Economic Council of Canada
Conseil économique du Canada

Technical Report No. 7

THE REGULATION OF TELECOMMUNICATIONS IN CANADA

Melvyn Fuss
and
Leonard Waverman
University of Toronto

...nical Reports Series
...ection des rapports techniques

Technical Reports are documents made available by the Economic Council of Canada, Regulation Reference, in limited number and in the language of preparation. These reports have benefited from comments by independent outside experts who were asked to evaluate an earlier version of the manuscript as part of the consultation process of the Regulation Reference.

TECHNICAL REPORT NO. 7

THE REGULATION OF TELECOMMUNICATIONS IN CANADA

by

Melvyn Fuss
and
Leonard Waverman
University of Toronto

The findings of this Technical Report are the personal responsibility of the authors, and, as such, have not been endorsed by members of the Economic Council of Canada.

ISSN-0225-8013

March 1981

TABLE OF CONTENTS

Page

Résumé

Une des principales questions que se posent les responsables des politiques publiques est de savoir dans quelle mesure le secteur des télécommunications constitue un monopole naturel où, par conséquent, la concurrence ne serait pas souhaitable. Une préoccupation secondaire concerne les tarifs que les entreprises de télécommunications exigent pour les divers services qu'elles rendent. Les taux sont-ils justes et suffisants ? Sinon, quels changements y aurait-il lieu d'apporter ?

Dans la présente étude, nous nous sommes penchés sur ces deux questions et, à cet effet, nous avons examiné les livres comptables et les données sur la production de la société Bell Canada. Nous avons effectué des essais empiriques de l'hypothèse du monopole naturel et analysé la structure actuelle des tarifs ainsi que diverses autres structures possibles, du point de vue du service à la clientèle. Notre étude nous a menés à deux conclusions principales :

1) D'après les essais effectués à ce jour, fondés sur le modèle très général de la technologie de production du secteur des télécommunications, rien n'indique que Bell Canada forme un

monopole naturel pour ce qui concerne les principaux services offerts. En particulier, les tests sur les économies d'échelle d'ensemble et d'autres tests sur les économies d'échelle et la portée des services de lignes privées, ne permettent pas de rejeter l'hypothèse selon laquelle les services de lignes privées pourraient, sans perte d'efficacité, être offerts sur une base concurrentielle. Nos tests ne peuvent prouver que Bell Canada n'est pas un monopole naturel, mais il est impossible, à partir des données, de rejeter l'hypothèse du système concurrentiel.

2) La structure actuelle des tarifs de Bell Canada est inefficace. D'après le critère d'efficacité, les tarifs s'appliquant aux appels interurbains sont trop élevés et les tarifs des services locaux sont trop bas. Ainsi, une diminution de 20 % des tarifs des communications interurbaines, accompagnée d'une majoration de 15 % des tarifs des services locaux, serait à l'avantage de la clientèle, sans diminution des bénéfices réalisés par Bell Canada.

Voici un bref résumé de l'étude. Dans le premier chapitre, nous passons en revue les divers documents publiés antérieurement à la présente étude, concernant les caractéristiques du monopole naturel dans le secteur des télécommunications. Les auteurs que nous avons étudiés croient, pour la plupart, en l'existence d'économies d'échelle dans le domaine des télécommunications. Cependant, faute de rigueur scientifique, les preuves accumulées dans ces ouvrages demeurent peu convaincantes. Nous présentons au chapitre 2 une évaluation de la demande relative aux services de Bell Canada. Nous constatons que la demande pour les services locaux est inélastique, alors que la demande pour les communications interurbaines et les services de lignes privées est élastique. Au chapitre 3, nous construisons un modèle dynamique applicable à une entreprise réglementée, dans le but de déterminer le coût à l'utilisateur pour les principaux services, tout en prévoyant le bénéfice approuvé pour la prestation de ces services. Le calcul du coût à l'utilisateur indique des montants inférieurs au rendement brut approuvé pour toutes les années durant la période de 1952 à 1978. Ainsi, le calcul de l'accumulation de capital pourrait comporter une erreur Averch-Johnson. Le chapitre 4 contient une appréciation de la structure des coûts de Bell Canada. La spécification de la fonction est plus générale que toutes celles qui ont été utilisées auparavant dans le secteur des télécommunications. Elle permet, en particulier, la vérification formelle des hypothèses concernant la portée et les économies d'échelle dans le cas de produits particuliers. Les résultats de ces tests ont été décrits ci-dessus. Au chapitre 5, nous tentons de vérifier l'hypothèse selon laquelle le calcul de l'accumulation de capital dans le processus de production de Bell Canada contiendrait une erreur Averch-Johnson (A-J). Le modèle, dont la performance a été pauvre, s'avère incapable de prouver l'existence d'une erreur A-J. En conséquence, la possibilité d'une erreur, telle que soulevée au chapitre 3, ne se trouve pas confirmée. Au chapitre 6, nous analysons la structure des tarifs de Bell Canada. Les résultats de cette analyse ont été résumés dans ce qui précède. L'étude se termine par une série de conclusions et de recommandations.

Summary

A primary concern of public policy decision-makers is the extent to which the telecommunications sector is a natural monopoly, hence perhaps rendering competition undesirable. A secondary concern relates to the structure of rates charged for the various services of telecommunications firms. Are these rates efficient and equitable? If not, what would be a desirable change in the rate structure?

This study has considered both of the above telecommunications issues using data drawn from Bell Canada's production and financial accounts. We have conducted empirical tests of the natural monopoly hypothesis and investigated the welfare aspects of the current rate structure and several potential alternative structures. Two main conclusions flow from our study:

(1) Hypothesis tests based on the most general model of the telecommunications sector's production technology estimated to date show that there is little evidence to suggest that Bell Canada is a natural monopoly with respect to all its principal service offerings. In particular, tests of overall economies of scale and tests of economies of scale

and scope with respect to private line services fail to reject the hypothesis that private line services can be provided on a competitive basis without efficiency loss. Our hypothesis tests cannot prove that Bell Canada is not a natural monopoly, but the competitive alternative is not rejected by the data.

(2) The current structure of Bell Canada's rates is inefficient. By the efficiency criterion, long distance message toll rates are too high and local service rates are too low. For example, a decrease in long distance rates of 20% accompanied by a 15% increase in local service rates leaves Bell Canada's profit level unchanged and results in a welfare improvement.

We now proceed to a brief summary of the study. Chapter one surveys the evidence, prior to this study, concerning the natural monopoly characteristics of telecommunications. The majority of studies surveyed support the existence of economies of scale in telecommunications production. However these studies are sufficiently flawed to render the past accumulation of evidence very weak. Chapter two is concerned with the estimation of the demand for Bell Canada's services. We find that demand for local service is inelastic, while demand for long distance message toll and private line services are elastic. In chapter three we build a dynamic model for a regulated firm which yields a user cost of capital services and an allowed return on capital services. Calculation of the user cost shows that it is less than the allowed gross return for all years in the 1952-78 period. Hence an Averch-Johnson capital accumulation bias could exist. Chapter four contains the estimation of Bell Canada's cost structure. The functional form specification is more general than any which has previously been used in the telecommunications sector. In particular it permits formal testing of hypotheses concerning product-specific economies of scale and scope. The results of these tests have been described above. Chapter five contains an attempt to test the hypothesis that an Averch-Johnson (A-J) capital accumulation bias is present in Bell Canada's production process. The model performed poorly and no evidence of an A-J bias was found. Hence the possibility of a bias, raised in chapter three, was not confirmed. In chapter six we analyse Bell Canada's rate structure. The results of this analysis have been summarized in the preceding discussion. A set of conclusions and policy recommendations complete the study.

THE REGULATION OF TELECOMMUNICATIONS IN CANADA

Chapter One

Natural Monopoly Characteristics of Telecommunications: A Review of Theory and Evidence*

1.0 Introduction

The Canadian telecommunications industry offers a wide variety of services to the public - local basic residential and local basic business services (basic telephone connections rental and calls that do not incur mileage charges); vertical local services (extensions, PBX, coloured telephones, etc.); long distance switched message toll service (MTS) - dialed calls between two parties at a time and distance charge; telegraph service; and a wide variety of business long distance services (telex, TWX, private line, foreign exchange lines, wide band, etc.).

The present structure of the industry involves a set of geographically distinct franchised monopolies in local and MTS services which together form the Trans-Canada Telephone System; a national joint venture franchised monopoly in telegraph services (CN-CP Telecommunications); and growing competition in some parts of the country for vertical and business long distance services; competition in the latter coming primarily from CN-CP. In the Challenge case, the Federal Court of Appeal upheld a CRTC decision limiting Bell Canada's tarrifing exclusivity provisions in regard to competitors as well as customers, thus opening entry by competitors into some vertical equipment markets then monopolized by Bell.[1] In the recent landmark Interconnection case, the CRTC approved an application by CN-CP for interconnection with Bell Canada's local facilities for the purposes of offering competitive business long distance services.[2] In August 1979 the Federal Cabinet refused to overturn this decision of the CRTC.

The CN-CP application for interconnection was contested by Bell Canada and other members of the Trans-Canada Telephone Systems, as well as several provincial governments (Quebec, and the four Atlantic provinces). The basic economic arguments put forward by Bell in opposition to that application, were that:[3]

1. Bell Canada was the least cost supplier

> "We at Bell Canada manage what we believe to be a natural monopoly. There are underlying factors which define the boundaries of any natural monopoly. Broadly speaking, those boundaries exist where the monopolist has just exhausted the total range of products and services which it can supply at costs lower than anyone else could achieve ....
>
> What are the ingredients which permit Bell Canada to be the least cost supplier, in real terms, of telecommunications? There are three integrally interrelated ingredients, which I shall describe in turn. These are: economies of scale, economies of scope and economies of technological change".

(De Grandpré Exhibit Bell 33, p. 2.3)

2. The competition from CN-CP would destroy the principle of route-average pricing. Route-average pricing insures that the rates for a particular service between points of equal distance are identical even though the density of traffic differs on the routes.

3. Entry by competitors would destroy the present cross-subsidization of local service rates by long distance services. Entry would lead to a revenue loss for Bell, which Bell would have to recover in order to earn a "fair" rate of return by raising rates for those non-competitive services with low demand elasticities.

> "The study of local and toll revenue/cost relationships shows that local revenues do not cover the causally related costs. Toll revenues, on the other hand, not only cover the causally related costs but cover the local service short fall and make a substantial contribution to 'common costs' as well".
>
> (Bell CRTC Interrogatory 403, att. p. 18)

4. CN-CP was a 'cream-skimmer' - a firm that would enter the high profit routes but leave the high cost (low profit) routes to the telephone companies.

The Bell Canada company position therefore was that entry by CN-CP would raise social costs, that local rates and low density routes would bear these costs and that only large business service users would gain.

The Commission in its decision stated that "... no conclusions can be drawn from the available evidence as to the existence or non-existence of economies of scale." (p. 199) and that

> "in the Commission's view, whether or not it is cream-skimming to provide service only to those centres within a carrier's territory where it considers there is sufficient demand or to offer service at prices which vary by route demand, the fact of the matter is that with respect to competitive services both carriers engage in these practices, and on a case by case basis they have been found to be in compliance with the appropriate provisions of the Railway Act. Nor is there any evidence in the present case that the granting of the application will create any significant changes in this respect that fail to comply with the relevant provisions of the Railway Act."
>
> (p. 229)

The Commission also stated that the revenue loss estimated by Bell had been overstated "to a considerable margin" that any "great increase resulting from the granting of the application should be borne primarily by those business users most likely to benefit from inter-connection" (p. 232).

The CRTC Interconnection decision was not based on an assessment of evidence detailing the magnitude of scale economies or cross-subsidization in the present network configuration. Bell asserted that its present offerings of telecommunications services constitute and will continue to constitute a natural monopoly and therefore the allowing of competition would increase social costs and distort socially accepted pricing principles. CN-CP did not prove the absence of natural monopoly aspects, but instead asserted that the onus of proof was on Bell but Bell had not proven that it was a natural monopoly, and therefore that competition should be allowed.

It is clear that the considerations outlined above are fundamental issues for regulatory policy - what services presently available or likely to be available in the future in telecommunications should be offered under monopoly or under competitive conditions; what amounts of cross-subsidization 'should' be allowed? To aid in examining these issues requires information on the economic characteristics of telecommunications production - what are the economies of scale, economies of scope, and economies of technological change? It is essential that empirical evidence be gathered to examine the issues.

In the remainder of the chapter we will define several crucial terms and survey the evidence which is available. As will become clear, available evidence is quite inadequate for the kinds of public policy decisions currently required in the telecommunications sector. In the remaining chapters of this study, we examine the natural monopoly characteristics of the operations of Bell Canada over the 1952 to 1978 period. Our goal was to correct the defects present in a number of other econometric studies of telecommunications firms. Unfortunately, data limitations prevented our completely correcting all these defects.

1.1 Definition of Scale Economies

The most commonly used definition of economies of scale is as follows: economies of scale are said to exist when an expansion of X% in the real volume of all inputs leads to a greater than X% increase in output. For example, if the firm utilizes only capital and labour inputs to produce output, economies of scale are present when doubling of both inputs (for example) leads to more than a doubling of output. If output (in this example) went up by 110%, the scale elasticity, a measure of scale economies, would be 1.10. This definition, however, presupposes that all inputs increase proportionally with scale.

An alternative and more general method of measuring scale economies, assuming the firm chooses the optimal input combination in order to minimize costs, is to examine the change in cost accompanying an increase in output. Thus, if output doubles and costs increase by only 91%, the scale elasticity is again 1.10. The definition in terms of cost changes is especially useful when a larger scale of output is characterized by changes in optimal input proportions. This appears to be the case in telecommunications, where a larger scale of operations is often accompanied by increased capital intensity.

There are diseconomies of scale when the increasing of inputs in some proportion leads to less than that proportionate change in output, or alternatively, when the change in output is accompanied by an increase in average cost. The intermediate case of constant returns to scale occurs when doubling all inputs, for example, just leads to a doubling of output, or when the doubling of output leaves average cost unchanged.

Note that these definitions assume two conditions: a) a simultaneous expansion of <u>all</u> inputs; b) no change in the underlying technology. <u>Holding one input fixed, and increasing all other inputs does not lead to a measure of economies of scale</u>. If capital is in excess capacity, such an experiment measures the economies of fill, i.e., the expansion of output assuming excess capacity exists in one input. Similarly, not correctly accounting for technological change overestimates scale elasticities. As an example of the first point, (a), holding capital constant and increasing labour would increase output, but this reflects two elements - the impact on output of increasing labour and the impact on output of utilizing the spare capacity of capital. This measure bears no relationship to the change in output due to a simultaneous expansion of both inputs, fully utilized. This fact has often not been recognized in telecommunications proceedings in Canada and the United States, where declining short-run or intermediate-run incremental costs have been presented as evidence of economies of scale. These phenomena are evidence of economies of fill (capacity utilization), not economies of scale.

