Volume Title: Studies in Public Regulation

Volume Author/Editor: Gary Fromm, ed.

Volume Publisher: The MIT Press

Volume ISBN: 0-262-06074-4

Volume URL: http://www.nber.org/books/from81-1

Publication Date: 1981

Chapter Title: Regulation and the Multiproduct Firm: The Case of Telecommunications in Canada

Chapter Author: Melvyn A. Fuss, Leonard Waverman

Chapter URL: http://www.nber.org/chapters/c11434

Chapter pages in book: (p. 277 - 328)

6

Regulation and the Multiproduct Firm: The Case of Telecommunications in Canada

Melvyn Fuss
Leonard Waverman

We report here the results of a theoretical and empirical investigation of the problems of public regulation when a regulated firm produces more than one output using common production facilities. Measurement and interpretation of economies of scale become complex, especially if one wants to attribute any existing scale economies to a particular product.[1] The problem of efficient pricing in the presence of joint or common costs is also an issue of concern. These problems, and others, require consideration of the production technology and of the costs faced by the firm. The natural vehicle for such an analysis is the multiproduct cost function, since its arguments are outputs and input prices. Recent advances in the econometric literature (Diewert 1971) have made possible the use of cost functions to represent general structures of technology. The advent of multiproduct generalized cost functions could provide an important econometric supplement to cost-separation studies in regulatory hearings.[2]

We discuss theoretical problems associated with analyzing the production technology of a multiproduct firm (aggregation of output, economies of scale, economies of scope, cost separation), give a detailed econometric specification, and develop a constrained profit-maximizing model of a regulated telecommunciations firm in which the level of local service output is chosen by the regulators rather than by the firm. We apply this model to data generated by Bell Canada during the period 1952–1975.

Rate-of-return regulation creates potential difficulties for estimation of the production technology. If such regulation is effective, estimates of parameters of the production technology that ignore this fact may be biased. We demonstrate how the theory of duality between cost and production can be used to specify a multiproduct cost function and associated derived demand functions which explicitly incorporate effects of rate-of-return regulation.

The Separation of Common and Joint Costs

When firms produce two or more outputs utilizing common or joint inputs, two problems arise: the allocation of these common or joint costs to the separate outputs, and the measurement of aggregate output for the firm.

Common costs are defined as the costs of common inputs utilized by two or more outputs, so that the multiproduct transformation function is represented by

$$F(Y_1,\ldots,Y_m;X_1,\ldots,X_n) = 0, \tag{1}$$

where Y_i ($i = 1,\ldots,m$) are outputs and X_j ($j = 1,\ldots,n$) are inputs. Were costs not common, (1) could be rewritten as

$$\begin{aligned} Y_1 &= F_1(X_1,\ldots,X_d), \\ Y_2 &= F_2(X_{d+1},\ldots,X_h), \\ &\ldots \\ Y_m &= F_m(X_{k+1},\ldots,X_n), \end{aligned} \tag{2}$$

where $j = 1,\ldots,d,\ldots,h,\ldots,k,\ldots,n$.

Joint costs as they are defined in the regulation literature occur when two or more outputs are produced in fixed proportions; it is impossible to produce some proportion λ of 1 without also producing λ of the others (Kahn 1971). The product transformation curve can then be represented as

$$F(X_1,\ldots,X_n) = \min(a_1 Y_1, a_2 Y_2,\ldots,a_m Y_m). \tag{3}$$

Note that (3) is the limiting case of (1). In the following discussion we will treat the problems of common and joint costs as the same issue, referring instead to *joint production*, as represented by equation (1). Whether costs are "joint" or whether production is characterized by fixed coefficients can be tested empirically.

When a firm produces heterogeneous products, there is no single unique index of output. For an index $h(Y)$ to be formed, the product transformation function must be written as

$$F(Y_1,\ldots,Y_m; X_1,\ldots,X_n) = F(h(Y),X_1,\ldots,X_n). \tag{4}$$

But (4) can be rewritten as $F(h(Y),g(x)) = 0$, since the existence of an output aggregate implies the existence of an input aggregate (Brown et al. 1976; Baumol and Braunstein 1977).[3] As a result, the dual joint cost function is separable in outputs. But then the relative marginal costs of the outputs are independent of input prices—a strong assumption (see Lau 1978).

Most regulated industries involve joint production. Electric utilities produce peak and off-peak kWh utilizing common generating transmission and distribution capacity. Railroads use the same roadbed for passenger and freight traffic. Telecommunications firms provide a wide variety of services: residential local switched calls, business local switched calls, residential switched toll calls, business switched toll calls, private wire service, teletypewriter exchange service, specialized common-carrier service, and a variety of broadband data services. All switched calls utilize local exchange switching equipment in common. Toll calls utilize common interoffice switching equipment and common intercity communications equipment. Business switched and nonswitched private wire services all use common intercity plant. These are but several of the many examples of joint production in the telecommunications sector.

It is difficult to find examples of true joint costs in telecommunications except in a temporal distribution sense. An increase in the number of circuits between two points provides increased capacity which is distributed between day and night-time calls in fixed proportions. But this same increase in peak circuit availability can provide a varying proportion of business and residential calls, hence the increase in plant is common to business and residential use but joint between peak and off-peak use. In the remainder of the paper when we use the term "jointness" we mean in the production sense, encompassing both common and joint costs as used in the regulation literature.

Effective regulation poses the following fundamental questions:

- What range of services is best supplied by a single firm? What are the production economies of scale (the change in average and marginal costs when firm size is increased)?[4] What are the economies of scope (the change in average and marginal costs when services are combined within a single firm)?[5] Economies of scope exist if and only if production is joint.
- What are the long-run marginal costs of producing one more unit of any one of the joint outputs $(Y_1, \ldots, Y_m)$?

The first set of analyses help determine the size of the firm and the degree of competition to be allowed;[6] the second analysis allows the examination of the efficiency of any rate structure.

Cost Functions, Economies of Scale, and Economies of Scope

The Multiproduct Cost Function

A multiproduct production process can be represented by the product transformation function

$$F(Y_1,\ldots,Y_m;X_1,\ldots,X_n) = 0, \tag{5}$$

where Y_i $(i = 1,\ldots,m)$ are outputs and X_j $(j = 1,\ldots,n)$ are inputs.[7]

In this section we consider only the case where rate-of-return regulation either is not used by the regulatory authorities or else is ineffective. In either case, we may assume that the firm pursues cost-minimizing behavior without regard to the possibilities of Averch-Johnson-type distortions. In that case, the theory of duality between cost and production (see Diewert 1971) ensures that for every transformation function of the type shown in equation (1) there exists a dual cost function of the form

$$C = C(Y_1,\ldots,Y_m;P_1,\ldots,P_n), \tag{6}$$

where $P_j(j = 1,\ldots,n)$ are the prices paid by the firm for the inputs X_j—as long as the product transformation function satisfies the usual regularity conditions (such as convex isoquants), the firm pursues cost-minimizing behavior, and the firm has no control over input prices.

Under these assumptions, the cost function is just as basic a description of the technology as the product transformation (joint production) function, and it contains all the required information, including information on jointness.

The properties of the multiproduct cost function (6) are that C is concave in P_j, linearly homogeneous in P_j, and increasing in Y_i and P_j; that $\partial C/\partial P_j = X_j$ (Shephard's lemma);[8] and that the own-price elasticities of factor demand are given by

$$\varepsilon_{jj} = P_j \frac{\partial^2 C}{\partial P_j^2} \Big/ \frac{\partial C}{\partial P_j}.$$

Economies of Scale and Incremental Costs

The fact that average cost (cost per unit of output) is not defined for multiple-output technologies makes analysis of returns to scale somewhat complex, since we can no longer just measure the effect of output increases on the average cost of production. In addition, we need to distinguish between economies of scale in some overall sense (that is, when all outputs are increased) and economies of scale associated with the expansion of a particular output (with all other outputs held constant). We begin by considering an overall measure of returns to scale.

"Overall" returns to scale can be obtained by computing

$$dC = \sum_{i=1}^{m} \frac{\partial C}{\partial Y_i} dY_i$$

or

$$d\log C = \sum_{i=1}^{m} \frac{\partial \log C}{\partial \log Y_i} \frac{dY_i}{Y_i}. \tag{7}$$

Equation (7) represents the total change in cost resulting from differential changes in the levels of the m outputs. Unfortunately, unless we add additional structure to equation (7) it is difficult to interpret changes in cost resulting from changes in outputs in terms of returns to scale. One common procedure is to assume that all outputs are increased in proportion; that is, $dY_i/Y_i = d\log Y_i = \lambda$. Then

$$\frac{d\log C}{\lambda} = \sum_{i=1}^{m} \frac{\partial \log C}{\partial \log Y_i}. \tag{8}$$

If $d\log C/\lambda > 1$, incremental overall costs are increasing and hence production is subject to decreasing returns to scale; if $d\log C/\lambda < 1$, the technology exhibits increasing returns to scale; if $d\log C/\lambda = 1$, overall constant returns to scale exist.

This overall description of returns to scale is somewhat less relevant in the multiple-output case than in the single-output case, since it requires all outputs to be increasing in strict proportion, which may not correspond to the optimal production plan. Nevertheless, the overall-returns-to-scale number can be a useful summary statistic for comparing results obtained from the present framework with returns to scale estimated from aggregate production or cost functions (that is, functions that aggregate all outputs into a single variable).

Now consider the concept of returns to scale with respect to a single output. From equation (7),

$$\left.\frac{d\log C}{d\log Y_i}\right|_{Y_j,\, j \neq i \text{ constant}} = \frac{\partial \log C}{\partial \log Y_i}. \tag{9}$$

The term $\partial \log C/\partial \log Y_i$, the output cost elasticity, represents incremental or marginal cost in percentage terms.

It is tempting to specify that if

$$\frac{\partial \log C}{\partial \log Y_i} = \frac{Y_i}{C}\frac{\partial C}{\partial Y_i} > 1,$$

returns to scale in producing the ith output are decreasing; that if

$$\frac{\partial \log C}{\partial \log Y_i} < 1,$$

they are increasing; and that if

$$\frac{\partial \log C}{\partial \log Y_i} = 1,$$

they are constant. This specification would yield the correct trichotomy for the case of a single-output production process. However, any attempt to use this definition can lead to a conflict with the definition of overall returns to scale introduced earlier. Consider the two-output Cobb-Douglas cost function

$$C = AY_1^{\alpha_1}Y_2^{\alpha_2}. \tag{10}$$

Overall decreasing returns to scale implies

$$\sum \frac{\partial \log C}{\partial \log Y_i} = \alpha_1 + \alpha_2 > 1. \tag{11}$$

However, $\partial \log C/\partial \log Y_i = \alpha_i$ can be <1 while $\alpha_1 + \alpha_2 > 1$, which leads to a contradiction between our overall and the proposed output-specific returns to scale measures. Thus, we must reject the use of the individual-product cost elasticities as indicators of returns to scale. However, if one of the cost elasticities exceeds unity, the sum of them will also exceed unity and therefore no contradiction will arise.

Can we define a meaningful indicator of the potential advantages or disadvantages of output expansion? Since cost elasticity cannot be used, the remaining intuitive concept is that of changes in incremental cost. It would appear, at first glance, that decreasing incremental cost ($\partial^2 C/\partial Y_i^2 < 0$) should indicate increasing returns to scale. However, we will show that this marginal-cost concept also does not provide a solution to the problem of developing an unambigous indicator of scale economies. It can easily be shown that

$$\frac{\partial^2 C}{\partial Y_i^2} = \frac{C}{Y_i^2}\left[\frac{\partial^2 \log C}{\partial \log Y_i^2} + \frac{\partial \log C}{\partial \log Y_i}\left(\frac{\partial \log C}{\partial \log Y_i} - 1\right)\right]. \tag{12}$$

The usual situation is that $\partial \log C/\partial \log Y_i < 1$. The more disaggregated the output vector, the more likely it is that $\partial \log C/\partial \log Y_i < 1$. If we accept this case, then an additional sufficient condition for $\partial^2 C/\partial Y_i^2 < 0$ is that $\partial^2 \log C/\partial \log Y_i^2 < 0$; that is, that cost elasticity is a decreasing function of output. In all such cases, a highly possible outcome is that marginal costs with respect to each output will decrease ($\partial^2 C/\partial Y_i^2 < 0$) and, at the same time, overall returns to scale will also decrease ($\Sigma_j(\partial \log C/\partial \log Y_i) > 1$). In addition, for any given degrees of jointness and overall returns to scale, it is possible (for the case of decreasing cost elasticity) to increase the rate at which marginal costs decline simply by further disaggregation of the outputs. These two possibilities should be sufficient warning against using any observed decreasing marginal cost as an indicator of subproduct-specific returns to scale. In fact, there exists no unambiguous measure of output-specific returns to scale except in the case of nonjoint production,[9] since separate cost functions cannot be constructed when common costs exist. Thus, we appear to be left with only the overall measure of returns to scale.[10] Unfortunately, this measure is of no value when one is attempting to evaluate the possible efficiency gains from increasing the scale of production of one of the outputs in a multiproduct production process.

