 (March, 1977).

Wilkinson, G.F. (1981), "The Estimation of Usage Repression Under Local Measured Service Empirical Evidence From the GTE Experiment," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Williams, J.B., The Theory of Investment Value, Cambridge, Massachusetts: Harvard University Press (1938).

Williamson, O., "Peak Load Pricing and Optimal Capacity with Indivisible Constraints," American Economic Review, Vol. 56, No. 4 (September, 1966), pp. 810-827.

Willig, R.D., "Consumer Equity and Local Measured Service," in Baude et al. (1979).

Willig, R.D., "Consumers' Surplus without Apology," American Economic Review, Vol. 66, No. 4 (September, 1976), pp. 589-597.

Willig, R.D., "Multi-Product Technology and Market Structure," The American Economic Review, Vol. 69, No. 2 (May, 1979), pp. 346-51.

Willig, R.D., "Pareto-Superior Nonlinear Outlay Schedules," Bell Journal of Economics, Vol. 9, No. 1 (Spring, 1978), pp. 56-69.

Willig, R.D. and Elizabeth E. Bailey, "Ramsey-Optimal Pricing of Long Distance Telephone Services," Pricing in Regulated Industries: Theory and Application, John T. Wenders (ed.). Paper presented at an Economic Seminar sponsored by The Mountain States Telephone and Telegraph Company (January, 1977).

Wilson, R., "On the Theory of Syndicates," Econometrica, Vol. 36, No. 1 (1968), pp. 119-32.

Wilson, R., "The Structure of Incentives for Decentralization Under Uncertainty," La Decision, Editions Du Centre National De Le Recherche Scientifique, Paris (1969).

Wohlstetter, Charles, "Government Meddling Can Ruin America's Telephone Industry," Telephony (July 16, 1973).

Wong, T.F. (1981), "Identifying Tariff-Induced Shifts in the Subscriber Distribution of Local Telephone Usage," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Yale, Janet, "Telesat Canada's Membership in Trans-Canada Telephone System: A Critique," paper prepared for a DOC Conference, Montreal (March 4 - 6, 1981).

Zellner, A., "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, Vol. 57 (June, 1962), pp. 348-68.

Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 58 (December, 1962), pp. 977-92.

Zajac, Edward E., Fairness or Efficiency: An Introduction to Public Utility Pricing, Ballinger Pub. Co., Cambridge, Mass. (1978).

Zajac, E.E., "Note on an Extension of the Ramsey Inverse Elasticity of Demand Pricing or Taxation Formula," Bell Laboratories Economics Discussion Paper No. 3 (January, 1974).

Zajac, E.E., Some Preliminary Thoughts on Subsidization, Conf. on Communication Policy Res., Office of Telecommunication Policy, Washington (1972).