Additions to capacity may however be lumpy. A unit of plant for one additional call cannot be added - one must add an additional radio channel on an underutilized microwave system or build a new coaxial link, for example. To correctly measure scale economies, corrections must be made for capacity utilization. One either examines data for different sized systems, each operating at capacity, or direct accounting for under-utilization must be made; the lumpiness of capital must be understood.

Changes in technology must also be accounted for. Imagine, a firm which operates with diseconomies of scale, but where substantial technical change occurs. This technical change involves a reduction in unit costs without altering the inherent diseconomies of scale present in the technology. However, were an attempt made to estimate scale elasticities without accounting for the effects of the change in technology on the unit cost reduction, one could easily find "economies of scale" where none in fact exists (see Section 1.4 for a more detailed examination of technical change).

1.2 Economies of Scope

Economies of scope are said to exist when the combination of two or more distinct services within a single firm results in lower costs than the production of each service separately by individual firms. Costs are lower when services are combined if there is some complementarity in production with respect to the inputs used to produce the services. For example, one could conceive of having two separate telephone networks - one devoted to local service and one for long distance service. Each home would then have two telephones, one for local calls and one for long distance calls. However, the two separate networks would duplicate some plant that could be used by the two services in common, since there are few homes where the telephone is continuously being used for either local or long distance calls. This sharing of common facilities leads to economies of scope - lower unit costs are involved for at least one of the two outputs - local and long distance service, by combining the facilities for the outputs within a single firm. Note, that while in this example there may be good reasons to combine facilities within one firm, the offering of services may still be done by more than one firm ('value added' utilities). Economies of scope can be achieved in a number of ways - not just the sharing of common capital as in the above example. The services can share labour or maintenance costs or there can be unit cost reductions in combining the two outputs within a single firm if there are differences in the characteristics of demand over some time interval (say, a day) leading to differences in the use of some facility at the particular service's peak demand. There may also be differences in the variability of demand over time leading to reductions in overall risk, and therefore reductions in unit costs from combining the two services.

In terms of telecommunications services, economies of scope involve the sharing of overhead and the use of common capital, possible maintenance savings and differences in peaking characteristics of the demand for the services. As the example described earlier, toll and basic local services use plant and other inputs in common and the two services have somewhat different demand patterns over the day. Similarly, basic business and basic residential services also share some plant in common. The vertical services however, do not share plant in common with other telecommunications services nor can there be any possibility of differences in peaking characteristics of demand or sharing of risk. Therefore, if there are any economies of scope between vertical services and other telecommunications services, they must be in the installation and maintenance of equipment. Business competitive services do share some facilities in common with the MTS system. Private line service, for example, does not utilize the switching network but does utilize local cable and long distance transmission facilities. The sharing of local loops is precisely the reason why interconnection was asked for by CN-CP. The sharing of long distance transmission facilities leads to economies of scope only so far as is cheaper to install and maintain the same facilities for both MTS and private line. But here the economies of scope may really be a question of the economies of scale in long distance transmission.

1.3 Sustainability

Recent theoretical research by Panzar and Willig, among others, has shown that a natural monopoly might not be able to 'sustain' itself against entry and as a result regulation may be necessary in order to prevent entry which raises social costs. Imagine a monopolist who offers a variety of services under conditions of substantial economies of scale and modest economies of scope. These services are to some extent substitutes for each other. A competitor wishing to enter one of the monopolist's markets also enjoys substantial economies of scale (but no economies of scope since the competitor is a single product firm). Under these conditions, entry by the competitor may be feasible if the competitor's costs are below the monopolist's price in the one market. This entry would, however, increase social costs since the monopolist will have to reduce his service offerings in this one market. The costs of the monopolist supplying the other markets however increases because of some loss in economies of scale and scope due to reducing this one output. This possibility is the central result of the "sustainability" literature. However, to this point the literature cited above is entirely theoretical and does not by itself suggest any motivation for regulation, monopoly or competition in the telecommunications market. The real questions are empirical - are the economies of scale substantial, are economies of scope modest and are the services substitutes?

If there are no economies of scale, sustainability is not likely to be an issue, since a number of firms can probably operate in the industry without any increase in social costs.[4] If economies of scope are substantial, then the monopolist will have a sufficiently large cost advantage that no competitor would dare enter. If economies of scope are non-existent, then again entry into "competitive" markets is feasible from the social perspective since social costs will not increase as a result of competition. If the services are not substitutes, then entry into one market does not reduce demand for other services offered by the monopolist.

Baseman has shown that a consideration of the sustainability problem suggests that the reverse problem may exist. A regulated monopolist may be able to maintain its position as a sole supplier of services for which it has no cost advantage, or may be able to enter markets in which it is at a cost disadvantage.[5]

1.4 Economies of Technological Change

To correctly measure the 'natural monopoly' aspects of telecommunications, one must consider the impact of technological change on the production of services. In addition, there are arguments which sugest that scale economies exist in the undertaking of research and development and in the implementation of technological change. We examine each of these two issues in turn.

Assume that 'data' on average costs per unit of output show a fall of 20% between two years, while output increased 30%. If one associated the fall in unit costs with the increase in outputs, then the conclusion would be that the increase in output (increase in scale) caused the decrease in unit costs - hence economies of scale would appear to exist. However that causal association could be incorrect, if the reduction in unit cost was due to technological change. In other words, the inherent relationship between output and unit cost may not have changed in the two years (or may even have worsened) but the introduction of a new means of producing output may have lowered unit cost.

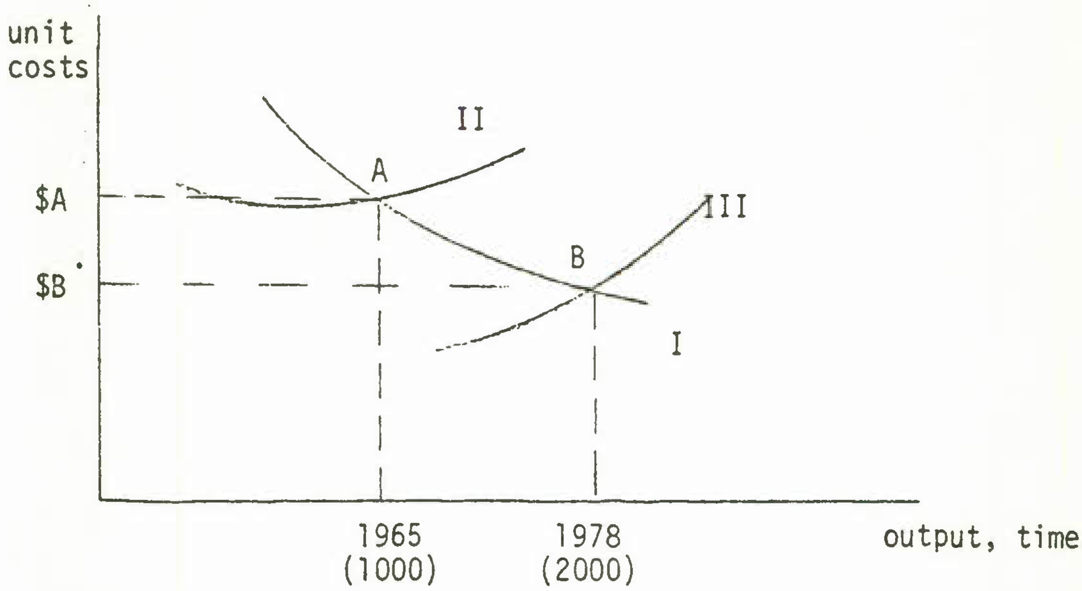


Figure 1

This point is illustrated in Figure 1. A represents unit cost in 1965 for an output level of 1000 units and B unit cost for an output level of 2000 units in 1978. Connecting A and B to show economies of scale (line I) will lead to incorrect policy decisions regarding the viability and social desirability of competition, if in fact the economies of scale relationships, after correcting for technological change, are as in lines II and III.

Unit costs for many services have fallen in telecommunications in the postwar period; these reductions in average costs have been accompanied by changes in the underlying technology of providing the services. Technological advance has been rapid in the provision of intercity transmission facilities (the movement from the use of open wire pairs to microwave and coaxial cable); the switching of calls (the introduction of crossbar and electronic switching in place of step by step); the provision of increased availability of direct distance dialing facilities; and many more specific technological advances.[6]

An important question is whether this technological advance in the telecommunications sector would have been as great or greater had more competitive entry into the provision of services been allowed in the postwar period. Schumpeter (1962) argued that monopolistic rather than competitive market structures yield greater incentives for innovation. Arrow (1962) showed that under certain simplifying assumptions, the incentive to invent is greater for competition than monopoly. However, the issue in telecommunications is not simply whether monopoly structure leads to more innovations than a competitive structure, but whether a regulated monopoly is more innovative than a competitive industry. The theoretical evidence as to whether regulation creates or destroys incentives for innovations is unclear.[7] In a recent study, Charles River Associates concluded that

> "In summary, the theoretical literature is not sufficiently developed to justify AT&T's strong conclusion about the beneficial effects of market structure on the rate of technological innovation."
>
> (Charles River Associates, The Economics of Competition in the Telecommunications Industry, August 1979, p. 238).

As part of FCC Docket 20003, several studies were undertaken which examined the extent of innovation in the various portions of the telecommunications industry, some much more competitive than others. The FCC commissioners' study[8] and the AT&T commissioned study[9] came to approximately the same conclusions - that innovation has not suffered due to competitive entry into portions of the U.S. telecommunications market. The FCC concluded, in fact, that competition had increased technological innovations.[10]

1.5 Econometric Studies of Economies of Scale

Econometric studies attempt to measure explicitly the characteristics of production (economies of scale, etc.) by estimating a production or cost function with past actual data on the operations of a firm, or an industry. By relating the changes in the use of inputs (capital of various types, labour, materials, etc.) to the changes in outputs over some past period, econometric studies determine the cost curves shown in Figure 1. Econometric studies are then essentially historical as compared, for example, to engineering studies which look at 'best practise' or hypothetical operaitons.

Econometric models deduce the technological relationships which are implicit in the results of firms' operations in markets. The prices of inputs used by the firm change, and the firm then changes the method of producing output, constrained by the technology which limits the substitution of some inputs for others. Two major assumptions are needed - that firms act rationally and that they operate so as to minimize the cost of producing any level of output.[11] Given these two assumptions, the changes in input prices that have occurred are treated by econometricians as experiments. Relating the changes in input mix to changes in input prices and changes in output trace out the underlying technology which constrains the firms actions. It is this description of technology which is the goal of the econometric studies of cost and production functions.

The economics of cost/production functions has a long history. Early work by Cobb and Douglas led to the use of the Cobb-Douglas production function. While easy to estimate, the Cobb-Douglas production function imposes severe limitations on the underlying technology - the elasticities of substitution between all pairs of factors is constant and unity.[12] In many Cobb-Douglas formulations, constant returns to scale and homotheticity are also imposed. Later work saw the development of CES production functions (the elasticities of substitution were constrained to be constant but not necessarily unity) and VES production functions (variable elasticities of substitution).[13] It became clear that production function analysis (i.e. relating outputs to inputs) in its most general form was difficult to implement empirically. However, it had been pointed out earlier by Shepherd that where firms minimize costs, the cost function was naturally dual to the production function. Analyzing the production problem of maximizing output given a set of inputs was identical to the cost problem of minimizing the cost of producing a given output. Therefore, the technological constraints which are implicit in the production function are also implicit in the cost function. General cost functions can be developed which do not impose any a priori conditions on the degree of substitutability between inputs or on the state of technology. Moreover, cost function analysis does not contain severe econometric problems. The last several years have witnessed numerous econometric examinations of technological conditions facing industries using cost function analysis.

To be useful for policy analysis, any econometric examination of telecommunications should shed light on the important public policy issues, namely the degree of competition to be allowed in the various service offerings of the sector. To be useful in regulatory hearings (as opposed to Journal publications), the results of econometric analyses must stand many tests, including robustness of results, in order for public policy to be based on these estimated technological parameters. Therefore, econometric analyses must disaggregate services, control for technological change and capacity utilization, and yield significant information on economies of scale, scope and sustainability. We turn now to a review of past studies, keeping these goals in mind.

U.S. Studies

All of the published U.S. studies of economies of scale assume a single aggregate output production process, and hence cannot provide evidence on the majority of interesting public policy questions which tend to be service-specific. In addition, all but one of the studies are the result of internal AT&T research. All of these studies use confidential data and hence cannot be subjected to the normal academic or regulatory scrutiny.

The internal AT&T evidence is presented in FCC Docket 20003.[14] In that docket, AT&T presented both econometric studies and engineering simulation network studies. The data used in Bell Exhibits 59 and 60 to FCC Docket 20003 are aggregate Bell System data for the 1949 to 1974 period. In AT&T Exhibit 60, five functional forms are used to estimate aggregate economies of scale. These functional forms are the Cobb-Douglas, the non-multiplicative, non-homogenous Cobb-Douglas, the additive non-homogeneous

Cobb-Douglas, the translog production function (a second-order Taylor-series expansion in the logs), and a variant referred to as XP-3 (a part of a third-order Taylor series expansion in the logs). The regressions are estimated both with and without attempting to account for technological change and also utilizing both ordinary least squares and a method known as the ridge regression technique.

AT&T's inference from these studies is presented in Exhibit 60:

> "The existence of economies of scale is indicated in all cases by the fact that the sums of the output elasticity coefficients are greater than unity both including and excluding technological change .... While this research is continuing at AT&T and Bell Telephone laboratories, the results presented here add considerable strength to the evidence previously reported that telecommunications exhibits economies of scale".

In fact, the above conclusion <u>cannot</u> be drawn from the available evidence. Both Exhibits 59 and 60 provide only point estimates of scale elasticities. No estimated standard errors are presented which could be used to test hypotheses concerning economies of scale. The statistical range of the results could well include diseconomies of scale, constant returns and increasing returns to scale.

Two other shortcomings limit the credibility of these Exhibits as evidence of the existence of economies of scale in the U.S. Bell system. First, as indicated before, economies of scale can be measured only after one corrects for utilization of capacity. No attempt has been made either in these studies (or in the Canadian studies, to which we will refer presently) to deal with the effects of variations in the utilization of capital. As a result one would attribute incorrectly to economies of scale the economies of fill that we have discussed earlier in this paper. Secondly, when the technique of ordinary least squares is used, and where technological change is accounted for, the point estimates of the scale elasticity are in fact less than one, suggesting diseconomies of scale. The point estimates exceed unity only when a technique called "ridge regression" is used. It is unusual to find a technique which is at the frontiers of econometric research playing an important role in regulatory proceedings. But AT&T has used ridge regression results as its primary econometric evidence that economies of scale exist in the Bell system. However, this assertion must be viewed with considerable suspicion.