Joint Production, Economies of Scope, and Subadditivity

To this point we have assumed that the production technology is truly a multiple-output one, in the sense that it is more efficient to produce the m outputs together than by separate production processes. This efficiency condition (jointness in production) is known in the industrial organization literature as *economies of scope* (see Panzer and Willig 1975, 1979; Baumol and Braunstein 1977).

For the purposes of this paper, we will say that the production technology exhibits economies of scope if

$$C(Y_1 \cdots Y_m) < C(Y_1, 0 \cdots 0) + C(0, Y_2, 0 \cdots 0) + C(0, 0, Y_3 \cdots 0) + \cdots + C(0, 0 \cdots 0, Y_m). \tag{13}$$

Panzer and Willig (1979) have shown that a sufficient condition for a twice-differentiable multiproduct cost function to exhibit economies of scope is that it exhibit cost complementarities, defined by

$$\frac{\partial^2 C}{\partial Y_i \partial Y_j} < 0 \qquad (i \neq j;\ i,j = 1, \ldots, m). \tag{14}$$

Conditions (14) provide a means of testing for the existence of economies of scope.

In a series of articles, Baumol (1977b), Baumol, Bailey, and Willig (1977), and Panzer and Willig (1977) have established the importance of "subadditivity of the cost function" as a production characteristic in the analysis of a regulated monopolist. Unfortunately, a test for subadditivity is difficult to devise, since one of the requirements is a knowledge of the cost function in the neighborhood of zero outputs. The closest we can come is a (weak) local test in the neighborhood of the point of approximation. It can be shown that the simultaneous existence of local economies of scale and local economies of scope is sufficient to ensure local subadditivity.

Cost Separation and Econometric Cost Functions

The transcripts of regulatory hearings in the telecommunications sector are replete with discussions of how to allocate joint and common cost.[11,12] In the United States, since some of the telephone plant is under state jurisdiction and some under federal jurisdiction, it became necessary to "separate" interstate from intrastate plant. When regulators became concerned with the structure of prices rather then just their average level, costs of different services had to be separated. The push for entry by specialized common carriers (SCCs) and AT&T's competitive response prompted the F.C.C. requests for guidance from affected parties. In Canada, competitive pressures between two transnational carriers and suggestions of "cream skimming" and predatory pricing prompted a lengthy cost inquiry.

Three basic separation formulas have been proposed: embedded direct cost (EDC), long-run increamental cost (LRIC), and fully distributed cost (FDC). These formulas are needed in order to assess "correct" prices. The FDC method is essentially one of long-run average cost pricing. Two methods have been proposed to fully distributed these average costs: relative use (revenue share, for example) and historical cost causation.[13]

The calculation of average embedded direct costs is equivalent to the measurement of short-run variable costs. Unless there are no fixed costs, long-run pricing at EDC will generate losses. FDC in principle takes into account the fixed costs. However, pricing on the basis of FDC is associated with a number of problems. Where production is characterized by long-run returns to scale, setting price equal to average FDC is inefficient since, for Pareto optimality, prices *ex ante* should be set equal to long-run incremental costs. Separating common costs by some measure of relative use is arbitrary and should not be used as a basis for price setting, since there is in general no connection between intensity of use and cost causation.

Cost separation is a bogus issue that exists because of regulatory commissions' reliance on historical average costs as a guide to setting price. But there is no method of correctly separating historical average costs. Pricing rules based on efficiency criteria should be set at long-run incremental costs, thus avoiding any need to "separate" costs.

One must be careful in defining the long run, in order to be certain that incremental costs are measured in terms of changes in capacity output, not changes in actual output at less than capacity.[14] As an example, let us examine the data-telecommunications market. Carriers have argued that the incremental costs of data service are low since it is an adjunct to the large monopoly switched message service that must be provided for voice transmission. However, if the quality of the entire system must be increased to accommodate an acceptable level of reliability for data transmission, these upgrading costs are attributable to data-service users and should be fully charged to them alone as part of the incremental costs of the change in the system. This higher quality of service is provided jointly to all users, but quality-upgrading joint costs should be solely allocated to the data-service users.

Can econometric analysis of the retrospective cost functions of telecommunications firms provide useful information for regulatory purposes? In theory, given highly disaggregated data corresponding to true economic variables in the absence of strong time trends, long-run incremental costs could be allocated on a historical assigned-cause basis.[15]

Econometric cost function analysis can be used to examine the response of a firm at the margin; that is, at the replacement values or opportunity costs of inputs. Costs are minimized subject to current factor prices. Regulators, while allowing the firm to use opportunity costs for labor and materials inputs, insist that the firm earn its allowed rate of return on the historical cost of the capital stock, not its replacement cost. As a result, costs and prices as determined under a regime of historical-cost rate base will differ from the incremental costs and prices determined by an econometric cost function.[16]

In a period of inflation, historical costs will be less than replacement costs. Incremental costs determined econometrically will then exceed historical "incremental" costs. One of the purposes of cost separation is to estimate the relative contributions of the various services—that is, the excess of price over historical cost. Comparing actual prices (those set by regulators on the basis of historical cost) with incremental replacement costs is no guide in determining the extent of cross-subsidization inherent in the actual price structure (when replacement and historical costs differ).

Incremental replacement costs are, however, a guide to efficient pricing behavior. Where economies of scale are not present, setting the price for each service equal to the incremental replacement cost (as determined by an econometric cost function) yields the set of subsidy-free Pareto-optimal prices. If there are economies of scale and the firm is constrained to at least break even, calculating the set of Ramsey prices on the basis of these incremental replacement costs leads to the most efficient "constrained" set of prices, assuming neutrality in interpersonal comparisons (that is, where distributional consequences are disregarded).

Econometric Specification of a Joint Production Process

Profit-Maximizing Model for a Multioutput Regulated Firm

Estimates of demand elasticities indicate that regulated telecommunications firms operate some services in the region of inelastic demand.[17] But profit-maximizing monopolies will never produce where marginal revenue is negative; to do so would require setting marginal revenue equal to negative marginal costs. If marginal costs are positive, an unregulated profit-maximizing monopoly finding itself in the inelastic region of demand would raise price (lower output) to increase total revenue.[18] However, many regulated utilities are not able to lower output, since the regulators insist that certain basic services be offered at prices that force the monopolist to remain in the inelastic region. One example is passenger train service. Most railroad firms would wish to reduce passenger service and raise its price. However, regulators in North America do not permit rail companies to raise the price of this service to the profit-maximizing level, although they do require the provision of passenger seats.

All studies of the telephone industry suggest that local telephone service is characterized by inelastic demand. The observed level of local service and the corresponding price are chosen not by a profit-maximizing monopolist, but by the regulators. As a result, it is reasonable to assume that the observed level of local service is not an endogenous choice variable to the firm. Instead, we consider the level of local service exogenous to the firm.

The firm's problem is to maximize profits,

$$\pi = \Sigma q_i Y_i - C(Y_1,\ldots,Y_m;P_1,\ldots,P_n), \tag{15}$$

subject to the constraint on the provision of certain services:

$$Y_i \geq \bar{Y}_i \qquad (i \in H),$$

where H is the class of outputs constrained by the regulators, assumed to be the first H outputs. (Note that if demand were inelastic in the region of Y_i, the firm would never decide to offer more than $\bar{Y}_i$.[19])

Substituting into the profit expression, we obtain

$$\pi = \sum_{i \notin H} q_i Y_i + \sum_{i \in H} q_i \bar{Y}_i - C(\bar{Y}_1 \cdots \bar{Y}_H, Y_{H+1} \cdots Y_m, P_1 \cdots P_n).$$

For profit-maximizing behavior we have the first-order conditions

$$\frac{\partial \pi}{\partial Y_i} = q_i + Y_i \frac{\partial q_i}{\partial Y_i} - \frac{\partial C}{\partial Y_i} = 0 \qquad (i = H + 1, \ldots, m), \tag{16}$$

or[20]

$$MR_i = MC_i \qquad (i = H + 1, \ldots, m).$$

In addition, the second-order conditions for maximization require that[21]

$$\frac{\partial MR_i}{\partial Y_i} \leq \frac{\partial MC_i}{\partial Y_i}. \tag{17}$$

For observed output levels such that demand is elastic, we assume that marginal revenue is set equal to marginal cost. For output levels such that demand is inelastic, we assume that those outputs are exogenous to the firm.

The Multiple-Output Translog Cost Function

The translog cost function is becoming an increasingly popular specification of the functional form of a cost function (Brown et al. 1976; and Fuss 1977). The function is quadratic in logarithms and is one of the family of second-order Taylor-series approximations to an arbitrary cost function. The multiple-output translog cost function, assuming capital-augmenting technical change (input n), takes the form[22]

$$\begin{aligned}
\log C = {} & \alpha_0 + \sum_{i=1}^{m} \alpha_i \log Y_i + \sum_{j=1}^{n-1} \beta_j \log P_j + \beta_n \log P_n^* \\
& + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{k=1}^{m} \delta_{ik} \log Y_i \log Y_k + \tfrac{1}{2} \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \gamma_{jk} \log P_j \log P_k \\
& + \tfrac{1}{2} \sum_{j=1}^{n} \gamma_{jn} \log P_j \log P_n^* + \sum_{i=1}^{m} \sum_{j=1}^{n-1} \rho_{ij} \log Y_i \log P_j \\
& + \sum_{i=1}^{m} \rho_{in} \log Y_i \log P_n^*,
\end{aligned} \tag{18}$$

where α_0, α_i, β_j, δ_{ik}, γ_{jk}, ρ_{ij} are parameters to be estimated, and $P_n^* = P_n e^{-\theta t}$ where θ is the rate of decline in the price of a nominal unit of capital due to capital-augmenting technical change. (In all the following equations, the asterisk notation is dropped unless it is important for interpretation.)

Using Shephard's lemma ($\partial C/\partial P_j = X_j$), we have

$$\frac{\partial \log C}{\partial \log P_j} = \frac{P_j X_j}{C} = M_j = \beta_j + \sum_k \gamma_{jk} \log P_k + \sum_i \rho_{ij} \log Y_i, \tag{19}$$

where $j = 1,\ldots,n$ and M_j is the cost share of the jth input.

Equations (18) and (19) make up the cost system. However, a number of features of the system reduce the number of parameters to be estimated. Since M_j is a cost share,

$$\sum_{j=1}^{n} M_j = 1;$$

this implies

$$\sum_j \beta_j = 1, \qquad \sum_j \gamma_{jk} = 0, \qquad \sum_j \rho_{ij} = 0.$$

In addition, the linear homogeneity property of cost functions outlined earlier implies the further parameter restrictions

$$\sum_k \gamma_{jk} = 0.$$

Finally, the fact that the function is a second-order approximation implies

$$\delta_{ik} = \delta_{ki}, \qquad \gamma_{jk} = \gamma_{kj}.$$

The translog cost function can be used along with the profit-maximizing conditions to generate additional equations representing the optimal choice of endogenous outputs.

Taking the derivations of the cost function with respect to endogenous outputs, we have[23]

$$\frac{\partial \log C}{\partial \log Y_i} = \frac{\partial C}{\partial Y_i}\frac{Y_i}{C} = MR_i \frac{Y_i}{C} = \frac{q_i(1 + 1/\varepsilon_i)Y_i}{C}, \tag{20}$$

where ε_i is the own price elasticity for the ith output. Denoting $q_i\ Y_i/C$ as R_i, the "revenue share," we obtain

$$R_i = \frac{\partial \log C}{\partial \log Y_i}\left(1 + \frac{1}{\varepsilon_i}\right)^{-1}. \tag{21}$$

For the translog system the sum of the R_i is not constrained to be unity, since the firm is not constrained to earn zero economic profits.[24] Using the translog cost function, the system of equations (21) become

$$R_i = \left(\alpha_i + \sum_k \delta_{ik} \log Y_k + \sum_j \rho_{ij} \log P_j\right)\left(1 + \frac{1}{\varepsilon_i}\right)^{-1}, \tag{22}$$

where $i = H + 1, \ldots, m$.