Ridge regression is a procedure, which unlike least squares, requires a substantial subjective input by the investigator. The investigator, in effect, adjusts data which he finds deficient in informational content in order to yield results which appear plausible *a priori*. There is a great danger in this procedure that the investigator will discover what he *expects* to discover.[15] The problems inherent in the use of this procedure are compounded when confidential data are utilized.

In summary, the least squares estimates of the scale elasticity contained in the AT&T Exhibits generally suggest that aggregate economies of scale do not exist in the Bell system. This conclusion is weakened or reversed only if the ridge regression technique is employed or if the scale elasticity includes the measured effects of technological change. As indicated, the ridge technique results in arbitrary and biased estimates, and is inappropriate for policy purposes. The inclusion of the effects of technical change in the scale elasticity is conceptually incorrect. Its inappropriate inclusion will overstate the true scale elasticity and be seriously misleading in an evaluation of the "natural monopoly" conditions.

One study which does indicate the existence of statistically significant aggregate scale economies for the Bell system is a recent study by Nadiri and Shankerman (1979). They estimated a four input (capital, labour, material, R.&D. input) one output translog cost function for the period 1947-1976.[16] However, this finding of aggregate scale economies is of little use to regulatory deliberations concerned with service-specific issues since the source of the increasing returns to scale cannot be determined.

If econometric studies show economies of scale, the specific service source of these scale effects must be pinpointed. Public policy towards competition is likely to be different, for example, if local services are the only services with scale economies as compared to the case where all services exhibit scale economies.

In none of these U.S. studies are economies of scope examined (since there is only one output, scope is not an issue). Nor is there an analysis of the elasticity of demand. These studies then do not shed any light on the degree of regulation or competition which maximizes social welfare in telecommunications.

Studies of Economies of Scale of Bell Canada

Several studies have been undertaken in order to estimate economies of scale in the Bell Canada aggregate system. In a study for the Telecommission, Dobell et. al. (1972) estimated a production function using aggregate Bell Canada data for the 1952 to 1967 period. Technological change was accounted for by including a variable denoting the extent of direct distance dialing measured by the percentage of station-to-station toll calls dialed by the customer. The production function used was the Cobb-

Douglas form. The scale elasticity, estimated by least squares, was 1.11, significantly greater than unity at a 97.5% confidence level. While the estimate is consistent with the hypothesis of economies of scale for Bell Canada, there are several problems with the economics of this study. First, no capital capacity utilization measure is incorporated, thus likely biasing the results. Second, the Cobb-Douglas formulation imposes a rigid structure on the firm, one not likely to be met. This form as stated earlier implies a unitary elasticity of substitution between capital and labour, and homogeneous production with respect to changes in inputs. The latter restriction requires the capital-labour ratio to remain constant as scale expands, and thus the Cobb-Douglas form cannot capture an important feature of telecommunications technology - that the capital intensity of production increases as scale increases. The translog function estimated in AT&T Exhibits 59 and 60 to FCC Docket 20003, by Nadiri and Shankerman, op. cit. and in two recent Canadian studies discussed, below is generally considered a sufficiently flexible functional form which overcomes the restrictiveness of the Cobb-Douglas form. This function, the translog function is a second order approximation to an arbitrary production (or cost) function.

Available statistical results indicate that the Cobb-Douglas function is not appropriate for the telecommunications industry.[17] To the extent that the Cobb-Douglas function cannot be accepted as appropriate, estimates of scale economies obtained from this function are based on biased estimates of parameters, and should not be used for public policy purposes.

A second study of returns to scale for Bell Canada, using aggregate data, is a study by Carr (1972). He divided the 1952-67 period into two technological epochs: 1952-58 and 1959-67. The break was chosen at

1958-59, since during these years a major technological innovation, direct distance dialing, was introduced. Carr estimated separate Cobb-Douglas functions for each of his two periods. In the first period, the scale elasticity was not statistically different from unity. In the second period, the estimated scale elasticity was significantly different from unity and substantial at 1.78. Carr's results would appear to indicate substantial returns to scale in the period 1959-67. However, his study is subject to several criticisms. First, Cobb-Douglas functions were estimated. Secondly, within each period, technical change is unaccounted for, thus probably biasing upwards the measure of economics of scale. Third, utilization problems are not considered. The introduction of direct distance dialing coincided with large increases in investment expenditures, and the utilization of this investment was spread over the subsequent years in Carr's sample. Carr's estimate of economies of scale may be to a large extent, attributable to economies of fill. Of course, this latter problem is shared by all econometric studies to date in telecommunications. Finally, Carr's study has only a single output.

Among the most recent studies of Bell Canada's technology are studies by Fuss and Waverman (1978), for the 1952-75 period, and Denny et. al. (1979) for the 1952-76 period.[18] These studies were the first to disaggregate output. The output categories chosen were local services, message toll services, and competitive services; the three categories most relevant for the majority of the regulatory issues (although competitive services include WATS revenues). The authors used slightly different specifications of the multi-product translog cost function to estimate the characteristics of the technology. The point estimates of the scale elasticities indicated economies of scale. This result was not statistically significant in the Fuss-Waverman study but was statistically significant in the Denny et. al. study, indicative of a lack of robustness to change in specification. In both studies the scale elasticity is monotonically trended upward over time.[19] This trend is almost entirely accounted for by the trend in the local service scale elasticity. The indications of aggregate increasing returns to scale are likely caused by increases in the provision of local services - evidence consistent with but not necessarily proving the hypothesis that the provision of local services is at the centre of any natural monopoly that exists with respect to Bell Canada's technology.

The two studies discussed above have some limitations in common with previous studies. Capacity utilization problems are not treated, due to the lack of appropriate data in the public domain. The treatment of technical change is not satisfactory. In the Fuss-Waverman paper, technical change is specified as a capital-augmenting exogenous time trend. The Denny et. al. paper improves on this situation somewhat. Explicit indicators of technical change (access to direct distance dialing facilities and central offices with modern switching) are utilized and permitted to affect the costs of the three service categories separately. However, technical change remains exogenously determined, as is the case with all other studies of the telecommunications technology. The firm is not hypothesized as constrained by regulation and the cost of capital, a crucial variable is rudimentarily estimated. These studies are not sufficiently developed for regulatory purposes.

A second set of recent studies of Bell Canada's technology are those of Breslaw, Corbo and Smith of Concordia University.[20] Smith and Corbo (1979) have examined economies of scale for Bell Canada over the 1952-76 period with one and two output cost and production functions. For the one output case they state "Examination of the scale parameter suggests that Bell has been characterized by moderate scale economies but that for the last years of the sample the extent of these economies has dropped." (p.45). They find similar results when output is disaggregated into regulated and competitive services (p.83).

In their December 1979 report, Breslaw and Smith extend the Concordia analysis to three outputs (local, MTS, competitive) and find economies of scale. In their March 1980 Final Report Breslaw and Smith utilize two outputs again, but this time local and monopoly toll. Competitive services are deleted due to "... some computational problems experienced during the simultaneous estimation of the cost and demand model ... [because] the demand for competitive services was insensitive to price variation. The explanation of this problem lay in the Bell construction of the (1967 constant dollar) output and price series." (p.10). They excluded competetive services by adjusting downward labour, capital and materials inputs by the proportion of revenues contributed to competitive services. In general, such an adjustment is incorrect. To the extent that there exists substitution possibilities among outputs the resulting parameter estimates will be biased. Moreover, it is precisely the eliminated competetive services which are at the centre of current public policy discussions concerning the proper boundary of natural monopoly in telecommunications.

In summary, the most recent econometric studies from the U.S. and Canada (prior to this study) appear to provide weak evidence in support of the claim that there exists overall increasing returns to scale in the provision of telecommunications services. Three surveys of the U.S. data have been made, two of these appearing in F.C.C. Docket 20003. T&E Inc., consultants to the F.C.C., concluded that the case for economies of scale had not been proven.[21] Stanford Research Institute, examining the same studies for AT&T, concluded that economies of scale existed in the range of 1.1 to 1.25.[22]

Evidence from Bell Canada Cases Before Regulatory Commissions 1945-78

The majority of cases involving Bell Canada before Federal regulatory authorities in the 1945-1978 period involved rate applications (1950, 1951/52, 1957, 1958, 1966, 1969, 1970, 1971, 1972, 1975, 1977, 1978). These rate applications were examined for any evidence, econometric or otherwise, of economies of scale.

In 1958, Bell witness Hambly testified that the cost per circuit of long distance circuits fell as volume rose and therefore that long distance rates should be kept at levels which would attract high volumes, especially for the long haul circuits.[23] No mention was made however of the changing technology including changes in transmission method and the introduction of direct distance dialing. This testimony is not then proof that economies of scale exist.

The issue of economies of scale do not appear to have been raised again until the 1972 rate case.[24] Counsel for the Hotel Association of Canada argued that there was no reflection in the tariff for message registries of economies of scale, unlike the tariffs for PBX. (Vol. 14, pp. 2245-50). The Ontario Government argued that the tariff for telepak channels exhibited the presence of economies of scale. (Vol. 18, p. 2012-6). When the Ontario Government questioned the lower rates on U.S. calls as compared to calls of the equivalent distance within Bell territory, Bell witnesses replied that the Canada-U.S. tariffs were lower because of the economies of scale 'enjoyed' by AT&T. These three statements were assertions with no evidence given to back them up.

1.6 Engineering Evidence on Economies of Scale

AT&T submitted a number of network engineering studies to Docket 20003 which purport to show economies of scale. There are a number of serious defects with these engineering studies as indicators of whether economies of scale in fact exist in the operations of the presently franchised carriers. First, they do not differentiate between short-run economies of fill and long-run economies of scale. Second, they are based on highly generalized network analyses utilizing costs that do not, in fact, apply to any actual operating system. Moreover, when these studies find that fragmentation of the service among many suppliers leads to higher costs, they assume that the cost of the fragments or the independent suppliers can be no lower than the costs for AT&T. As W. Melody in a detailed critique of the engineering studies in Docket 20003 has suggested.

> "specialization is used simply as one piece of a larger system which is optimized under single system planning. Thus, to measure attributes of specialization - the application of technology and different ways to service diversified market demand rather than general homogeneous demands, and the more rapid response to changing demand and technological opportunities - are assumed away. In this analysis, specialization simply reflects sub-optimization on a known system for a limited number of known services .... Inasmuch as specialization must begin by building in differentiations in demand, pursuit of the benefits of economies of specialization is severely restricted, if not rendered impossible".

(W.H. Melody, Comments in H. Trebing, ed., New Dimensions in Public Utility Pricing (Michigan State University: East Lansing, 1976, p. 391).

It should also be noted that these engineering studies do not treat systems costs other than investment in any systematic manner, i.e., operating and maintenance costs are incorporated by assuming generally that operating and maintenance costs remain in the same proportion to scale as investment costs for all levels of scale. Charles River Associates summarized the engineering studies that appeared in Docket 20003 in this way. (p. 213).

> "In summary, engineering studies which conclude that large economies of scale are present in long distance terrestrial transmission investment costs typically focus on only one component of such costs (basic transmission), excluding all related equipment such as multiplexing (and switching when needed) and related operating costs. When these other cost elements are included, overall economies of scale appear lower. Whatever scale economies there may be moreover, do not preclude the offering by "value-added" common carriers of nonconventional and innovative long distance services, if cost-saving techniques or valuable new features are made available to customers. Moreover, entry into long distance may be feasible on many higher density routes without incurring unduly large cost penalties, especially with the newly emerging satellite technologies".

It is our view that these engineering studies shed very little light on the issues of service specific economies of scale, economies of scope and sustainability.

1.7 Evidence on Economies of Scope

There have been several recent econometric studies of the existence of economies of scope in telecommunications. Our remarks in Section 1.2 suggested that cost complementaries likely exist between toll and basic local services, but not likely between competitive and toll services or competitive and basic local service. Fuss and Waverman (1978), using a relatively weak testing procedure, found some evidence of cost complementaries between local and toll services and toll and competitive services. However, no statistically significant evidence of cost complementaries was found. Breslaw and Smith, in their Interim Report, using the same procedure as Fuss and Waverman, found economies of scope between local and competetive services but "insufficient evidence to deduce the existence of scope economies between local and message toll services and message toll and other services." (p.17). In their Final Report, Breslaw and Smith concluded that economies of scope existed between monopoly toll and local services but that they were "unimportant" (p.41).

Footnotes to Chapter One

* An earlier draft of this chapter appeared as "The Regulation of Telecommunications in Canada: Notes on Evidence Pertaining to the Natural Monopoly Characteristics of Bell Canada", Law and Economics Workshop Series No. WS II-8, Faculty of Law, University of Toronto.

1. re: Bell Canada and Challenge Communications Ltd. (1978), 86 D.L.R (3rd) 351 (leave to appeal dismissed by the Supreme Court of Canada on 19 June 1978).

2. Telecom. Decision CRTC 79-11, Ottawa, 17 May 1979.

3. These and similar arguments have been put forward by the telephone companies in their general opposition to competition in the provision of telecommunications services, including vertical services.

4. Note however that a natural monopolist may operate under local decreasing returns to scale. For an analysis of this case see Panzer and Willig (1977).

5. See Baseman (1977) for an analysis of this case.

6. M. Denny, M. Fuss, C. Everson and L. Waverman (1979) examined the impact of technical advance on the unit cost of Bell Canada's operations over the 1952 to 1976 period. It was found that a 1% increase in the percentage of telephones with access to direct distance dialing facilities reduced total unit cost by .04% and involved a substitution of

capital for labour and materials services. A 1% increase in the number of telephones connected to modern switching equipment (crossbar and electronic) also reduced total cost by .04% by reducing the use of all factors - capital, labour and materials. However, at the present state of this research the results reported above are not sufficiently robust for the public policy purposes. Small changes in specification of the equations lead to relatively large changes in the measured impact of technological change (and hence the impact of scale economies vis à vis technological change in reducing unit costs).

7. For an evaluation see Westfield (1971).

8. The Report, Appendix A, FCC Docket 20003.

9. "Analysis of Issues and Findings in Docket 20003"; April 1977, Bell Exhibit 65A.

10. "In the Matter of Economic Implications and Interrelationships Arising from Policies and Practices Relating to Customer Interconnection, Jurisdictional Separations and Rate Structures", FCC Docket 20003, First Report, adapted August 20, 1976.

11. In this section we ignore the impact of rate of return regulation on the firm's desire to minimize costs and on the ability of econometricians to model technology correctly.

12. These assumptions have been rejected in analyses of telecommunications data, see Fuss and Waverman (1977).

13. See, for example, Beckmann and Sato (1969).

14. This F.C.C. Docket examined the potential for competition in intercity communications.

15. An attractive feature of the ridge regression estimator is that there exists, for any estimation problem a ridge estimator which is preferred to the least squares estimator in the sense that it provides more precise estimates. The problem is that there is no generally accepted procedure for finding the correct ridge estimate. It has recently been shown that the commonly used procedures are deficient in the sense that they do not yield estimates which are necessarily more precise than the least squares estimates (Ullah (1978)). In addition, ridge estimators yield biased estimates so that one must be careful in using the point estimates calculated. This is a particularly serious problem since no measures of precision are presented in the AT&T exhibits to FCC Docket 20003, nor in fact are any available. A characterization of the statistical distribution of the ridge estimator sufficient to yield interval estimates has not been found, and remains an unsolved problem of theoretical econometrics.

16. Although this study is not an internal AT&T study, once again the data base is confidential.

17. The Cobb-Douglas function is a special case of the translog function. Nadiri and Shankerman (1979), Fuss and Waverman (1978) and Denny et. al. (1979) provide empirical results leading to a rejection of the Cobb-Douglas specification for the U.S. and Canada respectively.

18. See also Denny, Fuss and Waverman (1979).

19. The same result occurs in the Nadiri-Shankerman U.S. study.

20. See Smith and Corbo (1979), and Breslaw and Smith (1979, 1980).

21. T&E Inc. "A Project to Analyze Responses to Docket 20003". Final Report, Deliverable B. Prepared for the F.C.C., 24 September 1971.

22. SRI, "Analysis of Issues and Findings in FCC Docket 20003", Part II, Section D-2, April 1977.

23. Vol. 1001, p.

24. CTC Decision C-955.182.

Chapter Two

The Demand for Bell Canada's Services

2.0 Introduction

Rates which are designed to maximize social welfare cannot be based on costs alone. Both cost and demand considerations are necessary for optimal rate design. A simple example designed by William Baumol will demonstrate the point:

> "But the fact is that no cost calculation can guarantee the profitability of a service; that depends also on the state of demand. The production of an item that has gone out of fashion cannot make ends meet no matter what cost accounting procedures it employs. A full cost calculation will bring in the revenues it is designed to obtain only if the demand expectations on which it is based turn out to be justified. . . .It is also easy to show that the profitability of a service can sometimes be increased by a reduction of a price to a level below fully allocated cost, but one which covers incremental cost. Consider an enterprise with a million dollars in fixed costs and a service whose variable cost is one dollar per unit of output. If at a price of $10 it sells 100,000 units the price will clearly not cover its full cost - it will bring in $1 million and its total cost will then be $1,100,000. Nevertheless, a reduction in price further below the initial full cost level of $5 will bring a profit to the firm if it trebles sales, for then total revenue will be $5 x 300,000 = $1,500,000 as compared to a total cost of $1,300,000."[1]

Therefore, to examine rate design, knowledge of the demand curves facing Bell Canada is needed. Since we are examining a multi-product cost function, demand curves are required for each separate service. Moreover, demand should be further disaggregated in order to distinguish between groups with different demand characteristics. It is normally thought that business and residential customers as groups exhibit different elasticities of demand. As a result, demand characteristics should be differentiated by class of service (local, MTS, competititive) and by class of customer (residential, business).

2.1 Past Studies

Several past studies have examined the elasticity of demand for telephone services in North America. Dobell et al (1972) examined the elasticity of demand for Bell Canada local and toll services for both residential and business customers over the 1952 to 1967 period Their results are shown below:

Elasticity of Demand (Bell Canada)

	Local		Long Distance		Total	
	SR	LR	SR	LR	SR	LR
Residential	0	0	-.3	-1.9	-.09	-1.2
Business	0	0	0	0	-.4	-1.3

The price of local telephone service (measured as the monthly charge in centres of over 250,000 telephones) did not influence the number of telephones in use. Nor, surprisingly did the price of long distance service (measured as the price of a 350 mile call) influence the business demand for long distance service. It is clear, however, that the prices of telephone service employed in all these regressions are simply proxies for the true implicit price indices, and likely incorrect proxies.

For example, the price changes of a 350 mile call would not adequately represent the total toll price index if the mix of calls varied over the period (less short calls, etc.) or if the relative prices of long distance calls of varying length changed. Nor does the proxy for a toll call price index used by Dobell et al incorporate changes in peak and offpeak rates. A correct toll price index would examine the changes in prices for calls at specific times over specific distances, weighted by their share in total calls.

The Concordia University group incorporate a detailed demand side in their modelling of Bell Canada. The demand for telephone service (local; MTS; residential local, residential business) is analyzed as a double logarithmic function of the real price, income, population, and the number of conversations (local). The estimated price elasticities are:

total local	-.519
residential local	-.395
business local	-.706
MTS	-1.292

The results of an econometric examination of price elasticities by AT&T were presented in the Spring 1973 Bell Journal and are shown below:[2]

	Price Elasticities of Demand (US)	
	SR	LR
Local	-.21	-.31
Message Toll	-.88	-1.02
WATS	-.14	-.83
Private Line Service	-.74	-.90
Private Line Telephones	-.01	-1.00

The price indexes used in this study attempt to correctly measure price changes of the various components of the aggregate measures. Note, however, that the demand functions are independent of each other. For example, it is assumed that the price of message toll service does not affect WATS demand. For public policy purposes it is crucial information whether services are substitutes or not; the independence of demands should be tested rather than assumed.

Littlechild and Rousseau (1975) utilize a complex mathematical programming model to examine price elasticities for telephone services over the day. Four time periods are examined - day (6 a.m. to 6 p.m.), evening (6 p.m. to 8 p.m.), night (8 p.m. to 12 p.m.) and after midnight (12 p.m. to 6 a.m.). Unlike the other studies which were based on telephone company revenue, Littlechild and Rousseau use actual traffic on three routes in the Illinois Bell territory for 1962 and 1963. Their overall interstate price elasticity of demand is -.99, while the weighted average day price elasticity is -.90, and the evening night elasticity is -1.7. Cross price elasticities between day and night calls range from .12 to .37.

In their simulation of the effects of competition on AT&T revenues, Charles River Associates use a range of price elasticities, a range they feel is representative of the demand estimation results to date[3]. For toll calls the elasticities are as follows:

Price Elasticities used by Charles River Associates in Simulations

	low	high
8 a.m. - 12 noon	-.3	-.7
12 noon - 5 p.m.	-.3	-.7
5 p.m. - 11 p.m.	-1.1	-1.3
11 p.m. - 8 a.m.	-1.1	-1.3
weekend	-1.1	-1.3

The other price elasticities used in the Charles River Associates study are not explicitly given but appear to be: -.2 for connections to the system and -.2 for local calls. The demands for services appear to be independent in the Charles River Associates study; wide changes in local and MTS rates are assumed not to affect the WATS and private line market demands.

These previous demand studies have a number of defects - not all differentiate business and residential traffic; only one (Littlechild and Rouseau) examines cross elasticities of demand; much of the price data used are rudimentary; only one (AT&T) separately breaks out private line and WATS services.

In our study of Bell Canada, we examined a number of different demand formulations utilizing various measures for the prices of services, since data are often not publicly available to enable the degree of disaggregation necessary. We begin by examining the data series available to us.

Ideally, we would have liked to obtain current and constant dollar revenue series for each of the service components. We would have preferred to use in our econometric estimation current dollar revenues for local (business and residential separated), toll (business and residential separated) and other (WATS, private line separated). The crucial variables missing were price indices to convert the residential and business current dollar revenue series into constant dollar series (or measures of output).

2.2 Data Series - Bell Canada Toll

Various bits of information were available to us, information which we hoped would piece together and allow the construction of toll price indices. This information was as follows:

1) Tariffs: The CRTC (and its predecessors) regulates Bell's intra-company toll tariffs. Tariffs for adjacent areas, Trans-Canada Telephone System traffic and U.S. and Overseas traffic are filed with the Commission, but are not effectively regulated. Until amendments to the Railway Act in 1970, rates for private line and leased circuits were unregulated.

For toll traffic, two tariff schedules exist, one for customer dialed calls; the other for operator handled calls (person to person, credit card, collect). Toll calls face varying discounts according to the time of day or day of the week that the call is placed. All toll calls involve a two step tariff - the initial price for a one minute call (the minimum call was a 3 minute call until 1970) and a lower price for each additional minute. Longer distance toll calls are priced above shorter haul calls for the same duration; however, the average price per mile generally declines with distance.

2) Distribution of toll revenues by time of day and by type of toll call. The tariffs alone are insufficient information to construct price indexes since the distribution of total toll calls has changed considerably, both over the day and over the week. Available to us, was a distribution of toll revenues by length of call (under and over 3 minutes), time of day, day of week and business/residential split for a sample period in 1967. If these revenue shares were constant over time, then a consistent meaningful toll price index could be developed utilizing the tariff schedules and the weights for different types of calls derived from the 1967 distribution of

revenue shares. However, over the 1950 to 1970 period, large changes occurred in the relative prices of calls both over the day and over the week. These price changes likely increased the total quantity of calls, shifted peak demand, and induced large changes in the distribution of calls for a number of years throughout the period. Therefore revenue share weights for any single year are unlikely representative of all the years.

3) Estimates of the number of toll calls (all types) by mileage band and length of call (one minute, two minutes, three minutes, four minutes, five minutes, six minutes and over) for 1972 through 1976. In response to interrogatory requests from the province of Quebec, Bell Canada provided data on the distribution of all toll calls according to distance and duration[4].

We detail below how we constructed toll price series using the data described in (1) and (2), why these price indexes were unsatisfactory and our attempts (largely unsuccessful) at using the data in (3) to generate information on the changing distribution of calls.

<u>Construction of Toll Price Series Using 1967 Weights</u>

We collected tariff data on 'representative' distance calls, since it would be difficult to compute price indices using all the toll tariffs filed by Bell Canada. For calls wholly within Ontario and Quebec (intra Bell territory), 9 mileage bands were chosen each representing a specific call between two points - 9 miles, 15 miles, 47 miles, 68 miles, 103 miles, 131 miles, 180 miles, 218 miles, 312 miles. For Trans-Canada tariffs, 5 mileage bands were chosen - 490, 1138, 1684, 1886 and 2093 miles, each representing a route between two points. For Canada U.S. traffic, 5 mileage bands were picked - 342 , 435, 742, 1132 and 2460 miles. For each tariff schedule and for each mileage band, charges are distinguished between station-to-station (customer dialed) and person-to-person (operator handled calls). There is a day rate and a separate rate for night and Sunday calls[5]. Each call involves a fee for an initial period and then a lower per minute charge for each additional minute duration of the call. Beginning in 1965/1966 sales tax was levied (at different rates) in Quebec and Ontario on toll calls.

For 1967 only, available to us was a distribution showing the percentage of Bell Canada toll revenue for the following call categories for each of intra Bell, Trans-Canada and U.S. revenues.

	Under 3 Minutes				Over 3 Minutes				
	Station-to-Station		Person-to-Person		Station-to-Station		Person-to-Person		
	day	night	day	night	day	night	day	night	
Residential	x	x	x	x	x	x	x	x	< 100 miles
	x	x	x	x	x	x	x	x	> 100 miles
Business	x	x	x	x	x	x	x	x	< 100 miles
	x	x	x	x	x	x	x	x	> 100 miles

In addition, we also had information on the revenue distribution between Intra Bell, Trans-Canada and U.S. traffic for each of residential and business calls.

We therefore used this one year's set of weights to attempt to generate toll price indices separately for both business and residential traffic for the entire 1952 to 1978 period.

Weights (revenue shares) are not available by length of call. We therefore had to find some weighting scheme to aggregate the different calls of different distances in each tariff group (Intra Bell, Trans Canada, U.S.). Two price indexes were calculated: one utilizing a geometric mean to weight calls of different distances (therefore giving greater weight in the index to the longest calls) and a 'reverse mileage band method' which gives greater weight to the calls over shorter distances.

We took account of tariff changes and changes in sales tax percentages which occurred mid way through a year. Where the Sunday rate differed from the night rate, a weighted average based on revenue shares was used.

The final toll price index is based on a Laspeyres index (1967 weights) calculated as follows:

$$p = \sum_{i=1}^{3} \sum_{j=1}^{2} pI_{ij} w_{ij} \qquad (1)$$

where pI is an intermediate price index (defined below)

i is Intra Bell, Trans Canada or U.S.

j is day or evening

The intermediate price index was formed from the actual tariffs as follows:

$$pI_I = w_u pI_u + w_o pI_o \qquad (2)$$

where w_u is the share of revenue for calls under 100 miles in length, pI_u is the price index for such calls

w_o is the share of revenue for calls over 100 miles in length, pI_o is the price index for such calls

To form pI_u and pI_o requires some method of aggregating the various tariffs for calls of different length when there are no data on the share of revenue for different length calls. We used two different methods, the geometric mean and the reverse mileage band method, (as shown below for the Trans Canada tariff schedule (where 9 representative calls were used, 4 under 100 miles; 5 over 100 miles). The procedure is the same for Trans Canada or U.S. tariff schedules.

1) geometric mean

$$pI_u = \sqrt[4]{\pi\, T_i} \;;\quad pI_o = \sqrt[5]{\pi\, T_i} \qquad (3)$$

where T_i is the tariff for the i-th length call

2) 'reverse mileage band method'

$$pI_u = \sum_{j=1}^{4} w_j T_j \;;\qquad w_1 = \frac{M_4}{\Sigma M_i}$$

$$w_2 = \frac{M_3}{\Sigma M_i}$$

$$w_3 = \frac{M_2}{\Sigma M_i}$$

$$w_4 = \frac{M_1}{\Sigma M_i} \qquad (4)$$

where M_i is the distance of the i-th call

As can be seen from 1), the geometric mean being the product of the tariffs gives greatest weight to the most expensive calls i.e. the longest calls. Method 2) gives greatest weight to the shortest calls.