Factor Price Elasticity, Incremental Costs, Overall Economies of Scale, and Economies of Scope with the Translog Cost Function

It can be shown that the own price elasticity of demand for factor j can be computed as (see Berndt and Wood 1975)

$$\varepsilon_{jj} = \frac{\partial \log X_j}{\partial \log P_j} = \frac{\gamma_{jj} - M_j + M_j^2}{M_j}. \tag{23}$$

Once γ_{jj} is estimated, the above price elasticities are determined. The incremental or marginal cost elasticity of producing output i is

$$\frac{\partial \log C}{\partial \log Y_i} = \alpha_i + \sum_k \delta_{ik} \log Y_k + \sum_j \rho_{ij} \log P_j. \tag{24}$$

The incremental cost curve for output Y_i can be obtained as

$$\mathrm{ICC}(Y_i) = \frac{C}{Y_i}\left(\alpha_i + \delta_{ii} \log Y_i + \sum_{k \neq i} \delta_{ik} \log \bar{Y}_k + \sum_j \rho_{ij} \log \bar{P}_j\right), \tag{25}$$

where $\bar{Y}_k$ and $\bar{P}_j$ are preassigned constant outputs and prices, respectively. The "overall" returns-to-scale number can be obtained from

$$\begin{aligned} \frac{d \log C}{\lambda} &= \sum_i \frac{\partial \log C}{\partial \log Y_i} \\ &= \sum_i \alpha_i + \sum_i \sum_k \delta_{ik} \log Y_k + \sum_i \sum_j \rho_{ij} \log P_j. \end{aligned} \tag{26}$$

Economies of scope or jointness in production can only be tested in the translog framework using the approximate tests discussed by Denny and Fuss (1977). The condition is that

$$\delta_{ij} = -\alpha_i \alpha_j \qquad (i \neq j),$$

when the data have been scaled so that all $Y_i = P_j = 1$ at the point of approximation.[25]

Specialized Descriptions of Technology and the Translog Cost Function

Three specialized descriptions of technology are often assumed in the estimation of production structures. The joint production function $F(Y_1,\ldots,Y_m;X_1,\ldots,X_m) = 0$ is said to have a *separable* input-output structure if it can be written in the form

$$G(Y_1,\ldots,Y_m) = H(X_1,\ldots,X_m). \tag{27}$$

It can be shown (see Denny and Pinto 1978) that the joint cost function can then be written in the form

$$C = C(h(Y_1,\ldots,Y_m),P_1,\ldots,P_n). \tag{28}$$

It is obvious from equations (27) and (28) that the test for separability is the test for the existence of an output aggregate. Following Denny and Fuss 1977, it can be shown that the separability constraints for the translog approximation used in this paper ($m = n = 3$) are

$$\begin{aligned}
\alpha_1 &= \alpha_3\frac{\rho_{11}}{\rho_{31}}, \\
\alpha_2 &= \alpha_3\frac{\rho_{21}}{\rho_{31}}, \\
\rho_{12} &= \rho_{11}\frac{\rho_{22}}{\rho_{21}}, \\
\rho_{32} &= \rho_{31}\frac{\rho_{22}}{\rho_{21}}.
\end{aligned} \tag{29}$$

A production function is said to be *homothetic* if input proportions are independent of scale. Shephard (1953) showed that the cost function for a homothetic production structure takes the form

$$C = g(P_1,\ldots,P_n)h(Y_1,\ldots,Y_m). \tag{30}$$

For the translog approximation used in this paper the homotheticity constraints are

$$\rho_{12} = \rho_{21} = \rho_{13} = \rho_{31} = \rho_{23} = \rho_{32} = 0. \tag{31}$$

Finally, the production structure will be of a Cobb-Douglas form if the joint cost function can be written as

$$C = AP_1^{\beta_1}P_2^{\beta_2}P_3^{\beta_3}Y_1^{\alpha_1}Y_2^{\alpha_2}Y_3^{\alpha_3}. \tag{32}$$

For the translog function used in this paper the Cobb-Douglas constraints are

$$\delta_{ik} = \rho_{ij} = \gamma_{jk} = 0 \qquad (i,j,k, = 1,2,3). \tag{33}$$

Data

Data pertaining to Bell Canada's operations during the period 1952–1975 were used to estimate the equations of our model. Bell Canada is the single largest telecommunications firm in Canada, serving in 1975 close to 8 million telephones in the provinces of Ontario and Quebec. For the majority of services offered, no competition is allowed. However, there is a range of so-called competitive services, such as data transmission (where competition with one other firm, Canadian National/Canadian Pacific Telecommunications, has always existed).

The basic source of all our data is a Bell Canada submission to the Canadian Radiotelevision and Telecommunication Commission entitled Response to Interrogatories of the Province of Ontario, Item 101, 12 February 1977, that presents constant-dollar revenues, skill-weighted manhours, net value of capital, and the associated prices. (This is referred to hereafter as the BCS.)

Output Data

The BCS gives both constant-dollar (1967 base) and current-dollar revenues for local services, directory advertising, and three subdivisions of message toll (intra-Bell, Trans-Canada and adjacent members, U.S. and overseas). Also included are other toll revenues and miscellaneous revenues. We used three output measures: local services, message toll services, and what we label competitive services (the remaining three services).[26] The aggregate measure of message toll output was derived as a Divisia index of the three constant-dollar subaggregates using arithmetic weights. The implicit price index for the toll aggregate service was determined from the division of current-dollar revenue by the Divisia quantity index.

An aggregate measure of the quantity of competitive services was computed in a similar fashion by calculating a Divisia quantity index of the directory advertising, other tolls, and miscellaneous services. The implicit price index was formed as above—by dividing total current-dollar revenue from the three services by the quantity index.

Input Data

The BCS lists four separate factors: cost of materials, services, rent, and supplies; indirect taxes; manhour input; and capital input. We call the first two groups *materials* and form the aggregate price index as a Divisia index (arithmetic weights) of the separate components. Table 6 of the BCS provides a series called "Labour Value, Adjusted for Quality," in constant 1967 dollars. An earlier memorandum indicates that the method of adjusting for quality consists of weighting actual man-hours in each of twenty-eight labor categories by the ratio of the average total hourly remuneration of that specific group in the base year to the average remuneration for all groups (the Kendrick method; see Olley 1970). All hours attributable to contruction are excluded, as are sick leave, vacations and holidays. This series represents the quantity of labor input for our study.

Because we could not acquire a comparable "adjusted" series for the nominal expenditure on labor, we used nominal labor expenditures as reported in Bell Canada annual reports. As these expenditures included management salaries, sick pay, vacation pay, and services not included in the quantity index, the implicit price index was slightly greater than unity in 1967. We normalized the labor price series to be unity in 1967.

Table 7 of the BCS includes a series entitled "Total Average Net Stock of Physical Capital in 1967 Values." The earlier Olley (1970) memorandum describes the process that generated this capital series. For each year from 1920 on, the age distribution of capital in place was determined in each of six categories (buildings, central office equipment, station equipment, outside plant, furniture and office equipment, and motor vehicles). Constant-dollar values were determined by reflating the physical stock by a Laspeyres price index (1967 = 1.0) for each capital type. Cash, accounts receivable, and short-term assets less short-term liabilities were excluded. Total net value of the capital stock in any year is the sum of the six individual constant-dollar categories.

The user cost of capital (P_k) for our study is calculated as

$$P_k = P_I(i + \delta),$$

where i is the expected long-run real after-tax rate of return applicable to Bell Canada (assumed to be constant at 6 percent), δ is the rate of real economic depreciation (from the BCS), and P_I is the telephone plant price index (from the BCS).

We have assumed that the expected real rate of return was constant over the period. Attempting to incorporate an increasing real rate led to implausible results. We chose 6 percent as the real rate, but also examined the effects of the alternative assumptions of 5 percent and 7 percent. Jenkins (1972, 1977) estimated the actual real after-tax rate of return for a wide cross-section of Canadian firms over the 1953–1973 period. The average rate was 6 percent. For the communications sector, the 1965–1969 average real rate was 5.2 percent including capital gains and 7 percent excluding capital gains. For the 1965–1974 period Jenkins estimated the average rate, excluding capital gains, to be 7.2 percent for the communications sector.

Empirical Results

Estimation Procedure

For both the demand functions and the augmented-cost-equations system the method of estimation was iterative three-stage least squares, an asymptotically efficient simultaneous-equations procedure. The instrumental variables used for the right-hand-side endogenous variables were formed from the exogenous variables of the two systems: local service output, input prices, and real income.

Estimates of Elasticities in Demand

We estimated log-linear demand functions of the form

$$\log Y_{it} = a_i + b_i \log(q_{it}/\mathrm{CPI}_t) + d_i \log N_t, \tag{34}$$

where Y_{it} ($i = 1,2,3$) are per capita outputs of local service, toll service, and competitive services, respectively; q_{it} ($i = 1,2,3$) are the corresponding output prices; CPI_t is the consumer price index; and N_t is real per capita disposable income in Bell Canada's operating territory.[27]

Equations (34) can be viewed as first-order approximations to arbitrary demand functions. We have also assumed that the service in question is weakly separable from all other goods and services in the individual's utility or production functions, and that the aggregate price index of these other commodities can be adequately represented by the consumer price index.

The three estimated demand functions are

$$\log Y_{1t} = \underset{(0.209)}{-0.489} - \underset{(0.264)}{0.721} \log(q_{1t}/\mathrm{CPI}_t) + \underset{(0.264)}{0.443} \log N_t;\ R^2 = 0.804,$$

$$\log Y_{2t} = \underset{(1.01)}{-1.902} - \underset{(0.451)}{1.435} \log(q_{2t}/\mathrm{CPI}_t) + \underset{(0.450)}{1.095} \log N_t;\ R^2 = 0.934,$$

$$\log Y_{3t} = -1.261 - 1.638 \log(q_{3t}/\text{CPI}_t) + 0.890 \log N_t; \; R^2 = 0.780.$$
$$\qquad\quad (0.200) \quad (0.166) \qquad\qquad\qquad (0.047)$$

(Standard errors of coefficients are given in parentheses.)

The above results confirm one of the assumptions of the previous sections: that Bell Canada operates in the inelastic region of the demand curve for local services. Therefore, we will assume in the augmented cost model estimated below that the level of the local service output (Y_1) is exogenous to the firm's profit-maximizing output choice decision.

Estimation of the Augmented Model

The system of equations consisting of the cost function (18), the cost-share equations (19), and the revenue-share equations (22) was estimated using the data described in the previous section. The estimated elasticities for toll (−1.435) and competitive services (−1.638) were used as extraneous estimates in the revenue-share equations.

Parameter estimates are given in table 6.1, along with the corresponding standard errors. Goodness-of-fit statistics are given in table 6.2. The ceofficient representing capital-augmenting technical change, θ indicates that augmenting technical change resulted in an annual decrease of 6.7 percent in the effective price of a nominal unit of capital. The rate of total cost diminution is given by $\partial \log C/\partial t = \theta M_n$, where M_n is the cost share of capital. The rate is 3.0 percent at the point of approximation (1964), and ranges from 2.3 percent in 1952 to 3.4 percent in 1975.

Table 6.3 presents the own-price elasticities of demand for the three factors for the beginning, midpoint, and terminal years, and table 6.4 gives the cross-price elasticities for the middle year. Note the fall in the capital own-price elasticity over time, the rise in the labor own-price elasticity, and the constancy of the materials own-price elasticity. These trends are due to the trends in Bell Canada's cost shares. To demonstrate this fact, we note that the own-price elasticity (ε_{ii}) is calculated as

$$\varepsilon_{ii} = \frac{\delta_{ii} + M_i^2 - M_i}{M_i}.$$

The derivative of the ith own-price elasticity with respect to share M_i is

$$\frac{\partial \varepsilon_{ii}}{\partial M_i} = 1 - \frac{\gamma_{ii}}{M_i^2}.$$

Since γ_{ii} is a constant, the elasticity will fall as the share increases and increase as the share falls—a property of any translog cost function.

Table 6.1 Parameter estimates for the augmented translog joint-cost function.

	Coefficient	Standard Error
α_0	0.0140	0.985
θ	0.0668	0.009
α_1	0.7521	0.332
α_2	0.3732	0.099
α_3	0.2295	0.057
β_1	0.0348	0.334
β_2	0.5008	0.326
β_3	0.4644	0.145
δ_{11}	0.1130	0.071
δ_{12}	−0.0857	0.022
δ_{13}	−0.0369	0.001
δ_{22}	0.0479	0.009
δ_{23}	−0.0061	0.003
δ_{33}	0.1662	0.002
γ_{11}	−0.0505	0.047
γ_{12}	0.0659	0.044
γ_{13}	−0.0154	0.024
γ_{22}	−0.1251	0.049
γ_{23}	0.0592	0.025
γ_{33}	−0.0435	0.031
ρ_{11}	0.1210	0.052
ρ_{12}	−0.0671	0.051
ρ_{13}	−0.0539	0.031
ρ_{21}	−0.0276	0.012
ρ_{22}	0.0349	0.013
ρ_{23}	−0.0073	0.012
ρ_{31}	−0.0369	0.007
ρ_{32}	0.1444	0.007
ρ_{33}	0.0225	0.007

Table 6.2 Goodness-of-fit statistics for augmented model.

Equation	R^2	SEE[a]
Cost function	0.9991	0.013
Capital cost share	0.9471	0.014
Labor cost share	0.9592	0.012
Toll revenue "share"	0.5260[b]	0.008
Competitive revenue "share"	0.9602	0.004

a. SEE: Standard error of estimate.
b. The toll revenue share is approximately constant (at 0.34), which accounts for the low R^2. The standard error of the estimate, 0.008, indicates a high degree of explanation in spite of the low R^2.

Table 6.3 Own-price elasticities of factor demand.

	Capital	Labor	Materials
1952	−0.800	−0.807	−1.04
1964	−0.671	−0.989	−1.02
1975	−0.589	−1.11	−1.05

Table 6.4 Own- and cross-price elasticities of factor demand (1964).

	Capital	Labor	Materials
Capital	−0.671	0.508	0.163
Labor	0.627	−0.989	0.363
Materials	0.365	0.658	−1.02

Note: In order to interpret the numbers, apply the convention that the effect of a change in the price of capital is contained in the first row, the effect of a change in the price of labor is contained in the second row, etc. Thus, the first row consists of elasticities of the form $d\log X_i/d\log P_k$ $(i = K, L, M)$.

As the cross-price elasticities indicate, all three inputs are substitutes, with output held constant. Thus, the cost-function concavity conditions will be satisfied at the point of approximation, 1964.

Tests of the Structure of Production

Table 6.5 summarizes the tests of the various specialized production structures discussed above. The tests were performed using the likelihood-ratio-test. The likelihood-ratio-test statistic, $-2\log(L_1/L_0)$, is distributed asymptotically as χ_r^2; where L_0 and L_1 are the values of the unconstrained and constrained likelihood functions, respectively, and r is the number of additional restrictions contained in the null hypothesis (the degrees of freedom of the test). Since we view the translog cost function as a second-order approximation, the tests presented in table 6.5 are approximate (see Denny and Fuss 1977).[28] The point of approximation chosen was

Table 6.5 Tests of the production structure.

Structure	Log Likelihood	Test Statistic $[-2\log(L_1/L_0)]$	Degrees of Freedom	Critical Value (5%)
Joint cost function (maintained hypothesis)	464.48		98	
Homotheticity	441.96	44.96	6	12.59
Cobb-Douglas[a]	n.c.[b]	n.c.	15	n.c.
Separability	450.04	28.88	4	9.49
Nonjointness (lack of economies of scope)	1952: 462.47	4.02	3	7.82
	1964: 464.07	0.62	3	7.82
	1975: 462.83	3.10	3	7.82
Constant returns to scale	1952: 461.84	5.28[c]	1	3.84
	1964: 464.47	0.73	1	3.84
	1975: 463.26	2.44	1	3.84

a. The Cobb-Douglas structure is a special case of the homothetic structure, and therefore is rejected along with the homothetic structure.
b. Not computed.
c. The rejection of constant returns to scale is a rejection in favor of decreasing returns to scale.

the middle year of the sample, 1964. Since the tests depend on the point of approximation, when a null hypothesis was not rejected tests were also performed using the initial year (1952) and the final year (1975) as points of approximation. Thus, three test statistics appear in table 6.5 for the nonjointness and constant-returns-to-scale hypotheses.

From the results contained in table 6.5 we can reject homotheticity, the Cobb-Douglas structure, and separability into outputs and inputs as descriptions of the telecommunications production structure. The rejection of separability is particularly important, since all previous estimates utilize this assumption (see, for example, Dobell et al. 1972; Vinod 1976a, b). A surprising result is the acceptance of nonjointness, which indicates a lack of economies of scope. We can take a closer look at this test by recalling that economies of scope imply $\partial^2 C/\partial Y_i \partial Y_j < 0$. It can easily be shown that

$$\frac{\partial^2 C}{\partial Y_i \partial Y_j} = \frac{C}{Y_i Y_j}\left(\frac{\partial \log C}{\partial \log Y_i}\frac{\partial \log C}{\partial \log Y_j} + \frac{\partial^2 \log C}{\partial \log Y_i \partial \log Y_j}\right). \tag{35}$$

If we scale the data so that the point of approximation is characterized by $P_j = Y_i = 1$ $(i,j = 1, 2, 3)$, then equation (35) is negative if $\alpha_i \alpha_j + \delta_{ij} < 0$ $(i,j = 1, 2, 3)$. Table 6.6 presents the estimate of $\alpha_i \alpha_j + \delta_{ij}$ and the associated approximate standard errors underlying the test of the null hypothesis when 1964 is the point of approximation.

Table 6.6 Estimated jointness parameters.

Formula	Estimate	Standard Error
$\alpha_1\alpha_2 + \delta_{12}$	−0.016	0.021
$\alpha_1\alpha_3 + \delta_{13}$	0.002	0.009
$\alpha_2\alpha_3 + \delta_{23}$	−0.002	0.003

Although none of the estimates of $\partial^2 C/\partial Y_i \partial Y_j$ is significantly different from zero, the point estimates of $\partial^2 C/\partial Y_1 \partial Y_2$ and $\partial^2 C/\partial Y_2 \partial Y_3$ are negative. Thus, there is some weak evidence of cost complementarities between local and toll services, and between toll and competitive services. The lack of cost complementarity between local and competitive services is reasonable, in that competitive services consist primary of private line services, which in Canada are not interconnected with the local switched network.

The hypothesis that production is subject to constant returns to scale is not rejected, except at the beginning of the sample period, where there exists evidence of decreasing returns to scale. The test statistic for the middle year (1964) is 0.13, which compares with a 5 percent critical value of 3.84. This implies that the contant-returns-to-scale hypothesis is a close description on the technology at the point of approximation usually chosen in studies that use translog functions.

We stated above that economies of scope and economies of scale are sufficient conditions for subadditivity. Since neither characteristic is strongly supported by our results, there is also no real evidence to support the contention that Bell Canada's production process is subadditive. However, we must emphasize again that our test is local, and hence not really suitable. In any case, necessary conditions cannot be investigated within our framework.

Overall Scale Elasticity

The overall scale elasticity (SE) can be measured as

$$\mathrm{SE} = \left(\frac{d \log C}{d\lambda}\right)^{-1} = \left(\sum_i \frac{\partial \log C}{\partial \log Y_i}\right)^{-1},$$

where λ is a proportionate increase in outputs (see Panzer and Willig 1975). Table 6.7 presents the estimated scale elasticity and approximate standard errors for 1952, 1964, and 1975.[29] The point estimates indicate decreasing returns for the first half of the period and increasing returns for the latter half of the period, but (as shown by the tests in table 6.5)

Table 6.7 Estimated scale elasticity (SE).

	SE	Standard Error
1952	0.845	0.068
1964	1.02	0.064
1975	1.15	0.093

are not estimated accurately enough to cause a rejection of the constant-returns-to-scale hypothesis for the latter period.

The monotonic rising trend in the returns-to-scale parameter is highly suspect. When we used a rising real rate of return instead of the constant 6 percent that yields the results of table 6.7, the parameter estimates indicated implausible scale elasticities (negative in some years, impossibly high in other years). Moreover, the estimates of scale elasticities presented in table 6.7 are sensitive to changes in the assumed real constant rate of return. Lowering the rate to 5 percent increased the scale elasticity somewhat (to 0.912 in 1952, 1.06 in 1964, and 1.20 in 1975); however, none of the values was significantly different from unity. Increasing the assumed rate to 7 percent lowered the estimate of scale elasticity (to 0.867 in 1952, 0.981 in 1964, and 1.12 in 1975); none of these estimates was significantly different from 1.0 except for the 1952 estimate.

There are two likely sources of the problems: incorrect specification of technical change, and omission of capacity-utilization measures. None of the data available to us allowed us to overcome these problems. Technical change is likely to be both factor-specific and neutral. When we tried to incorporate both Hicks neutral and factor-augmenting technical changes, the variance-covariance matrix of the estimating equations became ill-conditioned. This problem highlights the difficulty of separating measures of scale elasticity from general measures of the rate of technical progress in a highly trended time series.[30]

Measures of capacity utilization have both theoretical and empirical limitations. An aggregate output cannot be defined unless the production function can be separated into inputs and outputs. Thus, capacity output can only be defined under an assumption that has been rejected by our empirical results. One could, however, define capacity in terms of the physical limitations of utilizing capital stock. We had no information that would have allowed us to account for changes in capital-stock utilization. The inability to correct for changes in capacity utilization creates problems for the measurement of production characteristics, particularly economies of scale. If the utilization rate fell monotonically throughout the period, perhaps because of capital expansion involving lumpy expenditures on increasingly larger units, the inability to account for capacity utilization would bias the measure of scale elasticity upward.

The omission of some sources of technical change could also incorrectly attribute to scale expansions in output that should be attributable to changes in technology. Even with these two problems, which we feel bias the estimates of scale elasticity upward in the latter part of the sample period, no statistically significant economies of scale were found.

Incremental Costs and Their Relationships to Output Prices

Incremental cost curves for each of the three output can be estimated using equation (25). With the midsample (1964) observations taken as the fixed values of input prices and irrelevant outputs, the incremental cost curves were found to be downward-sloping for all three outputs.[31] Since the incremental cost elasticities are increasing for all three outputs ($\partial^2 \log C/\partial \log Y_i^2 = \delta_{ii} > 0$), at some output level incremental costs must increase. For all three services the regions of increasing incremental costs occur at output levels greater than that observed in the sample. We arbitrarily doubled each of the three outputs, and reestimated the incremental cost elasticities assuming that the observed relationship between output and costs still held. Marginal-cost curves continued to fall at these higher outputs.

We do not find it very useful to compare incremental costs with prices charged for these services for the purposes of examining relative contributions and "cream-skimming" as usually defined in regulatory hearings. As we indicated previously, these incremental costs are based on replacement or opportunity cost concepts, while the firm is regulated on a historical capital cost basis. Prices could then be below these incremental opportunity costs but be above historical fully allocated average costs as determined in regulatory hearings. Valuing inputs at opportunity costs (replacement value) might indicate losses when compared with actual revenue even though the firm was earning its allowed rate of return on its rate base. In our study, only the cost of capital and the value of the capital input exhibit differences between historical and opportunity costs. Though the firm has to pay each unit of labor the opportunity cost in each year, regulators do not allow the firm to revalue its capital (or rate base) every year on the basis of replacement value and the opportunity cost of capital.

However, it is of some interest to compare output price with incremental

Table 6.8 Actual prices and incremental costs (replacement value).

	Local Services		Toll Services		Competitive Services	
	Price	IC[a]	Price	IC	Price	IC
1952	0.92	2.16	1.05	0.40	0.78	0.32
1964	1.00	1.73	1.04	0.34	1.00	0.41
1975	1.18	1.48	1.19	0.22	1.56	0.30

Note: The index is 1967 = 1.0.
a. Incremental cost.

cost for the purposes of evaluating the efficiency aspects of rate setting. Table 6.8 compares prices and marginal costs, where the opportunity cost of capital used in the calculations is the nominal after-tax realized rate of return.

In all years of the sample, the actual price for local service was below the measured incremental replacement cost for that service. In all years, the actual prices for both toll and competitive services were substantially above the marginal costs of these respective services. If there were indeed no economies of scale, marginal-cost pricing would cover all costs. Pricing at the marginal costs of the services would substantially lower the prices for toll and competitive services and increase the price of local service. If there were increasing returns to scale, marginal-cost pricing could not cover all costs. We cannot calculate the second-best set of Ramsey prices, constraining the firm to break even, since one of the three services was estimated to have a constant elasticity of demand less than unity imply negative marginal revenue.[32]

Sensitivity of the Empirical Results to Alternative Elasticity Assumptions
The extraneous demand elasticities are estimated from rather *ad hoc* specifications. We therefore investigated the sensitivity of all our empirical results to alternative estimates of demand elasticities. We increased the absolute value of the point estimates of toll-service and competitive-service demand elasticities by two standard deviations. The adjusted demand elasticities are -2.30 for toll service and -2.03 for competitive services. The parameter estimates did change, but the characteristics of production did not change in any significant way. The estimates of the scale elasticity fell marginally (0.98 in 1964 as compared with 1.02 with the lower demand elasticities; 1.10 in 1975 as compared with 1.14 for our base-case results).