The resulting price indexes are given in Tables 2.1 and 2.2. Several points stand out. All price series show rising price indices for toll calls. Weighting shorter calls more (the 'reverse mileage band method') leads to greater

Table 2.1

LASPEYRES PRICE INDEX - LONG DISTANCE TEL. - INCLUDING SALES TAX
(Geometric Weights)

YEAR	RES.	RES.-DAY	RES.-NIGHT	BUS.	BUS.-DAY	BUS.-NIGHT
1949	0.843513	0.895697	0.813502	0.924430	0.931997	0.801736
1950	0.855573	0.900458	0.824759	0.931140	0.938169	0.817159
1951	0.915869	0.924264	0.911040	0.964689	0.969031	0.894277
1952	0.915869	0.924264	0.911040	0.964689	0.969031	0.894277
1953	0.921484	0.923259	0.920466	0.963666	0.967479	0.901849
1954	0.924291	0.922756	0.925174	0.963155	0.966702	0.905636
1955	0.924291	0.922756	0.925174	0.963155	0.966702	0.905636
1956	0.924291	0.922756	0.925174	0.963155	0.966702	0.905636
1957	0.924291	0.922756	0.925174	0.963155	0.966702	0.905636
1958	0.928727	0.927701	0.925317	0.965492	0.969407	0.910621
1959	0.967127	0.978283	0.960710	0.990425	0.993412	0.950592
1960	0.981850	0.981967	0.981786	0.996885	0.998628	0.966637
1961	0.974167	0.974954	0.973715	0.986975	0.988487	0.962476
1962	0.954081	0.950377	0.956211	0.953836	0.954054	0.950314
1963	0.951461	0.948016	0.953444	0.950375	0.950508	0.948208
1964	0.950282	0.946551	0.952424	0.948290	0.948341	0.947456
1965	0.964631	0.960633	0.966590	0.961950	0.961942	0.962069
1966	0.995595	0.989470	0.994115	0.991184	0.990819	0.997124
1967	1.000001	1.000000	1.000001	1.000001	1.000001	1.000001
1968	0.985771	1.002471	0.976155	1.000555	1.002024	0.977249
1969	0.928358	1.028375	0.970837	1.013427	1.021600	0.880908
1970	1.026803	1.156170	0.952403	1.096357	1.103273	0.984229
1971	1.102541	1.290239	0.994594	1.178328	1.187876	1.023688
1972	1.185845	1.429033	1.045955	1.265234	1.276872	1.076335
1973	1.212027	1.518147	1.082573	1.325438	1.341013	1.072915
1974	1.223346	1.575573	1.124775	1.353055	1.381765	1.059701
1975	1.251894	1.629192	[illegible]	1.415079	1.436097	1.143139
1976	1.359849	1.717113	1.154241	1.501777	1.520120	1.204372
1977	1.398439	1.784928	1.178104	1.552249	1.571987	1.232218

Table 2.2

LASPEYRES PRICE INDEX - LONG DISTANCE TEL. - INCLUDING SALES TAX

(Reverse Mileage Band Method)

YEAR	RES.	RES.-DAY	RES.-NIGHT	BUS.	BUS.-DAY	BUS.-NIGHT
1949	0.841445	0.892256	0.812223	0.916137	0.923169	0.802115
1950	0.853455	0.896892	0.828473	0.922922	0.929429	0.817421
1951	0.913507	0.920077	0.909728	0.956848	0.960727	0.893952
1952	0.913507	0.920077	0.909728	0.956848	0.960727	0.893952
1953	0.919053	0.918804	0.919196	0.955464	0.958789	0.901571
1954	0.921826	0.918168	0.923930	0.954772	0.957819	0.905381
1955	0.921826	0.918168	0.923930	0.954772	0.957819	0.905381
1956	0.921826	0.918168	0.923930	0.954772	0.957819	0.905381
1957	0.921826	0.918168	0.923930	0.954772	0.957819	0.905381
1958	0.925729	0.922924	0.927342	0.957535	0.960489	0.909651
1959	0.960649	0.971983	0.954130	0.982668	0.984981	0.945156
1960	0.974625	0.975756	0.973975	0.988655	0.990287	0.962202
1961	0.967205	0.969525	0.965872	0.979675	0.981148	0.955795
1962	0.947119	0.949599	0.945673	0.952069	0.952781	0.940526
1963	0.944363	0.947261	0.942697	0.948629	0.949270	0.938248
1964	0.943468	0.945824	0.942114	0.946617	0.947159	0.937830
1965	0.957915	0.959914	0.956764	0.960302	0.960782	0.952532
1966	0.993214	0.988730	0.995795	0.989896	0.989623	0.994334
1967	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1968	0.987080	1.002673	0.973111	1.000965	1.002307	0.979200
1969	0.938302	1.032220	0.884287	1.016973	1.024403	0.896524
1970	1.067114	1.186197	0.998630	1.116815	1.121581	1.039565
1971	1.141563	1.344607	1.024791	1.211901	1.221203	1.061070
1972	1.228429	1.508711	1.067236	1.312801	1.325662	1.104286
1973	1.262636	1.615469	1.059719	1.383395	1.400583	1.104709
1974	1.280178	1.685369	1.047148	1.428500	1.449070	1.094966
1975	1.363033	1.753848	1.138273	1.497765	1.516580	1.192698
1976	1.444707	1.860443	1.205613	1.595654	1.616004	1.265711
1977	1.479114	1.920537	1.225247	1.643467	1.665234	1.290534

price changes - thus the prices for shorter calls rose more over this period than did the prices for longer calls. In addition over this 1952-1978 period, the prices for day calls increased more than did the prices for evening, night and Sunday calls. Finally, the price to residential customers rose less than the price to business customers reflecting the predominance of station-to-station off-peak calls by residential customers.

In Table 2.3, we show a total toll price index (for both business and residential traffic) as developed by Bell Canada. There is a fundamental difference in the movement of prices over time as given by the Bell Canada toll price index and as given by the price indices we developed in Tables 1 and 2. The Bell Canada total toll price index shows a falling nominal price for toll calls between 1960 and 1963; our indices show rising nominal prices between those years. In addition, the Bell total toll price index indicates that the average toll message was priced slightly lower in 1967 than 1950; our toll indices for business and residential calls show substantial price increases in 1967 as compared to 1952. In addition, the Bell total toll price index grows more slowly between 1967 and 1977 than either of our toll price indices. We consider that the Bell Canada total toll index shows the correct direction of change. How then do we account for our finding of rising nominal prices when falling prices actually occurred?

The reason is simple - the weights used in the analysis are incorrect weights for the entire 1952 to 1978 period. Two sets of weights are used First, to aggregate calls of different length within one tariff schedule (Intra Bell, Trans-Canada, U.S.), we use either geometric weights or 'reverse mileage' weights. These weighting schemes may not reflect the actual distribution of the calls in any given year. Second, and likely of greater bias, in aggregating the various forms of calls (station-to-station, person-to-person,

Table 2.3

BELL CANADA MESSAGE TOLL PRICE INDEX

1952	1.04716
1953	1.04840
1954	1.04917
1955	1.04862
1956	1.04782
1957	1.04452
1958	1.05384
1959	1.09930
1960	1.11315
1961	1.10193
1962	1.04134
1963	1.03961
1964	1.03866
1965	1.03772
1966	1.00978
1967	1.00000
1968	.991215
1969	.994918
1970	1.07704
1971	1.09413
1972	1.11011
1973	1.13243
1974	1.14576
1975	1.18689
1976	1.25437
1977	1.28836
1978	1.35508
	1

intra Bell, Trans-Canada, U.S.; night, day) we used revenue share weights based on the 1967 distribution of revenue. But, it is clear that the large change in relative prices for various types of toll calls induced substantial shifts in toll calling patterns so that the 1967 revenue distribution by call was very different from that distribution of revenues in 1977 or 1952. We would expect the revenue distribution to shift away from the relatively expensive calls to the calls which because of tariff changes become relatively inexpensive. Indeed, the Bell Price index shows that the movement was substantial - the nominal price of an "average" toll call fell. As a result, toll price indices based on a single year's weights are misleading. We conclude that the toll price indices developed in Tables 2.1 and 2.2 for business and residential traffic are misleading. They are misleading because of the untenable assumption that the revenue shares of various types of toll calls remained constant over the perid. We therefore attempted to determine a method to generate the revenue share weights for years other than 1967.

2.3 Demand for Toll Calls by Distance and Length of Call

Data exists for four years showing the distribution of toll calls (all types - intra Bell, adjacent provinces, TCTS and other) for Bell Canada by mileage band and duration of the call. The data for 1975 are shown in Table 2.4. These data, if properly analyzed, could indicate two crucial points - first, the price elasticity of demand could be determined by mileage band and duration of call; second, the change in call distribution could be used to provide information on the changing revenue distribution (the changing weights needed to construct toll price indices). The method of analyzing demand characteristics from these data is not however transparent and requires careful discussion.

TABLE 2.4

COMPANY ϕ - 1975 (BELL CANADA)

ESTIMATED LONG DISTANCE MESSAGES (ALL TYPES) ACCORDING TO DISTANCE AND DURATION

(on an annual basis - in millions)

Rate Mileage	Duration in Charged Minutes						Total
	1 min.	2 min.	3 min.	4 min.	5 min.	6 min. & Over	Minutes Charged
0- 10	6.4	3.5	2.0	1.2	0.9	3.4	71.0
11- 14	12.2	7.1	4.1	2.6	1.7	5.9	125.3
15- 22	24.6	15.2	9.4	6.1	4.2	15.2	303.8
23- 30	15.1	10.0	6.5	4.4	3.1	11.8	222.3
31- 40	13.4	9.3	6.3	4.4	3.2	12.8	231.9
41- 50	7.6	5.5	3.9	2.7	2.0	8.6	151.9
51- 60	6.1	4.6	3.3	2.4	1.8	8.0	138.2
61- 80	6.5	5.0	3.7	2.7	2.1	10.0	169.0
81-100	4.2	3.3	2.5	1.8	1.5	7.3	123.9
101-130	5.4	4.1	3.1	2.4	1.9	9.7	163.2
131-160	3.5	2.6	2.0	1.5	1.2	6.3	106.3
161-200	2.1	1.7	1.4	1.1	0.9	5.0	83.6
201-250	3.8	2.9	2.3	1.7	1.4	8.4	142.8
251-300	1.5	1.2	0.9	0.7	0.6	4.1	68.2
301-400	5.9	4.6	3.7	2.9	2.3	13.6	229.7
401-500	1.5	1.2	1.0	0.7	0.6	4.0	67.7
501-Over	4.3	3.5	2.9	2.5	2.2	14.7	237.8

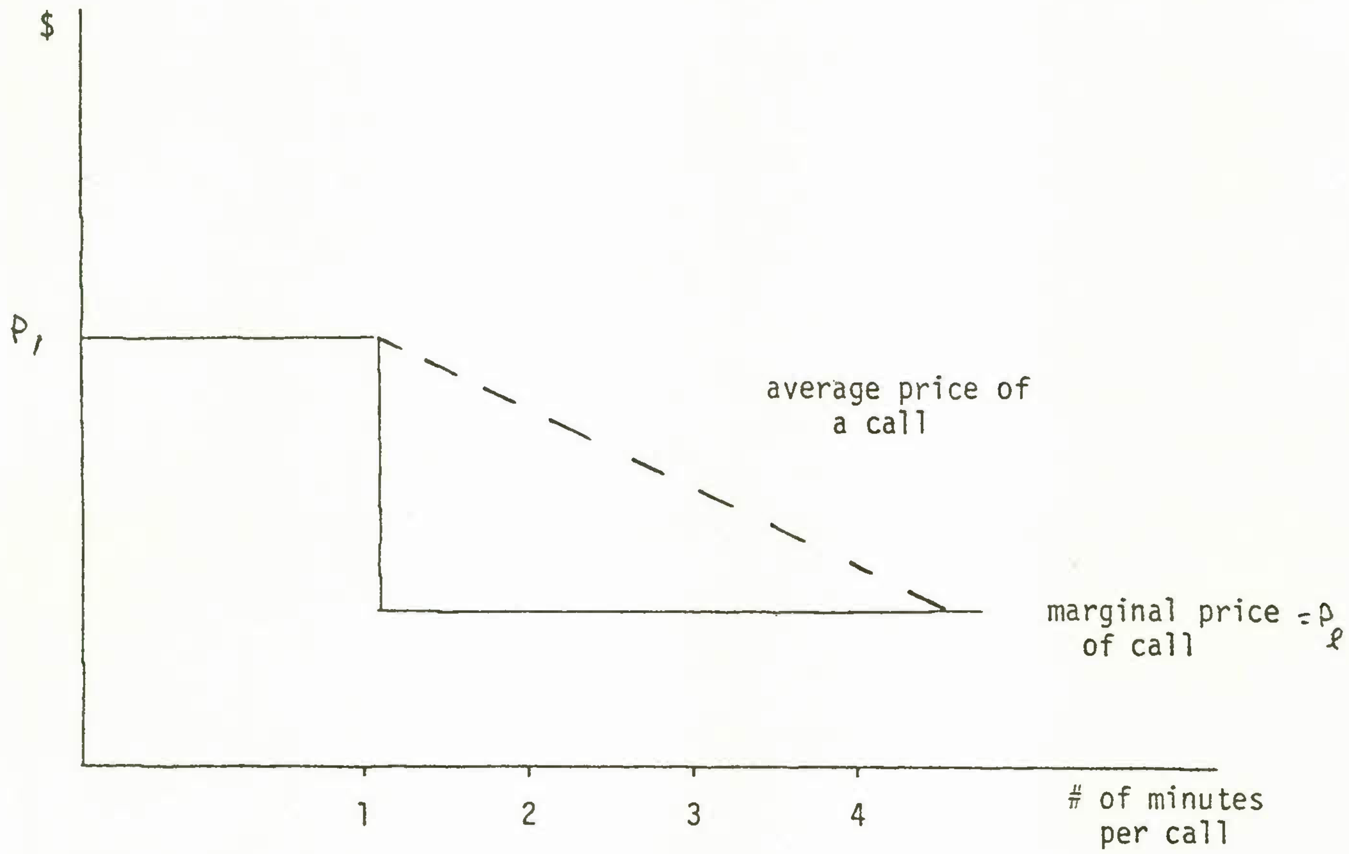
ϕ A breakout by province is not available

P(Q) 6 Jan. 77-58

Sheet 6 of 6

Figure 1

Individual Demand for Minutes per Call



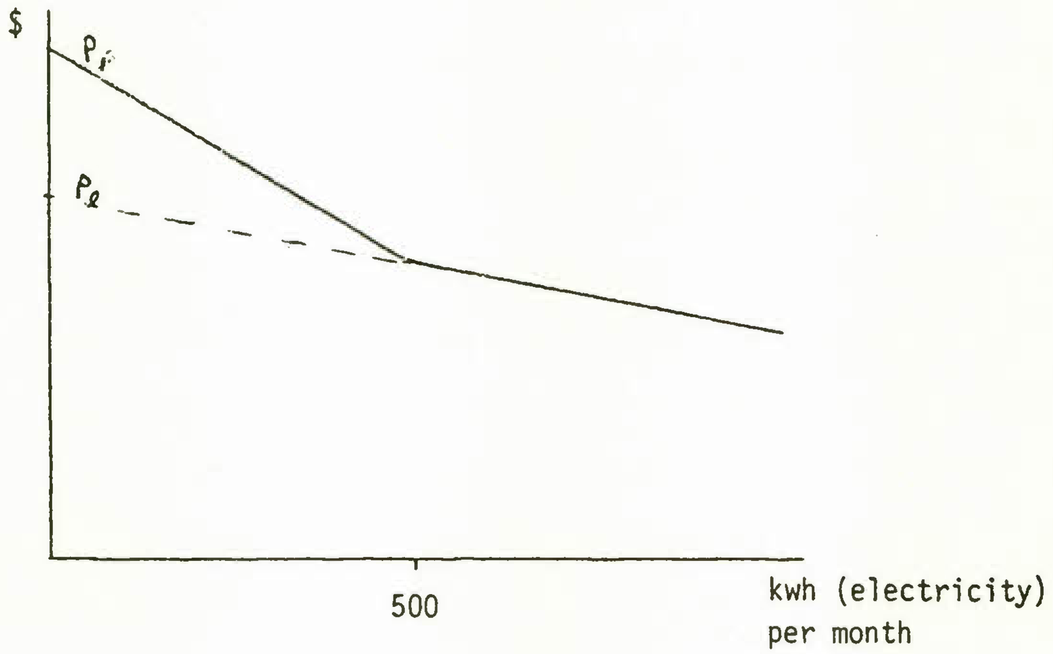
Declining Block Tariffs

The per minute rates for toll calls are not constant; instead the price schedule generally is one of a declining block tariff as shown in Figure 1. The quantity measured on the horizontal axis is the number of minutes per call demanded by an individual. A one minute call has a certain price (p_1) associated with it. After one minute has passed on each call, the marginal price falls to p_ℓ . The average price of a one minute call then declines as the price of the first block is amortized over more and more additional minutes. Every toll call involves, at least, a one minute call. Either the toll call lasts only one minute or a longer call is made, involving as a first step, a one minute call. The one minute call is then the entry price that must be paid in order to receive lower marginal and average prices for calls of greater time duration. The price schedule for toll calls is then a declining block tariff. For Bell Canada, the initial block is now one minute; before 1970 it was three minutes. Each additional minute beyond the initial block involves a constant marginal price, a price lower than the first block.

One cannot draw an exact analogy between analyzing the demand for toll calls under a declining block tariff and the amount of electricity demanded under a declining block tariff. In Figure 2, we show the aggregate monthly demand for electricity by a household where the price schedule has a break at 500 kwh; electricity consumed after that point facing a lower marginal price (p_ℓ).

Figure 2

Monthly Demand for Electricity



Households consume a certain amount of electricity per month and this will be either more or less than 500 kwh (but not both). We cannot however in a similar fashion aggregate the total monthly number of toll call minutes of a household in a given time period and draw a diagram analogous to Figure 2, since individual's make one and two and three and n minute calls in a month. Each of these calls involves the use of a price schedule as in Figure 1, but the total call minutes aggregated for a household over some time period such as a month are not themselves charged under a declining block tariff (as are total electricity hours consumed per month); each individual call faces a declining block tarrif. The total number of toll call minutes demanded is an inverse function of the price charged. But, there is no unique quantity on the horizontal axis of Figure 2 (total number of call minutes demanded per month) which represents the point at which the marginal price per minute changes and as a result leads to a kink in this monthly demand schedule for toll call minutes.

In Figure 3 and 4 we indicate the aggregate demand curves in any given time period for the total number of call minutes similar to the demand curve for electricity consumed in a month) and for the number of calls of any n minute duration. These demand curves are not kinked. If the marginal price after the first minute is decreased by the firm, people may shift from making one minute calls to calls of longer duration, but in aggregate the number of call minutes (Figure 3) will be increased. The demand curve for total call minutes is therefore downward sloping to the right. It is clear however, that the demand for one minute calls (Figure 4) can decrease when the price beyond the first block falls, since a longer call is a substitute for a one minute call. However, if the price of one minute calls falls, the demand for one minute calls will increase. (The demand curve for one minute calls or any n minute calls is downward sloping to the right.)

The analysis of the demand for toll calls would be identical to that of the demand for electricity if a different multi-part pricing system were practised by Bell Canada. A two-part price schedule which was related to a customer's total call minutes in a month would lead to a kinked aggregate demand curve for total call minutes in a month. For example, using Figure 1 to represent this hypothetical demand for total monthly call minutes under this alternative pricing system, all call minutes where the total demand

Figure 3

Monthly Demand for Call Minutes

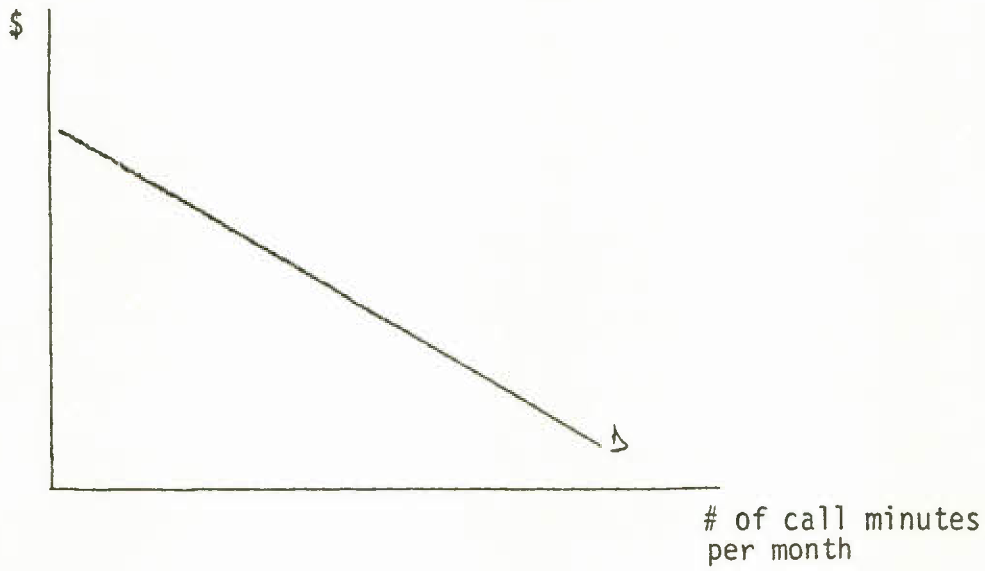
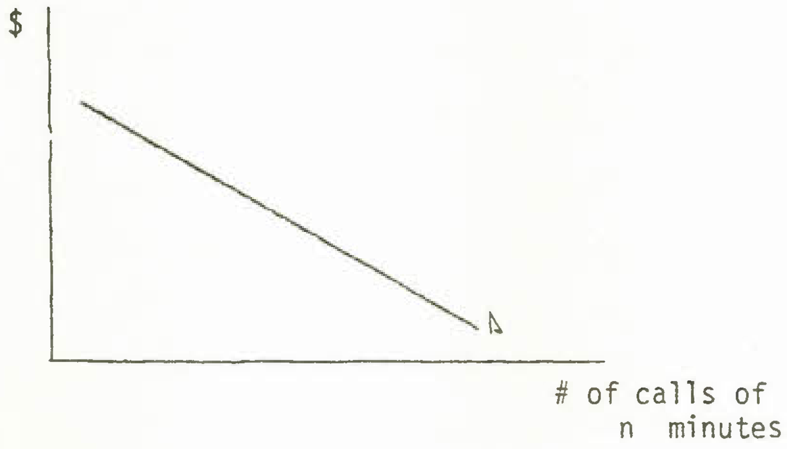


Figure 4

Demand for Calls, Each of n Minutes



was less than 500 minutes per month would face price p_1 ; all additional call minutes would be charged at a lower marginal price p_ℓ .

Therefore an exact equivalence does not exist between modelling the telephone call minute data as given in Table 2.4 and existing analyses of electricity demand. However, much of the general development is the same.

Partition the demand for call minutes by the length of the call, x_1 being the demand for one minute calls, x_2 the demand for two minute calls etc. Two marginal prices exist, p_1 for calls of one minute in length and a lower marginal price p_ℓ for each additional minute beyond the first minute. The demand for one minute calls is assumed to be a function of p_1 . The number of two minute calls, x_2 yield $2x_2 = X_2$ call minutes, the first x_2 minutes of these charged at p_1 , the second x_2 minutes charged at p_ℓ . Lower case symbols will represent the number of calls in the block, upper case the number of call minutes. For a customer making a two minute call, an increase in the price of a one minute call (p_1) will be similar to a change in income, as long as the customer does not decrease the length of his call to one minute. As a result, the demand function for two minute calls (relating quantity consumed and price) must include as explanatory variables, the marginal price facing the customer in the second minute block (p_ℓ) as well as the intramarginal expenditure on toll calls in the first block $((p_1-p_\ell)x_2)$ - the expenditure necessary to get into the second block. The lower per month price for the second minute of a two minute call can only be received once the higher per minute price has been paid for the first minute call. The difference in price for the first minute call and the lower price for an additional minute call is in essence a tax that must be paid to receive this lower price. The coefficient on this intramarginal expenditure in a demand equation for any duration call beyond one minute should be equal in magnitude but opposite in sign to the coefficient on income. A change in the price for a one minute call for customers making an n minute call is a change in the intramarginal tax, i.e., a change in "net" income.

For one minute calls, a demand equation could be written as:

$$x_1 = D'(p_1, E) \qquad (5)$$

where x_1 = demand for one minute calls

p_1 = price of one minute calls relative to the CPI

E = expenditure on all toll calls

For a two minute call, a demand equation could be written as:

$$x_2 = D^2(p_\ell, E^*, E) \qquad (6)$$

where p_ℓ = marginal price for an additional minute

E^* = marginal expenditure in the first block

E = expenditure on all toll calls

E^* measures the cost of paying the higher price in the first block in order to make a call of two minutes in duration. As noted above

$$E^* = (p_1 - p_\ell)x_2 \qquad (7)$$

In general, demand equations for calls of n minutes in duration have the form:

$$x_n = D^n(p_\ell, E^*, E) \qquad n \neq 1 \qquad (8)$$

As discussed above, the coefficient on E^* should be equal in magnitude but opposite in sign to the coefficient on E. The data available to us differentiates calls not only by the duration of the call but also by the distance of the call. In our analysis we examine whether demand characteristics vary according to the distance of the call. We also utilize share equations, attempting to explain the share of toll expenditures in a year (t) on a call, where the call is denoted by both the duration of the call (i) and the distance of the call (j) .

The demand equations estimated were of the following form:

$$S_{ijt} = \alpha_{ij}D_{ij} + \beta_i \ln p_{ijt} + \gamma_i \ln E^*_{ijt} + \delta_i \ln E_t \qquad (9)$$

where i = mileage band

j = time band

t = year

S_{ijt} = share of expenditure on the (i,j) block in total toll expenditures

E^*_{ijt} = marginal expenditure up to marginal block

E_t = total expenditure on toll calls

D_{ij} = dummy variable for the ith duration and the jth mileage band

p_{ij} = marginal price in the (i,j) block

Price and income (expenditure) elasticities can be calculated from these share equations, as follows.

Price Elasticities

First Block (Marginal expenditure is zero)

$$E_{x,p_x} = \frac{\beta_i}{S_{ij}} - 1 \qquad (10)$$

Other Blocks (Marginal expenditure is positive)

The expenditure shares can be written in the form:

$$S = \frac{p_\ell X + (p_1 - p_\ell)x}{E} \qquad (11)$$

where p_ℓ is the marginal price in the block

X is the number of call minutes in the block

x is the number of calls in the block

$X = nx$, where n is the number of minutes a call lasts

p_ℓ is the marginal price in the intramarginal block

$(p_1 - p_\ell)x$ is the marginal expenditure (the amount paid over and above p_ℓ to get into the marginal block)

and the i,t subscripts have been removed for simplicity.

The shares can be expressed in the compact notation.

$$S = \frac{M + M_\ell}{E} = \frac{Z}{E} \qquad \text{where } M_\ell = (p_1 - p_\ell)x \text{ and } Z = M + M_\ell \qquad (12)$$

The own price elasticities of demand can be obtained as follows. From (12) we can form the equation

$$\ln S = \ln Z - \ln E \qquad (13)$$

$$\frac{1}{S}\frac{\partial S}{\partial \ln p_\ell} = \frac{\partial \ln Z}{\partial \ln p_\ell}$$

$$= \frac{p_\ell}{Z}\left(X + p_\ell \frac{\partial X}{\partial p_\ell} + \frac{\partial M_\ell}{\partial p_\ell}\right)$$

$$= \frac{p_\ell}{Z}\left(X + p_\ell\left(\frac{\partial X}{\partial p_\ell} + \frac{\partial X}{\partial M_\ell} \cdot \frac{\partial M_\ell}{\partial p_\ell}\right) + \frac{\partial M_\ell}{\partial p_\ell}\right)$$

$$= \frac{p_\ell \cdot X}{Z} + \frac{p_\ell}{Z} \cdot \varepsilon_{X,p_\ell} - \frac{x(p_\ell)^2 X}{Z \cdot M_\ell} \cdot \varepsilon_{X,M_\ell} - x\frac{p_\ell}{Z}$$

where $\varepsilon_{X,M}$ is the elasticity of X with respect to M_ℓ

$$\text{Hence } \varepsilon_{X,p_\ell} = \frac{p_\ell \cdot X + M_\ell}{p_\ell \cdot X} \cdot \frac{\beta_i}{S} - 1 + \frac{p_\ell \cdot x}{M_\ell} \cdot \varepsilon_{X,M_\ell} + \frac{x}{X} \qquad (14)$$

By the same procedure as above we can show that

$$\varepsilon_{X,M_\ell} = \frac{Z}{p_\ell \cdot X} \left(\frac{\gamma_i}{S} - \frac{M_\ell}{Z} \right) \tag{15}$$

which can be substituted into (14) to obtain the price elasticities.

Note that since $X = nx$; $\varepsilon_{X,p_\ell} = \varepsilon_{x,p_\ell}$

and $\varepsilon_{X,M_\ell} = \varepsilon_{x,M_\ell}$

<u>Income (Expenditure) Elasticities</u>

Equation (15) provides the marginal expenditure elasticities.

Since demand in a block is a function of marginal price and income available for expenditure in the block.

$$x = D(p_\ell, E-M_\ell) \tag{16}$$

Let $E-M_\ell = Y$

$$\frac{\partial \ln x}{\partial \ln M_\ell} = \frac{\partial \ln x}{\partial \ln Y} \cdot \frac{\partial \ln Y}{\partial \ln M_\ell} = -\frac{\partial \ln x}{\partial \ln Y} \tag{17}$$

$$\frac{\partial \ln x}{\partial \ln E} = \frac{\partial \ln x}{\partial \ln Y} \cdot \frac{\partial \ln Y}{\partial \ln E} = +\frac{\partial \ln x}{\partial \ln Y} \tag{18}$$

Comparing (13) and (14) we see that the elasticity of toll calls with respect to total toll expenditures is equal in magnitude and opposite in sign to the marginal expenditure elasticity.

We attempted to estimate the share equations as given in (9) and to compute price and income elasticities [(14) (17)]. We began by examining each of the 17 individual mileage bands. For each band we had 24 observations available (four years 1972 through 1976 and 6 different durations of calls - 1, 2, 3, 4, 5, 6 and over minutes). We assumed that each observation was a point on a single demand curve. These attempts proved unsuccessful, and we pooled mileage bands as well, ending up with four divisions - mileage bands 0-30 miles, 31-80 miles, 81-200 miles and over 200 miles. Each of these pooled regressions then contains 96 observations (4 years, 6 time slots, 4 mileage bands) except the last which contains 120 observations. The results for these regressions are shown as column A, Table 2.5 below. Several of these elasticities are the wrong sign but only one of the elasticities is significantly different from zero.