The Behavior of the Multiproduct Firm Subject to Rate-of-Return Regulation—A Duality Approach

Behavioral Model

Since 1966, Bell Canada has been subject to regulation limiting the maximum rate of return that may be earned on invested capital. It is well known that rate-of-return regulation can bias the choice of inputs away from the cost-minimizing mix. This hypothesis (known as the Averch-Johnson, or A-J effect) has been tested, somewhat inconclusively, by Spann (1974), Peterson (1975), and Cowing (1978), among others. If the hypothesis is correct, then parameters (and hence technological characteristics estimated from econometric cost functions) will be biased owing to misspecification of the behavioral model. In this section we demonstrate the way in which the A-J effect can be explicitly incorporated into econometric cost functions and the derived cost-share and revenue-share equations. A unique feature of the derivation is the extensive use of modern duality theory.[33] Suppose the product transformation function is

$$F(Y_1,\ldots,Y_m;K,X_2,\ldots,X_n) \leqslant 0, \tag{36}$$

where $K = X_1$ is the capital stock used to determine the allowed return. Then the firm's problem is to maximize

$$\sum_{i=1}^{m} q_i Y_i - \sum_{j=2}^{n} p_j X_j - p_k K \tag{37}$$

subject to (36) and

$$\Sigma q_i Y_i - \Sigma p_j X_j \leqslant sK, \tag{38}$$

where q_i $(i = 1,\ldots,m)$ are endogenous output prices and s is the allowed rate of return. The appropriate Lagrangian expression is

$$\begin{aligned}\mathscr{L} = \sum_i q_i Y_i - \sum_j p_j X_j - p_k K + \lambda_1\left(sK - \sum_i q_i Y_i + \sum_j p_j X_j\right)\\ + \lambda_2[-F(Y_1,\ldots,Y_m;K,X_2,\ldots,X_n)].\end{aligned} \tag{39}$$

If production is technologically efficient and the firm earns exactly the allowed rate of return, (36) and (38) become equalities. Further, if we assume that the optimal solution results in nonzero Y_i and X_j for all i and j, then the first-order Kuhn-Tucker conditions for a maximum of (37) subject to (36) and (38) will involve no inequalities. These conditions are

$$\frac{\partial \mathscr{L}}{\partial X_j} = -p_j(1-\lambda_1) - \lambda_2 \frac{\partial F}{\partial X_j} = 0 \qquad (j = 2,\ldots,n), \tag{40}$$

$$\frac{\partial \mathscr{L}}{\partial K} = -(p_k - \lambda_1 s) - \lambda_2 \frac{\partial F}{\partial K} = 0, \tag{41}$$

and

$$\frac{\partial \mathscr{L}}{\partial Y_i} = \left(q_i + Y_i \frac{\partial q_i}{\partial Y_i}\right)(1-\lambda_1) - \lambda_2 \frac{\partial F}{\partial Y_i} = 0$$

or (42)

$$\mathrm{MR}_i(1-\lambda_1) - \lambda_2 \frac{\partial F}{\partial Y_i} = 0,$$

where MR_i is the marginal revenue of the ith output.

Differentiating (39) with respect to λ_1 and λ_2 gives

$$\frac{\partial \mathscr{L}}{\partial \lambda_1} = sK - \sum_i q_i Y_i - \sum_j p_j X_j = 0 \tag{43}$$

and

$$\frac{\partial \mathscr{L}}{\partial \lambda_2} = -F(Y_1,\ldots,Y_m;K,X_2,\ldots,X_n) = 0. \tag{44}$$

From (40) and (41) we obtain

$$\frac{\partial F}{\partial X_g}\bigg/\frac{\partial F}{\partial X_l} = \frac{p_g(1-\lambda_1)}{p_l(1-\lambda_1)} = \frac{p_g^*}{p_l^*} \qquad (g,l = 1,\ldots,n), \tag{45}$$

and

$$\frac{\partial F}{\partial X_g}\bigg/\frac{\partial F}{\partial K} = \frac{p_g(1-\lambda_1)}{p_k - \lambda_1 s} = \frac{p_g^*}{p_k^*} \qquad (g = 1,\ldots,n), \tag{46}$$

where p_g^*, p_l^*, and p_k^* are shadow prices of the inputs. Equations (45) and (46) state that in the optimal solution the firm sets the marginal rate of technical substitution equal to the ratio of shadow prices. But this condition is just the usual cost-minimization condition, except for the fact that the prices are shadow prices instead of market prices. The firm can be viewed as acting *as if* it minimized cost subject to the shadow prices. Therefore, by solving equations (43)–(46) we can obtain the producer's constrained multiproduct cost function,

$$C^* = C^*(p_1^*,\ldots,p_n^*,Y_1,\ldots,Y_m). \tag{47}$$

Alternatively, utilizing the theory of duality between production and cost, we can start with the cost function (47) and assume that the producer acts *as if* he minimizes cost subject to the outputs and shadow prices appearing in (47). We know from the marginal conditions (45) and (46) that this basic duality property is not affected by the use of shadow prices for the inputs. Of course, the p_j^* are endogenous. However, the point of the above analysis is to demonstrate that we can treat the producer as behaving as if the p_j^* were exogenous. The endogenous nature of p_j^* will be taken into account in equations (56)–(58). Using Shephard's lemma once again, we have

$$\frac{\partial C^*}{\partial p_j^*} = X_j \qquad (j = 1,\dots,n). \tag{48}$$

Equation (48) will be used to generate the cost-share equations for the rate-of-return-regulated firm.

From the above analysis it is clear that equations (43)–(46) determine the cost-minimization solution subject to the production technology and the rate-of-return constraints. We will now show that equation (42), which determines the choice of Y_i, is just the marginal-cost-equals-marginal-revenue condition necessary for profit maximization.

From the technology constraint we obtain

$$\sum_{i=1}^{m} \frac{\partial F}{\partial Y_i} dY_i + \frac{\partial F}{\partial K} dK + \sum_{j=2}^{n} \frac{\partial F}{\partial X_j} dX_j = 0. \tag{49}$$

Using (40) and (41), equation (42) becomes

$$\sum_{i=1}^{m} \frac{\partial F}{\partial Y_i} dY_i - \frac{1}{\lambda_2}(p_k - \lambda_1 s) dK - \frac{1}{\lambda_2} \sum_{j=2}^{n} p_j (1 - \lambda_1) dX_j = 0,$$

or

$$\lambda_2 \sum_{i=1}^{m} \frac{\partial F}{\partial Y_i} dY_i - \left[\left(p_k dK + \sum_{j=2}^{n} p_j dX_j\right) - \lambda_1\left(s dK + \sum_{j=2}^{n} p_j dX_j\right)\right] = 0. \tag{50}$$

Since

$$C = p_k K + \sum_{j=2}^{n} p_j X_j,$$

we have

$$dC = p_k dK + \sum_{j=2}^{n} p_j dX_j. \tag{51}$$

In addition, from (43),

$$s\,dK + \Sigma p_j dX_j = d(\Sigma q_i Y_i) = dR. \tag{52}$$

Now suppose that only Y_i changes, so that $dY_k = 0\,(k \neq i)$. Then equation (50) becomes, using (51) and (52),

$$\lambda_2 \frac{\partial F}{\partial Y_i} dY_i - (dC - \lambda_1 dR) = 0, \tag{51'}$$

where

$$dR = d(\Sigma q_i Y_i) \qquad (dY_k = 0, k \neq i).$$

We can write equation (51) in the form

$$\lambda_2 \frac{\partial F}{\partial Y_i} = \frac{dC}{dY_i} - \lambda_1 \frac{dR}{dY_i} = \mathrm{MC}_i - \lambda_1 \mathrm{MR}_i. \tag{52'}$$

Substituting for $\lambda_2 \partial F/\partial Y_i$ in equation (42), we obtain

$$\mathrm{MR}_i(1-\lambda_1) - [\mathrm{MC}_i - \lambda_1 \mathrm{MR}_i] = 0,$$

or

$$\mathrm{MR}_i = \mathrm{MC}_i. \tag{53}$$

Thus, equation (42) is just the MR = MC condition in somewhat disguised form.

The above interpretation of the first-order conditions suggests that the overall optimization problem can be subdivided into two sequential problems. First, for any outputs, minimize cost subject to the technology and rate-of-return constraints. This defines the output expansion path in terms of shadow-price tangency conditions from equations (40), (41), (43), and (44). Second, conditional on the optimal input proportions, choose outputs so as to equate marginal revenue to marginal cost (from equation (42)).

Because a sequential analysis can be applied in the case of rate-of-return regulation, the approach used in the earlier sections of this paper is relevant. That is, first use the cost function to obtain the input demand equations and then use the profit-maximizing conditions to determine the optimal Y_i.

The constrained cost function C^* can be written as

$$C^* = p_k^* K + \sum_{j=2}^{n} p_j^* X_j, \tag{54}$$

where it is understood that K and X_j are optimal (cost-minimizing) inputs, given p_k, p_j, s, and Y_i. C^* can also be written as

$$\begin{aligned} C^* &= (p_k - \lambda_1 s)K + \sum_{j=2}^{n} p_j(1-\lambda_1)X_j \\ &= p_k K + \sum_{j=2}^{n} p_j X_j - \lambda_1\left(sK + \sum_{j=2}^{n} p_j X_j\right) \\ &= C(p_k, p_2, \ldots, p_n, s, Y_1, \ldots, Y_m) - \lambda_1 \Sigma q_i Y_i, \end{aligned} \tag{55}$$

or

$C = C^* + \lambda_1 \Sigma q_i Y_i$, where C depends only on observable variables which are exogenous to the cost-minimization problem. Now

$$\frac{\partial C}{\partial p_l} = \sum_j \frac{\partial C^*}{\partial p_j^*}\frac{\partial p_j^*}{\partial p_l} + \frac{\partial C^*}{\partial p_k^*}\frac{\partial p_k^*}{\partial p_l} + \left(\sum q_i Y_i\right)\frac{\partial \lambda_1}{\partial p_l} \qquad (l = 2,\ldots,n), \tag{56}$$

where we have explicitly recognized the endogenous nature of p_j^*, p_k^*, and λ_1.

Taking derivatives of (55) and substituting in (56) yields

$$\begin{aligned} \frac{\partial C}{\partial p_l} &= X_l(1-\lambda_1) - \left(\sum_j p_j X_j\right)\frac{\partial \lambda_1}{\partial p_l} - sK\frac{\partial \lambda_1}{\partial p_l} + \left(\sum q_i Y_i\right)\frac{\partial \lambda_1}{\partial p_l} \\ &= X_l(1-\lambda_1) - \frac{\partial \lambda_1}{\partial p_l}\left(sK + \sum_j p_j X_j - \sum_i q_i Y_i\right) \\ &= X_l(1-\lambda_1), \text{ using equation (43).} \end{aligned}$$

Thus we have a modified Shephard's lemma:

$$\frac{\partial C}{\partial p_j} = X_j(1-\lambda_1) \qquad (j = 2,\ldots,n). \tag{57}$$

We can obtain additional components of the factor-demand equations by differentiating C with respect to p_k and s:

$$\begin{aligned} \frac{\partial C}{\partial p_k} &= \sum_j \frac{\partial C^*}{\partial p_j^*}\frac{\partial p_j^*}{\partial p_k} + \frac{\partial C^*}{\partial p_k^*}\frac{\partial p_k^*}{\partial p_k} + \left(\sum q_i Y_i\right)\frac{\partial \lambda_1}{\partial p_k} \\ &= -\frac{\partial \lambda_1}{\partial p_k}\left(\sum_j p_j X_j\right) + K - \frac{\partial \lambda_1}{\partial p_k}(sK) + \left(\sum q_i Y_i\right)\frac{\partial \lambda_1}{\partial p_k} \\ &= K, \end{aligned} \tag{58}$$

$$\begin{aligned}\frac{\partial C}{\partial s} &= \sum_j \frac{\partial C^*}{\partial p_j^*}\frac{\partial p_j^*}{\partial s} + \frac{\partial C^*}{\partial p_k^*}\frac{\partial p_k^*}{\partial s} + \left(\sum_i q_i Y_i\right)\frac{\partial \lambda_1}{\partial s}\\ &= -\frac{\partial \lambda_1}{\partial s}\left(\sum p_j X_j\right) - \lambda_1 K - \frac{\partial \lambda_1}{\partial s}(sK) + \left(\sum q_i Y_i\right)\frac{\partial \lambda_1}{\partial s}\\ &= -\lambda_1 K. \end{aligned} \tag{59}$$

In summary, we can generate the input demand functions and the Lagrangian multiplier from the cost function using the modified Shephard's lemma:

$$\begin{aligned}\frac{\partial C}{\partial p_j} &= X_j(1-\lambda_1) \qquad (j = 2,\ldots,n),\\ \frac{\partial C}{\partial p_k} &= K,\\ \frac{\partial C}{\partial s} &= -\lambda_1 K.\end{aligned} \tag{60}$$

This last result was also derived by Peterson (1975), who apparently did not recognize the additional behavioral equations that could be obtained from the cost function.

Actual estimating equations can be formed by noting that

$$X_j/X_l = \frac{\partial C}{\partial p_j}\bigg/\frac{\partial C}{\partial p_l}, \tag{61}$$

which eliminates the unknown Lagrangian multiplier λ_1. This multiplier can be obtained from the above equations as

$$\lambda_1 = -\frac{\partial C}{\partial s}\bigg/\frac{\partial C}{\partial p_k}. \tag{62}$$

The remaining equations in the profit-maximizing model can be obtained from the equations $\partial R/\partial Y_i = \partial C/\partial Y_i$, where C is defined as in (55).