TABLE 2.5

PRICE AND INCOME ELASTICITIES, MILEAGE BAND DATA

Mileage Band	Price Elasticity		Income Elasticity	
	A	B	A	B
1- 30	-.313	-1.416	.150*	-.108
31- 80	-.658	- .609*	.01	.225*
81-200	-.935	- .638*	-.02	.334*
200 & over	.251	-3.213*	.221	.663*

Elasticities evaluated at the means

A. static share model

B. adjustment model

* significantly different from zero at 5% level.

We altered the model by introducing the lagged share as an explanatory variable, thus incorporating a Koyck adjustment process, namely:

$$S_{ij,t} = \alpha_0 S_{ij,t-1} + \alpha_1 D_{ijt} + B_i \ln p_{ijt} + \gamma_i \ln E^*_{ijt} + \delta_i \ln E_t \qquad (19)$$

The resulting elasticities are shown in columns B in Table 2.5. Seven of the eight elasticities are of the expected sign, while 6 of the 8 elasticities are significantly different from zero. If we ignore the results for the first four mileage bands, where the elasticities are insignificant, both price and income elasticities increase with distance - an interesting finding. However, the results are not sufficiently robust to be used in analyzing the changing mix of calls, and we turned to an examination of demand elasticities based on more aggregate data.

2.4 The System of Demand Equations Based on Aggregate Data

We begin by assuming that the long-run demand equation for output i is of the form:

$$\log \hat{Q}_{it} = a_i + \sum_j b_{ij}\left(\frac{x_{jt}^{\lambda}-1}{\lambda}\right) + \sum_k c_{ik} z_{\kappa t} \qquad i = 1,2,3 \qquad (20)$$

where x_{jt} is the j-th exogenous variable. The above specification transforms the exogenous variables x_{jt} in accordance with the Box-Cox transformation in order to generalize the logarithmic or linear specification. If $\lambda = 0$, $\frac{x_{jt}^{\lambda}-1}{\lambda} = \log x_{jt}$. If $\lambda = 1$, $\frac{x_{jt}^{\lambda}-1}{\lambda} = x_{jt} - 1$, leading to a linear specification. In general, λ is a parameter to be estimated. the variables z_{kt} may be either linear or logarithmic.

We now present the list of x and z variables used in the three demand equations.

Monopoly Toll Long Distance Service:

Q_1 = constant dollar message toll revenue per capita

x_1 = price of message toll service divided by the CPI

x_2 = income per capita divided by the CPI

x_3 = number of telephones in Bell Canada's territory

z_1 = percentage of telephones with access to direct distance dialing facilities (in linear form)

Competitive Toll Service

Q_2 = constant dollar competitive toll revenue per unit real domestic product

x_1 = price of competitive toll service divided by the Real Domestic Product (RDP) deflator for Ontario and Quebec

x_2 = real domestic product (Ontario and Quebec)

Local Service:

Q_3 = constant dollar local service revenue per capita

x_1 = price of local service divided by the consumer price index (CPI)

x_2 = income per capita divided by the CPI

x_3 = households in the Bell Canada territory

We assume that only partial adjustment to long-run demands $\hat{Q}_i$ occurs within one year. We assume further that the partial adjustment mechanism can be written in the form

$$\log Q_{it} - \log Q_{i,t-1} = \Theta_i(\log \hat{Q}_{it} - \log Q_{i,t-1}) \tag{21}$$

The final form of the demand equations estimated can be obtained (by substituting (20) into (21)) as

$$\log Q_{it} = (1-\Theta_i)\log Q_{i,t-1} + \Theta_i a_i + \sum_j \Theta_i b_{ij}\left(\frac{x_{jt}^{\lambda}-1}{\lambda}\right) + \sum_k \Theta_i c_{ik} z_{kt} \tag{22}$$

$$i = 1,2,3$$

Equation (22) was estimated by searching over λ . Table 2.6 presents the results of the search procedure using data for 1952-78. The short-run own price elasticities of demand can be shown to be

$$SRE_i = \Theta_i b_{i1} p_i^{\lambda} \qquad i = 1,2,3 \qquad (23)$$

where p_i = the relative price of the i-th service .

The long-run price elasticities can be calculated as

$$LRE_i = b_{i1} p_i^{\lambda} \qquad i = 1,2,3 \qquad (24)$$

From Table 2.6 it can be seen that the likelihood function is very flat over a wide range of values of λ ; and these values of λ imply a wide range of elasticity estimates.

The cost model (to be presented in Chapter 4) contains, as parameters, elasticities of demand for monopoly and competitive toll services. In the hope that we could improve the accuracy of the estimated demand elasticities, we turned to simultaneous estimation of the cost and demand model. In this chapter we present parameter estimates for the demand model. The cost model estimates are presented in Chapter 4.

Computer software limitations precluded the estimation of λ , so we specified $\lambda=0$ which implies that $\frac{x_{jt}^{\lambda}-1}{\lambda} = \log x_{jt}$. Based on our results and those of the Concordia group, there is no evidence that such a restriction is unwarranted. Also due to computer software limitations, we have estimated the local services demand equation separately. This equation has no parameters in common with the rest of the demand/cost system. However

Table 2.6

Parameter and Elasticity Estimates for Chosen Estimate of Demand Equations

	Estimate	Standard Error
$\Theta_1 a_1$	-2.67	0.42
$\Theta_1 b_{11}$	-0.97	0.17
$\Theta_1 b_{12}$	0.04	0.01
$\Theta_1 C_{11}$	0.06	0.04
Θ_1	0.37	0.09
$\Theta_2 a_2$	0.37	0.04
$\Theta_2 b_{21}$	-0.12	0.24
$\Theta_2 b_{22}$	0.00	0.00
Θ_2	0.94	0.02
$\Theta_3 a_3$	-6.36	3.34
$\Theta_3 b_{31}$	-0.21	0.06
$\Theta_3 b_{32}$	0.009	0.006
$\Theta_3 b_{33}$	-19.1	10.1
Θ_3	0.96	0.04
SRE_1	-1.19	0.21
LRE_1	-1.91	0.17
SRE_2	-0.11	0.23
LRE_2	-1.74	3.85
SRE_3	-0.22	0.07
LRE_3	-5.16	2.65

there will be some loss of efficiency, if as is likely, the error terms are contemporaneously correlated with the error terms in the rest of the system.

Preliminary estimation indicated that serial correlation was a problem in the demand side of the model. Hence we specified the error structure to be that generated by a first order autoregressive process. The demand system estimated became

$$\log Q_{it} = (1-\Theta_i+\rho_i)Q_{i,t-1} + \Theta_i(1-\rho_i)a_i + \sum_j \Theta_i b_{ij}(\log x_{it} - \rho_i \log x_{i,t-1})$$

$$+ \sum_k \Theta_i c_{ik}(z_{kt} - \rho_i z_{k,t-1}) \qquad i = 1,2,3 \qquad (25)$$

where ρ_i is the autocorrelation coefficient for the ith equation. The simultaneous cost/demand system was estimated using iterative 3SLS . A list of the instrumental variables is presented in Chapter 4.

The following hypotheses could not be rejected at conventional significance levels, and were imposed to improve the accuracy of the estimated price elasticities:

(i) $b_{13} = c_{11} = 0$

(ii) $\Theta_1 = 1$ (full adjustment in 1 year - monopoly toll)

(iii) $b_{22} = 0$ (unitary output elasticity for competitive toll)

Table 2.7 presents the parameter estimates for the demand system. The summary statistics are contained in Table 2.8. Table 2.9 presents the estimated short and long run price and income (output) elasticities. It can

Table 2.7

Parameter Estimates - Demand Model

(Standard Errors in Brackets)

Parameter	Estimate	Parameter	Estimate
a_1	-5.070 (0.095)	a_3	-2.170 (0.966)
b_{11}	-1.386 (0.081)	b_{31}	-0.279 (0.278)
b_{12}	1.262 (0.110)	b_{32}	0.465 (0.244)
ρ_1	0.702 (0.081)	b_{33}	1.020 (0.684)
a_2	-6.201 (0.110)	θ_3	0.580 (0.211)
b_{21}	-2.047 (0.391)	ρ_3	0.916 (0.063)
θ_2	0.872 (0.182)		
ρ_2	0.795 (0.048)		

Table 2.8

Summary Statistics

Equation	R^2	D.W. Statistics*
Monopoly Toll	0.9977	1.38
Competition Toll	0.9936	2.33
Local	0.9996	1.53

* Note that in the case of lagged endogenous variables the D.W. Statistic is biased towards 2.

Table 2.9

Price and Income (Output) Elasticities

Service	Price		Income		Output	
	SR	LR	SR	LR	SR	LR
Local	-0.137	-0.279	0.270	0.465	--	--
Monopoly Toll	-1.386	-1.386	1.262	1.262	--	--
Competitive Toll	-1.784	-2.047	--	--	0.870	1.000

be seen from Table 2.9 that the long-run demands for monopoly and competitive toll services are elastic while the demand for local service is inelastic. An inelastic demand for local service indicates that Bell is not exercising its monopoly power in local services to maximize profits. In the cost model developed in Chapter 4 we assume that this result is due to the fact that the basic local service price is constrained by the regulatory commission. The estimated income elasticities indicate that local service is a "necessary" good, since the income elasticity is less than unity, while monopoly toll is a "luxury" good (income elasticity greater than unity). Hence any equity weighting of the rate structure in favour of lower income groups is likely to lead to a moderation of those local service price increases which would be dictated by the application of a pure efficiency standard.

Footnotes to Chapter Two

1. Baumol (1971), p.

2. See Davis et al (1973).

3. Charles River Associates, "The Economics of Competition in the Telecommunications Industry", August 1979.

4. C.T.C. Dossier C.955.183, Bell Canada P(Q) 27 juin, 75-50; Bell Canada P(Q) 6 Jan. 77-58, updated.

5. There are a number of other discounts in some years e.g. for after midnight calls, that were ignored because the revenue shares were small in those time zones.

Chapter Three

THE COST OF CAPITAL FOR A REGULATED UTILITY

A. Derivation of the User Cost of Capital Services and the Allowed Gross Return on Capital Services for a Regulated Utility

In this chapter we derive expressions for the user cost of capital services and the allowed gross return on capital services for a regulated utility which are consistent with the neoclassical model of capital accumulation, and which take into account the following institutional details:

(1) the taxation system: corporate taxation rates and the personal tax rates of the utility's shareholders differ.

(2) financial capital is raised from multiple sources - retained earnings, bond issues, common and preferred stock issues - with different implications for the calculation of corporate income taxes.

(3) depreciation for tax purposes may be in excess of true economic depreciation.

(4) the firm is subject to rate of return regulation which places an upper limit, ex ante, on the allowed gross return on capital.

The model developed in this section is an extension, to take account of the above factors, of the model proposed by Boadway and Bruce (1979). They suggested that the cost of capital services be derived from an intertemporal model which analyses the maximization of a consumer's preferences subject to the constraints imposed by his ability to borrow funds and the production activities of the firm in which he has an ownership interest.

We begin with some notation. Let

r_s = the interest rate at which a shareholder can borrow on a personal basis

r_B = the (long-term) bond rate at which the utility can borrow

r_E = the utility's cost of equity capital (a weighted average of common and preferred)

θ = the proportion of the utility's liabilities held in the form of debt (assumed exogenous)

U_c = corporate tax rate

U_p = personal tax rate on dividends

It is assumed that the shareholder wishes the firm's intertemporal production plan to be chosen so as to maximize the present value of the utility of his dollar consumption stream (C):

$$\int_0^{\infty} e^{-\gamma t} \cdot U(C)dt \qquad (1)$$

where γ is the (instantaneous) rate of time preference of the share holder.

The consumption stream C consists of the sum of: (a) after tax income from all sources except dividends from utility ownership, (b) after tax dividends from equity owned in the utility, and (c) the change in personal debt; less interest payments on previously accumulated personal debt. That is, the consumption stream is

$$C = Y + (1 - U_p)V + \dot{B} - r_s B = Y + T_p \cdot V + \dot{B} - r_s B \qquad (2)$$

where Y = after tax income from all sources except dividends from the utility (assumed exogenous)

V = dividends from equity ownership

B = personal debt

$T_p = 1 - U_p$

and the dot over a variable indicates the change in the variable per unit time.

We now develop an expression for the flow of dividends from utility ownership. Suppose the production function can be written in the implicit form

$$F(\underline{Q}, \underline{x}, K) = 0 \tag{3}$$

where $\underline{Q}$ is a vector of outputs, $\underline{x}$ is a vector of expensed factor inputs and K is the capital stock. At any point in time, conditional on K, the firm chooses $\underline{Q}$ and $\underline{x}$ so as to maximize variable profits subject to the production function (3). Denoting the maximizing values as $\underline{Q}^*$, $\underline{x}^*$ we obtain the variable profit function[1]

$$F(K) = [\underline{p}(\underline{Q}^*(K))] \cdot \underline{Q}^*(K) - \underline{w} \cdot \underline{x}^*(K) \tag{4}$$

where $\underline{p}$ is a vector of output prices and $\underline{w}$ is a vector of the prices of inputs other than capital. Note that $F(K)$ is just the difference between operating revenues and those operating costs not associated with real capital (depreciation and corporate income taxes). The flow of dividends from firm ownership can be written as

$$V = F(K) - qI - (\theta r_B + (1-\theta)r_E)A - U_c[F(K) - qD - r_B\theta A] + \dot{A} \qquad (5)$$

where A = corporate liabilities (equity at market value, debt at face value)

I = real investment in the capital stock

q = asset price of capital

D = depreciation for tax purposes

Equation (5) states that dividends accruing to an individual shareholder equals maximized variable profit less investment expenses less financial expenses[2] less income taxes, plus any increase in liabilities not used for other purposes.

Depreciation for tax purposes is based on an "accounting" stock of capital $\hat{K}$ which will differ from the "economic" stock of capital K when accelerated depreciation schedules are used. Let the rate of depreciation for tax purposes be α and the rate of economic depreciation be δ . Then real investment I can be expressed in either of the two forms:

$$I = \dot{K} + \delta K \qquad \text{or} \qquad (6)$$

$$I = \dot{\hat{K}} + \alpha\hat{K} \qquad (7)$$

Rearranging (5) into the form

$$V = F(K)\cdot T_c - q(I - U_c D) - [\theta T_c r_B + (1-\theta) r_E]A + \dot{A} \tag{8}$$

and using (7),we can write the consumption flow as

$$C = Y + T_p[F(K)\cdot T_c - q(\dot{\hat{K}} + \alpha\hat{K}T_c) - (\theta T_c r_B + (1-\theta) r_E)A + \dot{A}] - r_s B + \dot{B} \tag{9}$$

where $T_p = 1 - U_p$

The utility management's problem is to maximize (1) subject to (9) and three additional constraints which we will now consider: (i) a real capital accumulation constraint, (ii) a liabilities accumulation constraint and (iii) a rate of return constraint.