The Translog Econometric Model Under Rate-of-Return Constraint

The cost function can be written in the form

$$C = C(p_k, p_2, \ldots, p_n, s, Y_1, \ldots, Y_m). \tag{63}$$

For ease of notation, let $p_k = p_1$ and $s = p_{n+1}$. Then the translog approximation to the cost function (63) is

$$
\begin{aligned}
\log C = {} & \alpha_0 + \alpha_T T + \sum_{i=1}^{m} \alpha_i \log Y_i + \sum_{j=1}^{n+1} \beta_j \log p_j \\
& + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{k=1}^{m} \delta_{ik} \log Y_i \log Y_k \\
& + \tfrac{1}{2} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \log p_j \log p_k \\
& + \sum_{i=1}^{m} \sum_{j=1}^{n+1} \rho_{ij} \log Y_i \log p_j .
\end{aligned} \tag{64}
$$

The cost-share equations become

$$
\begin{aligned}
\frac{\partial \log C}{\partial \log p_j} & = \frac{p_j(1-\lambda_1)X_j}{C} \\
& = (1-\lambda_1) M_j \\
& = \beta_j + \sum_{k=1}^{n+1} \gamma_{jk} \log p_k + \sum_{i=1}^{m} \rho_{ij} \log Y_i \qquad (j = 2, \ldots, n),
\end{aligned} \tag{65}
$$

$$
\begin{aligned}
\frac{\partial \log C}{\partial \log p_1} & = \frac{p_1 X_1}{C} \\
& = M_1 \\
& = \beta_1 + \sum_{k=1}^{n+1} \gamma_{1k} \log p_k + \sum_{i=1}^{m} p_{i1} \log Y_i ,
\end{aligned} \tag{66}
$$

$$
\begin{aligned}
\frac{\partial \log C}{\partial \log p_{n+1}} & = \frac{p_{n+1}(-\lambda_1 X_1)}{C} \\
& = -\lambda_1 M_{n+1} \\
& = \beta_{n+1} + \sum_{k=1}^{n+1} \gamma_{n+1,k} \log p_k + \sum_{i=1}^{m} \rho_{i,n+1} \log Y_i ,
\end{aligned} \tag{67}
$$

where $M_{n+1} = p_{n+1} X_1 / C$ is the allowed rate-of-return "cost share."

The cost system to be estimated consists of equations (64) and (66) and equations of the form

$$
\frac{M_j}{M_2} = \frac{\beta_j + \sum_{k=1}^{n+1} \gamma_{jk} \log p_k + \sum_{i=1}^{m} p_{ij} \log Y_i}{\beta_2 + \sum_{k=1}^{n+1} \gamma_{2k} \log p_k + \sum_{i=1}^{m} \rho_{i2} \log Y_i} \qquad (j = 3, \ldots, n). \tag{68}
$$

Once the parameters have been estimated, λ_1 can be obtained from the ratio of (66) and (67). In addition to the cost-share equations, we have, as before, the revenue "share" equations obtained from the $\mathrm{MC}_i = \mathrm{MR}_i$ optimality conditions:

$$R_i = \frac{q_i Y_i}{C}$$

$$= \left(\alpha_i + \sum_{i=1}^{m} \delta_{ik} \log Y_k + \sum_{j=1}^{n+1} \rho_{ij} \log p_j\right)\left(1 + \frac{1}{\varepsilon_i}\right)^{-1}, \tag{69}$$

where ε_i is the own-price elasticity of demand.

Conclusions

Two important issues in regulation are the extent to which a utility is a natural (that is, competition-excluding) monopoly and the appropriate rate structure to be used by a multioutput regulated firm. Both issues require a knowledge of the firm's production technology. We believe that the most appropriate vehicle with which to estimate technology with these issues in mind is the multioutput cost function. We have estimated this function for Bell Canada, assuming no Averch-Johnson distorting effect.

A number of interesting empirical results emerge. First, the estimates of the overall scale elasticity are not sufficiently precise to enable one to reject the hypotheses of increasing, constant, or decreasing returns in scale. Second, we have not been able to reject the hypothesis of nonjoint production. The hypotheses of separability between outputs and inputs, homotheticity, and a Cobb-Douglas framework were, however, all rejected. Third, if the underlying technology is in fact a constant-returns-to-scale technology, efficient (marginal cost) pricing would lead to an increase in local-service rates and a decrease in toll-service and competitive-service rates.

The rejection of the most commonly specified functional forms suggests a need to use flexible forms, such as the translog form, in any estimation of telecommunications technology. However, for such an application a more extensive data base than the one available to us is desirable. This is demonstrated by the fact that our results are not as robust in the face of alternative assumptions as we would have preferred. What is needed is time-series data for a cross-section of telephone companies. The need for this more extensive data base is particularly obvious when one contemplates the estimation of the multioutput cost-function model which takes into account Averch-Johnson effects.

We thank the University of Toronto Institute for Policy Analysis for providing computer funds, and Angelo Molino for excellent research assistance.

Notes

1. For attempts to define multiproduct scale economies see Baumol and Braunstein 1977, Baumol 1977, and Panzar and Willig 1978.

2. For an application of multiproduct cost functions to U.S. railroad data see Brown et al. 1976.

3. This proof is taken from Brown et al. 1976, p. 9.

4. We are ignoring for the moment the problems of defining economies of scale, of measuring output for a firm producing a number of products, and of defining average cost.

5. Baumol et al. (1976) and Baumol and Braunstein (1977) refer to this change as "economies of scope," such as the economies of integrating a number of products within a firm.

6. They only help since the benefits of a single firm must be compared to the benefits of competition, product diversity, innovation, and cost minimization.

7. In the case of a single output, $m = 1$ and equation (1) can be solved explicitly as $Y_1 = f(X_1, \ldots, X_n)$, which is the usual form of the production function.

8. For an introduction to duality theory and the use of Shephard's lemma see Baumol 1977a, chap. 14.

9. If production is not joint, then

$$C(Y_1, \ldots, Y_m; P_1, \ldots, P_n) = \sum_{i=1}^{m} C^i(Y_i, P_1, \ldots, P_n)$$

(see Hall 1973). In this case the multiple-output technology is just a collection of single-output technologies, and

$$\frac{\partial \log C}{\partial \log Y_i} = \frac{\partial \log C^i}{\partial \log Y_i}$$

is an unambiguous measure of returns to scale.

10. Panzar and Willig (1978) developed a measure of product-specified economies of scale which may provide a solution to the consistency problem. However, their measure requires knowledge of the cost function in the region where one or more outputs are zero, and such output levels are generally unobservable.

11. See Kahn 1970, p. 53; FCC docket 20003 (Cost Separation); FCC docket 18128 (Telepek); FCC Docket 19919; Canadian Transport Commission "Cost Inquiry."

12. We are indebted to M. A. Schankerman for a memorandum entitled "Contributions and Cream-Skimming in Telecommunications Services."

13. The FCC in 1967 allocated the costs of local loops and switches to inter- and intrastate service by their relative minutes of use. When this formula resulted in "too low" a figure for interstate use, it was multiplied by three to reflect the higher "value" of interstate service (Kahn 1970, p. 153). Actually, to utilize demand elasiticities to determine the

"opportunity costs" of various joint services may be correct. However, demand studies indicate that the demand for interstate service is more elastic than that for intrastate service. Efficient pricing rules would suggest a relative decrease in the price for interstate service.

14. Economies of fill are not related in any way to long-run economies of scale.

15. In the above example, econometric analysis could associate the upgrading changes in cost with the data users if residential voice traffic did not increase at the same time. However, were business and residential traffic to increase by the same amount, system-upgrading costs would be associated with both types of demand. The typical time-series data used in studies of the telecommunications sector, including data available to us, are too highly aggregated and time-trended to permit these causal types of relationships to be estimated with any degree of precision. The required information would be time-series, cross-sectional data for a number of telephone companies.

16. One cannot use some historical embedded average cost of capital in econometric analysis, since the firm does not face these average embedded costs at the margin. Maximizing profits dependent on these embedded costs would yield the incorrect factor proportions and output decisions based on actual current market prices.

17. See Dobell et al. 1972; Waverman 1977; Houthakker and Taylor 1970.

18. This result is correct as long as demand complementarities among the monopolist's products are not sufficiently strong that an increase in the price of the product with inelastic demand so reduces demand for other products that the monopolist's total revenue falls.

19. See note 18 for an applicable qualification.

20. In the case of substitutability or complementarity among the monopolist's products, marginal revenue would be given by

$$\mathrm{MR}_i = q_i + \sum_j Y_j \frac{q_j}{Y_i}.$$

Since our empirical demand functions did not indicate the existence of the required interrelationships, this more general case has been relegated to a note.

21. In the actual estimation, the stability conditions

$$\frac{\partial \mathrm{MR}_i}{\partial Y_i} < \frac{\partial \mathrm{MC}_i}{\partial Y_i}$$

were satisfied at all data points.

22. We attempted to incorporate Hicks's neutral technical change both instead of and in addition to capital-augmenting technical change. The attempts proved unsuccessful.

23. It can be shown that if demand interdependence is present,

$$\frac{\partial \log C}{\partial \log Y_i} = \frac{q_i Y_i}{C}\left[1 + \sum_j \left(\frac{1}{\varepsilon_{ij}}\right)\left(\frac{q_j Y_j}{q_i Y_i}\right)\right],$$

where ε_{ij} is the cross-price elasticity of demand for product Y_i with respect to price q_j.

24. Since C is based on opportunity or replacement cost while the firm's actual revenue is constrainted by historical cost, $\sum_i R_i$ can be less than C with the firm still earning its allowed rate of return.

25. See Denny and Pinto 1978. A more detailed description of this test is provided below.

26. The three subcategories contained in the aggregate "competitive" service are other toll (private line, data communications, broadband, TWX); miscellaneous (consulting and other services), and directory advertising. Directory advertising was spun off to form a

separate company at the end of 1971. Our estimation procedure took this exogenous shift in revenues into account.

27. Although we attempted other formulations incorporating cross-elasticity effects, none of these attempts proved successful. There was no indication of significant interdependence among the demands for the three services.

28. The tests for homotheticity and the Cobb-Douglas structure are also exact in the sense that they do not depend on the point of approximation.

29. Standard errors were calculated as linear approximations. See Kmenta 1971, p. 444.

30. The measures of cost-diminution technical change and returns to scale are both monotonically trended. However, their combined effect on total factor productivity is to produce an almost constant rate. Ohta (1974) showed that the rate of total factor productivity can be measured as ${\varepsilon_{CY}}^{-1} \cdot \varepsilon_{Ct}$, where ε_{CY} is the scale elasticity and ε_{Ct} is the rate of total cost diminution. The resulting rates of total factor productivity are 2.7 percent in 1952, 2.9 percent in 1964, and 3 percent in 1975.

31. We reemphasize that a falling incremental cost curve is not necessarily associated with overall increasing returns to scale.

32. Only if an even number of services were subject to inelastic demands would we be able to calculate a Ramsey price vector. The solution, of course, is to relax the assumption of constant demand elasticities so that increasing the price of local services will eventually place the product in an estimated elastic demand region.

33. The approach taken in this section is similar to that used by Cowing (1978). However, we make more explicit use of duality theory to obtain the estimating equations, and do not need to assume profit-maximizing behavior with respect to the production of all outputs.

References

Baumol, W. J. 1977a. *Economic Theory and Operations Analysis*, fourth edition. Englewood Cliffs, N.J.: Prentice-Hall.

———. 1977b. "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry." *American Economic Review* 67: 809–822.

Baumol, W. J., and Braunstein, Y. M. 1977. "Empirical Study of Scale Economies and Production Complementarity: The Case of Journal Publication." *Journal of Political Economy* 85: 1037–1048.

Baumol, W. J., Bailey, E. E., and Willig, R. D. 1977. "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly." *American Economic Review* 67: 350–365.

Berndt, E. R., and Wood, D. W. 1975. "Technology, Prices and Derived Demand for Energy." *Review of Economics and Statistics* 57: 259–268.

Brown, R., Caves, D., and Christensen, L. 1976. Estimating Marginal Costs for Multi-Product Regulated Firms. Social Systems Research Institute working paper 7609, University of Wisconsin, Madison.

Cowing, T. 1978. "The Effectiveness of Rate-of-Return Regulation: An Empirical Test using Profit Functions." In M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.

Denny, M., and Fuss, M. 1977. "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates." *American Economic Review* 67: 404–418.

Denny, M., and Pinto, C. 1978. "An Aggregate Model with Multi-Product Technologies." In M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.

Diewert, W. E. 1971. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function." *Journal of Political Economy* 79: 481–507.

Dobell, A. R., Taylor, L. D., Waverman, L., Liu, T. H., and Copeland, M. D. G. 1972. "Communications in Canada." *Bell Journal of Economics and Management Science* 3: 179–219.

Fuss, M. A. 1977. "The Demand for Energy in Canadian Manufacturing: An Example of the Estimation of Production Structures with Many Inputs." *Journal of Econometrics* 5: 89–116.

Hall, R. E. 1973. "The Specification of Technologies with Several Kinds of Outputs." *Journal of Political Economy* 81: 878–892.

Houthakker, H. S., and Taylor, L. D. 1970. *Consumer Demand in the United States*. 2nd edition. Cambridge, Mass.: Harvard University Press.

Jenkins, G. P. 1972. Analysis of Rates of Return from Capital in Canada. PhD. diss., Dept. of Economics, University of Chicago.

———. 1977. Capital in Canada: Its Social and Private Performance 1965–1974. Economic Council of Canada discussion paper 98.

Kahn, A. E. 1970, 1971. *The Economics of Regulation*. 2 vols. New York: Wiley.

Kmenta, J. 1971. *Elements of Econometrics*. New York: Macmillan.

Lau, L. J. 1978. "Applications of Profit Functions." In M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.

Ohta, M. 1974. "A Note on the Duality between Production and Cost Functions: Rate of Return to Scale and Rate of Technical Progress." *Economic Studies Quarterly* 25: 63–65.

Olley, R. E. 1970. Productivity Gains in a Public Utility—Bell Canada 1952 to 1967. Paper presented at the annual meeting of the Canadian Economics Association.

Panzar, J. C., and Willig, R. D. 1975. Economics of Scale and Economies of Scope in Multi-Output Production. Bell Laboratories discussion paper 33.

———. 1979. Economies of Scope, Product Specific Economies of Scale, and the Multiproduct Competitive Firm. Bell Laboratories economics discussion paper 152.

Peterson, H. C. 1975. "An Empirical Test of Regulatory Effects." *Bell Journal of Economics* 6: 111–126.

Spann, R. 1974. "Rate of Return Regulation and Efficiency in Production: An Empirical Test of the Averch-Johnson Thesis." *Bell Journal of Economics and Management Science* 5: 38–52.

Vinod, H. D. 1976a. "Application of New Ridge Regression Methods to a Study of Bell System Scale Economies." *Journal of the American Statistical Association* 71: 929–933.

———. 1976b. Bell Scale Economies and Estimation of Joint Production Functions. Bell Laboratories, Holmdel, N.J. Submitted in the Fifth Supplemental Response, FCC docket 20003.

Waverman, L. The Demand for Telephone Services in Britain. Mimeographed. University of Toronto Institute for Policy Analysis.

Comment

Ronald Braeutigam

The article by Fuss and Waverman represents a major step forward in empirical research in the telephone industry. Their study generates a number of interesting conclusions, but even more important is the fact that it applies solid theoretical and empirical techniques to a difficult problem. The use of a flexible-form cost function and the theory of duality enables the authors to test a number of propositions which earlier studies have employed as maintained hypotheses. In addition, the econometric methods used in the paper are novel applications of well-known techniques, generally well justified, and help to identify areas where future research may lead to an even better understanding of the underlying technology.

Among the most interesting empirical results are the following. The authors cannot reject the null hypothesis that there are constant returns to scale in the industry. Neither can they reject the null hypothesis that there are no economies of scope. They do reject the proposition that the production function is homothetic, as well as the separability of inputs and outputs. The authors have also integrated a technical-change coefficient in their analysis, and conclude that capital-augmented technical change helped to reduce total cost at an annual rate of about 3 percent over the period 1952–1975. Finally, they have produced estimates of marginal cost for three types of services: local service, message toll, and "competitive" services (private line, broadband, and TWX). They suggest that local-service tariffs have been less than marginal cost for the period 1952–1975, while the tariffs for the other two service categories have exceeded marginal costs. They conclude that more efficient pricing would therefore result from some increase in local-service tariffs and some decrease in the other tariffs.

The techniques and the conclusions both represent important contributions to our understanding of the technology of telecommunications. Additionally, they provoke the reader to ask a number of questions which it may now be possible to answer with the advances in technique that are used in the article.

Usefulness of Cost Data

It is interesting that Canadian regulators are involved in a cost inquiry much like the American FCC docket 18128. The alternatives they are investigating have familiar names: embedded direct cost (EDC), long-run incremental cost (LRIC), and fully distributed cost (FDC). The most useful applications of a cost function of the sort estimated by Fuss and Waverman are in the determination of marginal and total costs, rather than in the calculation of the three aforementioned costs. The authors do not claim that the translog function will aid in the quest for EDC and FDC methodologies. However, they do attempt to describe an incremental cost function, and certain problems arise in the process. In equation (25) they state that the "incremental cost curve for output Y_i can be obtained as

$$\mathrm{ICC}(Y_i) = \frac{C}{Y_i}\left(\alpha_i + \delta_{ii}\log Y_i + \sum_{k\neq i}\delta_{ik}\log \bar{Y}_k + \sum_j \rho_{ij}\log \bar{P}_j\right),$$

where $\bar{Y}_k$ and $\bar{P}_j$ are preassigned constant outputs and prices, respectively," and where C is total cost and the parameters α_i, δ_{ii}, δ_{ik}, and ρ_{ij} are estimated.

Fuss and Waverman have used the terms "incremental cost" and "marginal cost" interchangeably. They are correct in describing equation (25) as a local description of the *marginal* cost of producing Y_i. However, it is necessary to point out that the *incremental* cost of producing Y_i is usually defined in regulatory circles as the difference in cost incurred when producing Y_i at some level, as opposed to producing a zero level of Y_i. Formally, this concept of incremental cost is represented for a two-product case as

$$\mathrm{ICC}^*(Y_1) = C(Y_1, Y_2) - C(0, Y_2),$$

where the level of Y_2 remains unchanged and factor prices are held constant and therefore suppressed in the definition of $\mathrm{ICC}^*(Y_1)$.

There is a relationship between the ICC used by Fuss and Waverman and the more conventional ICC*. In particular, if the estimated parameters of the translog function (α_i, δ_{ii}, δ_{ik}, and ρ_{ij}) were constant over the range of output of Y_i from zero to $\hat{Y}_i$, then

$$\mathrm{ICC}^*(\hat{Y}_i) = \int_{Y_i=0}^{\hat{Y}_i} \mathrm{ICC}(Y_i)\,dY_i + F_i,$$

where, in addition to the variables already defined, F_i represents a fixed cost that is not incurred unless Y_i is positive.

The distinction between ICC*($\hat{Y}_i$) and ICC($\hat{Y}_i$) emphasizes two major points about the translog function. First, since it is a Taylor-series approximation of a function, it will generally be accurate only locally. The parameters α_i, δ_{ii}, δ_{ik}, and ρ_{ij} will generally not be constant as Y_i varies from zero to $\hat{Y}_i$. Second, the translog function cannot accommodate a zero as one of its arguments (in particular, $Y_i = 0$), since it involves a logarithm of zero. Moreover, if avoidable fixed costs are attributable to a service (such as F_i in the example above), the cost function is not continuous at $Y_i = 0$. If it is not continuous, then neither is it differentiable at $Y_i = 0$.

Fuss and Waverman are aware of these limitations. They note that the local nature of the estimation prohibits their testing for subadditivity in the cost function. Our purpose here is simply to point out why one will not be able to infer the usual kind of incremental cost information referred to by regulators—ICC*—from the estimated cost function.

Joint Production and Economies of Scope

One part of the article that may cause some confusion is the treatment of "joint production." The authors describe joint production as meaning that the production function can be written as (equation (3))

$$F(X_1,\ldots,X_n) = \min(a_1 Y_1, a_2 Y_2, \ldots, a_m Y_m),$$

but cannot be written as (equation (2))

$$Y_1 = F_1(X_1, \ldots, X_d)$$

$$Y_2 = F_2(X_{d+1}, \ldots, X_h)$$

$$Y_m = F_m(X_{h+1}, \ldots, X_n),$$

where the Y variables refer to output levels and the X variables to input levels. In other words, the authors are characterizing a joint production process as one in which each input cannot be uniquely attributed to the production of a particular output.

The first point needing clarification involves the statement that "economies of scope exist if and only if production is joint." Economies of scope may exist in some cases without joint production; thus the "only if" part of the statement does not appear to be true. For example, consider the following cost function:

$$C = 10 + Y_1 + Y_2.$$

Note that the process involves costs which cannot all be attributed unambiguously to individual outputs, but production is not joint. However, also note that

$$C(4,5) = 19,$$

$$C(0,5) = 15,$$

$$C(4,0) = 14.$$

Thus,

$$C(4,5) < C(0,5) + C(4,0),$$

and the cost function does exhibit economies of scope, even though production is not joint.

The second point involves the statement that "economies of scope imply $\partial^2 C/\partial Y_i \partial Y_j < 0$." The implication is not generally true of cost functions. For example, consider the cost function

$$C = F + Y_1 + Y_2,$$

where F is a fixed, shared cost. Then

$$C(Y_1, Y_2) = F + Y_1 + Y_2 < C(0, Y_2) + C(Y_1, 0) = 2F + Y_1 + Y_2$$

and there are economies of scope, even though $\partial^2 C/\partial Y_1 \partial Y_2 \nless 0$.

Model Specification

At the heart of the article is an estimation of a translog cost function. The theoretical basis for the work is presented clearly. Specifically, it is assumed that the firm minimizes total cost in producing any observed vector of output, with no effect to the contrary such as an Averch-Johnson bias. The importance of this assumption is recognized explicitly by the authors, and a theoretical discussion of the A-J variation is presented in the concluding section. With this in mind, we confine our comments to the cost-minimizing model estimated in the article.

Those who have worked with translog cost functions are no doubt familiar with the property that the number of parameters to be estimated increases rapidly as the number of outputs and factor prices increases. Some aggregation is needed to make the model tractable. For example, the thought of having a translog function that includes factor prices for each of the twenty-eight types of labor mentioned in the paper boggles the mind! While one might like to see how the empirical results are affected by, for example, using four or five outputs instead of three, the authors are constrained by the amount of data they have available. In light of this, the categories of outputs chosen seem reasonable.

We address two areas of specification here. The first involves the following statement:

> For observed output levels such that demand is elastic, we assume that marginal revenue is set equal to marginal cost. For output levels such that demand is inelastic, we assume that those outputs are exogenous to the firm.

In other words, Fuss and Waverman have run a set of separate regressions to determine the demand schedules for each of the three output categories selected, and then calculated an elasticity of demand for each. They find that local service had an inelastic demand, and conclude that price (or quantity) regulation has in effect made the level of local service exogenous to the firm.

They find the other two service categories to have elastic demands, and therefore conclude that these output levels are endogenous to the firm, determined by profit-maximizing behavior (setting the marginal revenue equal to the marginal cost) in each market. There exists the uncomfortable possibility that regulators have specified the price (or quantity) level in each market at a level such that marginal cost exceeds marginal revenue, where the marginal revenue remains positive. Restated, the observation of a nonnegative marginal revenue is a *necessary*, but not *sufficient* condition for the profit-maximizing behavior assumed by the authors. (This assumes that the outputs are not complements; the issue of cross elasticities of demand will be addressed below.) Fuss and Waverman could have reinforced the validity of their endogeneity assumptions in these two markets with some additional institutional description about the way in which the regulatory process does in fact work.

A second potential problem of specification arises from the estimation of the log-linear demand functions used to provide additional information for the estimation of the cost function itself. Fuss and Waverman have assumed that each service "is weakly separable from all other goods and services in the individual's utility or production functions, and that the aggregate price index of these other commodities can be adequately represented by the consumer price index." Specifically, they have assumed that the cross-elasticities of demand among the three types of services are zero. This may or may not be true in the Canadian industry. It is an empirical matter that preferably should be tested rather than asserted as a maintained hypothesis. For example, competitive services may be imperfect substitutes for message toll services, at least for some users. Also, local service may exhibit some complementarity with respect to message toll services. At least in principle, if these complementarities were strong enough, then the observation of an inelastic demand for local service would not necessarily mean that the level of local service is exogenously specified. (This possibility is recognized in note 18, but the empirical test for demand complementarity is not made.)

More broadly, Fuss and Waverman employ a set of revenue-share equations (equation 22) as part of the system used to estimate the cost function. The revenue-share equations do not admit the possibility of cross-elasticities of demand; consequently, the model may be misspecified. In summary, the effects of this could range from incorrect assumptions about the endogeneity or exogeneity of each of the various services to incorrectly specified revenue-share equations. Both could alter the estimates of the coefficients of the cost function.

Pricing and Economic Efficiency

Fuss and Waverman note that the price for local service appears to be less than the incremental (or, as I have argued above, the marginal) cost, over the time period in question. For the message toll and competitive services the reverse is suggested. If there are no economies of scale (a possibility not rejected in the paper), then in the quest for economically efficient prices we need not worry about second best. If we were to set prices equal to marginal costs, then the firm would break even and economic efficiency would be maximized.

If there are economies of scale, then second best may be of interest. Fuss and Waverman state that second-best (Ramsey-optimal) prices cannot be calculated since local service has a constant elasticity of demand whose absolute value is less than unity. This by itself does not appear to be a problem. It is true that the log-linear specification of demands used by the authors does assume a constant elasticity of demand. Assume for the moment that there are zero income effects (an assumption contradicted by the empirical results in the article) and zero cross-elasticities of demand. Then, at a Ramsey optimum, two conditions would hold:

$$\Pi = 0$$

and

$$\left(\frac{p^i - \mathrm{MC}^i}{p^i}\right)\varepsilon_i = \left(\frac{p^j - \mathrm{MC}^j}{p^j}\right)\varepsilon_j \quad \text{for all } i \text{ and } j,$$

where

Π = overall level of profit for the firm,
p^i = price of ith service,
MC^i = marginal cost of ith service,
ε^i = price elasticity of demand for ith service.

The constant-elasticity assumptions would simply require that the percentage deviations of price from marginal cost—the term $(p^i - \mathrm{MC}^i/p^i)$—would be proportional in each market. The fact that the absolute value of the price elasticity in any market is less than unity does not preclude the determination of Ramsey-optimal prices.

The general assertion of Fuss and Waverman that the calculation of second-best prices may not be possible is correct. The estimates of parameters may not be constant over the range of outputs between the prevailing levels and the second-best levels. The calculation would be complicated further by the existence of nonzero income effects, as indicated by the present estimates, and by nonzero cross-elasticities of demand, if they exist.

Conclusion

The analysis of Fuss and Waverman casts doubt on a number of prior studies of the telephone industry, many of which have assumed unrealistic and restrictive production functions (such as the Cobb-Douglas, a form they reject), and most of which have characterized telephone operations with a single, highly aggregated output. They have shown how it is possible to apply flexible-form cost functions to the telephone industry, and have suggested that, at least in the Canadian case, there is a real question as to whether the industry is a natural monopoly. Perhaps more data and further refinement in the estimating techniques will enable us to say more about that question. However, the present work alone has considerably advanced the state of the art of empirical work in the telecommunications industry.

Comment

Bridger M. Mitchell

Melvyn Fuss and Leonard Waverman have given us a stimulating article, reemphasizing the importance of closely examining the behavior of the multiple-product firm and focusing our attention on the central difficulties of analyzing common costs in a regulated market. Their analytic methods —particularly when developed in more extensive form—promise to be a welcome addition to the economists' tool kit for addressing public-policy issues of rate structure, entry and competition, and the appropriate extent of monopoly supply. However, the article falls short of providing empirical results bearing on the policy questions that arise in regulation of the telecommunications sector.

I will briefly review Fuss and Waverman's methodology, and will comment on the appropriateness of their data as well as their modeling of regulatory constraints and of demand. I will then suggest how their econometric methods might be combined with an alternative procedure for estimating joint cost functions, an approach that holds greater promise for empirically establishing the long-run parameters of the production technology in the telecommunications sector.

The Model

Fuss and Waverman assume that three aggregate services (Y_i) are produced from three aggregated inputs (X_j) by a general multiproduct transformation function with convex isoquants:

$$F(Y,X) = 0, \qquad Y = (Y_1, Y_2, Y_3), \quad X = (X_1, X_2, X_3). \tag{1}$$

Provided that input prices (p) are exogenous, a cost-minimizing firm will have the dual cost function

$$C(Y,p), \qquad p = (p_1, p_2, p_3), \tag{2}$$

which embodies all of the technical information contained in (1).

Fuss and Waverman particularize this model by specifying (2) to be a general quadratic function in the logarithms of total cost, output, and price variables (the so-called translog cost function). They assume that the firm sells its services at uniform prices q, and maximizes profits

$$\Pi = \sum q_i Y_i - C \tag{3}$$

subject to a regulatory constraint to be specified. This behavior implies a particular structure for the cost shares,

$$\frac{p_j X_j}{C} = f_j(Y,p), \qquad j = 1,2,3 \tag{4}$$

and the "revenue shares" of the outputs

$$\frac{q_i Y_i}{C} = g_i(Y,p,\eta), \qquad \eta = (\eta_1,\eta_2,\eta_3), \qquad i = 1,2,3. \tag{5}$$

In their application of this model the authors simultaneously estimate the parameters of the seven equations in (2), (4), and (5) by iterated three-stage least squares, assuming that technological change is Hicks-neutral and that per capita income and Y_1 (local telephone service) are exogenous variables. The demand elasticities, η_2 and η_3, for the endogenous outputs, Y_2 and Y_3, are estimated from independent demand equations.

The principal empirical findings are that

- the overall returns-to-scale elasticity is not significantly different from unity,
- incremental costs are falling for all services,
- all inputs are substitutes,
- there is a lack of economies of scope, and
- price is below incremental cost for local service.

Data

Fuss and Waverman use annual time-series data for Bell Canada for the period 1952–1975. Output is aggregated into three services: local (Y_1), toll (Y_2), and "competitive" (Y_3). The last category includes private-line, data, TWX, and wide-area telephone service (WATS) as well as directory advertising. For Y_2 and Y_3, constant-dollar quantity indices were available, and prices (q_2,q_3) were computed as implicit indices from the ratios of current revenues and the quantity indices. The output measure for local service (which is unexplained) could be either the number of main telephone stations or the quantity of local calls. Since flat-rate charges were in effect for most customers during this period, it would be inappropriate to use demand estimates based on an average price per call.

Inputs are similarly aggregated into three broad categories—materials, labor, and capital—by constructing annual price indices for each category.

The time-series nature of the data employed by the authors poses a fundamental question for the interpretation of their results: Are year-to-year changes in total cost, factor cost shares, and output prices a reliable basis for estimating the long-run multiproduct production function? Unfortunately, annual data are dominated by short-run behavior and often reflect disequilibrium conditions. This extremely capital-intensive industry is characterized by lumpy and large-scale investment projects requiring years for systemwide installation. Capacity utilization ("fill") shows significant annual variation. Moreover, the structure of the firm's output prices is not readily adjusted to reflect year-to-year changes in both demand conditions and relative prices of inputs. Finally, the precision of the parameter estimates is restricted by the limited variation in prices, factor shares, and output proportions observed in time-series data. At best, then, the data used by the authors could enable them to estimate the local properties of *short-run* cost and production functions.

Regulatory Constraint

The impact of regulation on the multiproduct firm is a challenging topic, one worthy of extended investigation. Although the Fuss-Waverman article represents a start in this direction, it does not reach the goal of analyzing the effect of regulation on a utility's actual behavior.

To formally introduce regulation into a model of a profit-maximizing multiproduct firm, the authors take two approaches. The first, used in their empirical work, is to assume that regulation constrains the price of one output (local service) and leaves the firm free to set the levels of the remaining prices. Fuss and Waverman choose local service as the exogenous output because their demand equation obtains an inelastic price coefficient, and the model cannot function with negative marginal revenue. The authors' alternative approach, which they develop theoretically for the translog cost function, is to assume that the firm is subject to classic rate-of-return regulation on total invested capital.

The difficulty, of course, is that regulators exert (or attempt to exert) control at many points. Indeed, the accounting procedures for separating common costs between services are intended to influence the rate levels of several services, and it is possible that such price regulation would force a regulated monopoly to operate at inelastic levels of demand in all markets. However, in an aggregative model one hesitates to use up more than one degree of freedom to specify the regulatory constraint, and there is no readily-available theory of regulators' preferences to specify the degree of price control exerted in each market. It would, therefore, be interesting to have the results of the authors' second approach —for a profit-maximizing firm constrained only by rate-of-return regulation—to examine how closely such a firm's output prices resemble those observed in the data.

Finally, the authors note that rate-of-return regulation was imposed on Bell Canada in the middle of the sample period. Did this change in the firm's regulatory environment influence its costs and product behavior?

Demand

Fuss and Waverman estimate a log-linear equation for the quantity of each of the three aggregate telecommunication services as a function of each service's own price and of real income per capita. The authors' cursory attention to the demand side of the market—they suggest that their equations be regarded as first-order approximations to arbitrary demand functions—stands in sharp contrast to their systematic development of the implications of joint products on the supply side. In principle, the demands for local, toll, and "competitive" services are interdependent and must be represented in the form of a joint demand function,

$$D(Y,p) = 0. \qquad (6)$$

Several types of interdependencies are likely to be important for the particular output aggregates Fuss and Waverman use. First, the price of local service is both the cost to the customer of local calling and the price of access to obtain toll services, since he must first have a local telephone in order to place long-distance calls. The demand for toll services will, therefore, be a function of both local- and toll-service prices, especially for high-volume toll users who must purchase additional local lines in order to obtain access to the toll network. Second, some of the competitive services, particularly WATS, are close substitutes for message-toll service for high-volume users. The demands for Y_2 and Y_3 should, therefore, each be functions of both q_2 and q_3. Finally, telephone service has long been recognized as a case in which important externalities exist in the demand functions of individual consumers. In the demands for both local service and message toll service, the number of persons connected to the telephone network positively affects the demand for access as well as the number of calls at given prices. This externality, when incorporated into the firm's profit-maximizing calculus, can result in positive marginal revenues from supplying local service even when the demand function is apparently inelastic.

A satisfactory empirical investigation of the system of demand equations relevant to the Fuss-Waverman production structure will be a challenging undertaking. Over the period 1952–1975 important structural changes have occurred in telephone services. The introduction of direct distance dialing and of off-peak discount rates for toll calls has had a pronounced effect in stimulating message-toll traffic. Similarly, the availability of WATS and discount rates for bulk toll service and the recent emergence of competitive carriers for specialized intercity services have affected the conventional message-toll market as well as the market for other telecommunication services. Realistic estimates of the demand structure will need to incorporate at least the basic nature of these trends.

One can well sympathize with the authors for avoiding the additional difficulties posed by the simultaneity of demand and cost functions in a complete representation of their model. Nevertheless, it is clear that if the multiproduct firm's price policy is endogenous in its production decisions, then one may not without further investigation assume that prices are predetermined in using market data to estimate the system of demand equations that it faces.

An Alternative Approach

As an alternative to the direct econometric estimation of production or cost functions for the multiproduct firm, one may incorporate the major characteristics of the firm's joint production technology into a *process model*. In the telephone industry such an approach has been developed by S. C. Littlechild ("Peak-Load Pricing of Telephone Calls," *Bell Journal of Economics and Management Science* 1 (1970): 191–210). A process model incorporates multiple outputs Y_i and a technology consisting of items of equipment K_j with maximal capacities $\bar{K}_j$. In such a model, particular pieces of equipment are in common use for multiple outputs. For example, K_1, the local loop and switch, provides capacity that is shared between local and toll services. The capacity constraints are that usage, when summed over the set of outputs B_j that use each item of equipment, not exceed the available capacity,

$$\sum_{i \in B_j} Y_i \leqslant \bar{K}_j. \qquad (7)$$

Each item of equipment is associated with a long-run cost function that, in general, may incorporate economies of scale and factor substitution,

$$c_j(K_j,p). \tag{8}$$

When a particular set of regulatory constraints is imposed, the process model is optimized for a specified objective function. For example, the firm may be assumed to maximize profits:

$$\Pi = \sum_i q_i Y_i - \sum_j c_j(K_j,p). \tag{9}$$

As outputs, the process model yields the quantities and prices of market services, the levels of inputs, and the shadow price of each service. An important feature of the process-model approach is that the shadow prices may be interpreted as the long-run marginal costs of expanding the output of each service.

It is possible to combine the econometric and process approaches. Beginning with the process model, one may vary the input prices orthogonally over a wide range to generate minimum-cost solutions for joint production. (The model may also be used to introduce new technologies, different regulatory constraints, or the entry of competitors). The prices and outputs of these solutions constitute pseudodata that embody the firm's long-run responses to differing market conditions. One can then use these pseudodata to econometrically estimate a smooth multiple-output cost or production function, such as the translog function. Such a function will summarize the production technology into a small number of parameters representing the degree of scale economies, the nature of input substitution, and the character of expansion paths. This alternative approach of combining process information with econometric specifications avoids many of the problems inherent in time-series data, including multicollinearity, limited sample variation in key variables, and the presence of disequilibrium behavior.

J. M. Griffin's initial study of multi-output production in the electric power industry ("Long-run Production Modeling with Pseudo Data: Electric Power Generation," *Bell Journal of Economics* 8 (1977): 112–127) suggests that the psuedodata approach can elucidate policy questions that depend on the empirical measurement of long-run parameters. Although Fuss and Waverman's time-series analysis does not yield reliable information on these questions, the application of their methodology in conjunction with a process model is a promising line of future inquiry.

This comment is an elaboration of my discussion of the paper "Multiproduct, Multiinput Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada," by Melvyn Fuss and Leonard Waverman, presented at the conference. Preparation of these remarks has been supported by a fellowship from the German Marshall Fund and by National Science Foundation grant APR 77-16286 to the Rand Corporation.