The real capital accumulation constraint can be expressed as (using (6) and (7))

$$\dot{K} + \delta K - \dot{\hat{K}} - \alpha\hat{K} = 0 \tag{10}$$

We will assume that the regulatory authorities will not allow the utility to engage in liabilities accumulation in excess of net capital accumulation,[3] which implies

$$\dot{A} - q\dot{K} \leq 0 \quad . \tag{11}$$

Finally, we form the rate of return constraint. Variable profits, or gross net returns, must be no greater than bond financing costs plus the allowed return on equity plus depreciation plus income tax liabilities. Hence the rate of return constraint can be written as

$$F(K) \leq [\theta r_B + (1-\theta)s_E]\cdot A + \delta qK + U_c[F(K) - q\delta K - r_B\theta A] \qquad (12)$$

where s_E = allowed rate of return on equity.

Note that the allowed income tax liabilities are calculated assuming economic depreciation, so that when depreciation for tax purposes exceeds economic depreciation the regulated utility is permitted to accumulate a reserve for deferred income taxes.

Constraint (12) can be rewritten in the form:

$$T_c\ F(K) \leq [(1-\theta)s_E + \theta T_c r_B]\cdot A + T_c\delta qK \qquad (13)$$

In order to maximize its shareholders' present value of consumption the utility's management must maximize (1) subject to (9), (10), (11), and (13). This problem is an optimal control problem in which K, $\hat{K}$, B, and A are state variables, and C, $\dot{K}$, $\dot{\hat{K}}$, $\dot{B}$ and $\dot{A}$ are control variables. A solution can be obtained using Pontryagin's maximum principle. We first form the Lagrangian expression

$$\begin{aligned} L = {} & [U(C) + \lambda_o\dot{K} + \lambda_1\dot{B} + \lambda_2\dot{\hat{K}} + \lambda_3\dot{A}]e^{-\gamma t} \\ & - \phi_o\Big\{C - Y - T_p \cdot \Big[F(K)T_c - q(\dot{\hat{K}} + \alpha\hat{K}T_c) - [\theta T_c r_B + (1-\theta)r_E]A + \dot{A}\Big] + r_s\cdot B - \dot{B}\Big\} \\ & - \phi_1\{\dot{K} + \delta K - \dot{\hat{K}} - \alpha\hat{K}\} \\ & - \phi_2\{\dot{A} - q\dot{K}\} \\ & - \phi_3\{T_cF(K) - T_c\delta qK - [(1-\theta)s_E + \theta T_c r_B]A\} \end{aligned} \qquad (14)$$

where λ_o, λ_1, λ_2, and λ_3 are co-state variables and ϕ_o, ϕ_1, ϕ_2, and ϕ_3 are Lagrangian multipliers.

The first order conditions for a maximum are:[4]

$$\frac{\partial L}{\partial C} = U'(C)e^{-\gamma t} - \phi_o = 0 \tag{15}$$

$$\frac{\partial L}{\partial \dot{K}} = \lambda_o e^{-\gamma t} - \phi_1 + q\phi_2 = 0 \tag{16}$$

$$\frac{\partial L}{\partial \dot{B}} = \lambda_1 e^{-\gamma t} + \phi_o = 0 \qquad \text{or} \qquad \lambda_1 = - \phi_o e^{\gamma t} \tag{17}$$

$$\frac{\partial L}{\partial \hat{K}} = \lambda_2 e^{-\gamma t} - qT_p\phi_o + \phi_1 = 0 \qquad \text{or} \qquad \phi_1 e^{\gamma t} = qT_p\phi_o e^{\gamma t} - \lambda_2 \tag{18}$$

$$\frac{\partial L}{\partial \dot{A}} = \lambda_3 e^{-\gamma t} + \phi_o T_p - \phi_2 = 0 \tag{19}$$

It is convenient to use (15) - (19) to obtain a differential equation in the co-state variables. This is obtained through the following series of manipulations. Using (15) and (17) we obtain

$$U'(C) + \lambda_1 = 0 \tag{20}$$

Combining (16), (17) and (18) yields

$$\lambda_o + qT_p\lambda_1 + \lambda_2 + q\phi_2 e^{\gamma t} = 0 \tag{21}$$

Subtracting (17) from (19) yields

$$\lambda_3 e^{-\gamma t} + \phi_o T_p - \phi_2 - \lambda_1 e^{-\gamma t} - \phi_o = 0 \qquad (22)$$

$$\lambda_3 e^{-\gamma t} + \phi_o T_p - \phi_2 - \lambda_1 e^{-\gamma t} - \phi_o = 0$$

$$\lambda_3 + \phi_o e^{\lambda t}(T_p - 1) - \phi_2 e^{\gamma t} - \lambda_1 = 0$$

$$\lambda_3 - \lambda_1 (T_p - 1) - \phi_2 e^{\gamma t} - \lambda_1 = 0$$

$$\lambda_3 - T_p \lambda_1 - \phi_2 e^{\gamma t} = 0$$

$$q\lambda_3 - qT_p \lambda_1 - q\phi_2 e^{\gamma t} = 0 \qquad (23)$$

Adding (21) and (23) we obtain

$$\lambda_o + q\lambda_3 + \lambda_2 = 0 \qquad (24)$$

Differentiating (24) w.r.t. time provides us with the required differential equation:

$$\dot{\lambda}_o + \dot{\lambda}_2 + q\dot{\lambda}_3 = 0 \qquad (25)$$

The Euler necessary conditions are obtained as $(\lambda_i e^{-\gamma t}) = - \frac{\partial L}{\partial X_i}$ where $i = 0,1,2,3$ and $X_0 = K$, $X_1 = B$, $X_2 = \hat{K}$, $X_3 = A$. Applying these conditions in turn we obtain:

$$\dot{\lambda}_o e^{-\gamma t} - \gamma\lambda_o e^{-\gamma t} = - [+ \phi_o T_p T_c F'(K) - \phi_1 \delta - \phi_3 T_c F'(K) + \phi_3 T_c \delta q]$$

$$\dot{\lambda}_o = \gamma\lambda_o - \phi_o e^{\gamma t}[T_p F'(K)(T_p - \frac{\phi_3}{\phi_o})] + \phi_1 e^{\gamma t}\delta - \phi_3 T_c \delta q$$

$$\dot{\lambda}_o = \gamma\lambda_o + \lambda_1 T_c [F'(K)(T_p - \frac{\phi_3}{\phi_o})] - qT_p\lambda_1\delta - \lambda_2\delta - \phi_3 e^{\gamma t} T_c \delta q \qquad (26)$$

$$\dot{\lambda}_1 e^{-\gamma t} - \gamma\lambda_1 e^{-\gamma t} = - [- \phi_o r_s] = \phi_o r_s$$

$$\dot{\lambda}_1 = \gamma\lambda_1 + \phi_o e^{\gamma t} r_s$$

$$\dot{\lambda}_1 = \lambda_1(\gamma - r_s) \qquad (27)$$

$$\lambda_3 e^{-\gamma t} - \gamma\lambda_3 e^{-\gamma t} = - \{- \phi_o T_p(\theta T_c r_B + (1-\theta) r_E) + \phi_3((1-\theta)s + \theta T_c r_B)\}$$

$$\dot{\lambda}_3 = \gamma\lambda_3 + \phi_o e^{\gamma t} T_p(\theta T_c r_B + (1-\theta) r_E) - \phi_3 e^{\gamma t}((1-\theta)s + \theta T_c r_B)$$

$$\dot{\lambda}_3 = \gamma\lambda_3 - \lambda_1 T_p(\theta T_c r_B + (1-\theta) r_E) - \phi_3 e^{\gamma t}((1-\theta)s + \theta T_c r_B) \qquad (28)$$

$$\dot{\lambda}_2 e^{-\gamma t} - \gamma\lambda_2 e^{-\gamma t} = -[- \phi_o T_p q\alpha T_c + \phi_1 \alpha]$$

$$\dot{\lambda}_2 = \gamma\lambda_2 + \phi_o e^{\gamma t} T_p T_c q\alpha - \alpha\phi_1 e^{\gamma t}$$

$$\dot{\lambda}_2 = \gamma\lambda_2 - \lambda_1 T_p T_c q\alpha - \alpha[- qT_p\lambda_1 - \lambda_2]$$

$$\dot{\lambda}_2 = \lambda_2(\gamma+\alpha) + \alpha q T_p \lambda_1 (1-T_c)$$

$$\dot{\lambda}_2 = \lambda_2(\gamma+\alpha) + T_p U_c \lambda_1 \alpha q \qquad (29)$$

We next substitute the Euler conditions (26) - (29) into (25) in order to obtain an expression for $F'(K)$, the marginal increase in variable profits due to capital accumulation.

$$\gamma\lambda_o + \lambda_1 T_c[F'(K)(T_p - \frac{\phi_3}{\phi_o})] - qT_p\lambda_1\delta - \lambda_2\delta - \phi_3 e^{\gamma t}T_c\delta q$$
$$+ \lambda_2(\gamma+\alpha) + T_p U_c \lambda_1 \alpha q$$
$$+ q\gamma\lambda_3 - q\lambda_1 T_p(\theta T_c r_B + (1-\theta)r_E) - \phi_3 q e^{\gamma t}((1-\theta)s_E + \theta T_c r_B)$$
$$= 0 \tag{30}$$

Solving (30) for $F'(K)$ yields

$$F'(K)[T_p - \frac{\phi_3}{\phi_o}] = qT_p\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \frac{\delta}{T_c}\right] + \frac{\lambda_2(\delta-\alpha)}{\lambda_1 T_c} - \frac{U_c\alpha q T_p}{T_c}$$
$$+ \frac{\phi_3 e^{\gamma t}q}{\lambda_1 T_c}[T_c\delta + ((1-\theta)s_E + \theta T_c r_B)] \tag{31}$$

Now $\frac{\theta_3 e^{\gamma t}}{\lambda_1} = \frac{\theta_3 e^{\gamma t}}{-\phi_o e^{\gamma t}} = -\frac{\phi_3}{\phi_o}$ so that

$$F'(K)[T_p - \frac{\phi_3}{\phi_o}] = qT_p\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \frac{\delta}{T_c}\right] + \frac{\lambda_2(\delta-\alpha)}{\lambda_1 T_c} - \frac{U_c\alpha q T_p}{T_c}$$
$$- \frac{\phi_3}{\phi_o}\frac{q}{T_c}[T_c\delta + ((1-\theta)s_E + \theta T_c r_B)]$$

or

$$F'(K)[1 - \frac{\phi_3}{\phi_o T_p}] = q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \frac{\delta}{T_c}\right] + \frac{\lambda_2(\delta-\alpha)}{\lambda_1 T_c T_p} - \frac{U_c \alpha q}{T_c}$$

$$- \left[\frac{\phi_3}{\phi_o T_p}\right] \cdot \frac{q}{T_c} [T_c\delta + ((1-\theta)s_E + \theta T_c r_B)] \qquad (32)$$

Equation (32) contains the expressions for the user cost of capital services and the allowed gross return inherent in our intertemporal model. This fact, however, is not obvious. To demonstrate the validity of this statement it is useful to digress briefly and consider the simpler, atemporal Averch-Johnson model. The owner of the firm wishes to choose K so as to maximize after tax profits

$$\pi = T_p[F(K) - cK] \qquad (33)$$

where c is the user cost of capital and we have assumed, for simplicity, no corporate income tax and a zero rate of depreciation. The rate of return constraint is

$$F(K) \leq sK \qquad (34)$$

where s is the allowed rate of return

To maximize (33) subject to (34) we form the Lagrangian expression

$$L = T_p[F(K) - cK] + \mu[sK - F(K)] \qquad (35)$$

The first order condition for a maximum is

$$T_p[F'(K) - c] + \mu s - \mu F'(K) = 0$$

or

$$F'(K)[1 - \frac{\mu}{T_p}] = c - (\frac{\mu}{T_p})\cdot s \tag{36}$$

Comparing (36) and (32) we can see that the first order conditions coincide, where $\mu = \frac{\phi_3}{\phi_o}$, the user cost of capital services is

$$c = q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \frac{\delta}{T_c}\right] + \frac{\lambda_2(\delta-\alpha)}{\lambda_1 T_c T_p} - \frac{U_c \alpha q}{T_c} \quad , \tag{37}$$

and the allowed gross return on capital services is

$$s = \frac{q}{T_c}[T_c\delta + ((1-\theta)s_E + \theta T_c r_B] \tag{38}$$

The next step in the derivation is to eliminate the unknown co-state variables from equation (37). We begin by integrating equation (27) to yield

$$\lambda_1 = \lambda_1(0)e^{(\gamma-r_s)t} \tag{39}$$

where $\lambda_1(0)$ is the constant of integration.

Substituting (39) into (29) and integrating we obtain

$$\lambda_2 = \lambda_2(0)e^{(\gamma+\alpha)t} + \lambda_1(0)T_p U_c \alpha q\left[\frac{1}{(r_s+\alpha)}(e^{(\gamma+\alpha)t} - e^{(\gamma-r_s)t})\right] \tag{40}$$

Substituting (39) and (40) into (37) yields the user cost of capital equation

$$c = q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \frac{\delta}{T_c}\right] - \frac{U_c\alpha q}{T_c}$$

$$+ \frac{(\delta-\alpha)}{T_p}\left[\frac{\lambda_2(0)e^{(\alpha+r)t}}{\lambda_1(0)T_c} + \frac{U_c\alpha qT_p}{(r_s+\alpha)T_c}\ e^{(\alpha+r_s)t} - \frac{U_c\alpha qT_p}{(r_s+\alpha)T_c}\right]$$

$$= q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \frac{\delta}{T_c}\right] - \frac{U_c\alpha q}{T_c} - (\delta-\alpha)\frac{U_c\alpha q}{(r_s+\alpha)T_c}$$

$$+ \left\{\frac{(\delta-\alpha)}{T_pT_c}\left[\frac{\lambda_2(0)}{\lambda_1(0)} + \frac{U_c\alpha qT_p}{(r_s+\alpha)}\right] e^{(\alpha+r_s)t}\right\} \qquad (41)$$

Following Boadway and Bruce (1979) we note that a stationary solution to the optimal control problem requires the term in { } brackets of (41) to be zero. Since non-stationary solutions can be shown to be non-optimal the constants of integration $\lambda_1(0)$, $\lambda_2(0)$, must satisfy this condition. Given the implied initial conditions, the user cost of capital services becomes

$$c = q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c}\right] + q\delta + q\delta\left(\frac{1}{T_c} - 1\right) - \frac{U_c\alpha q}{T_c} - \frac{(\delta-\alpha)U_c\alpha q}{(r_s+\alpha)T_c}$$

$$= q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \delta\right] + \frac{U_cq\delta}{T_c} - \frac{U_c\alpha q}{T_c} - (\delta-\alpha)\frac{U_c\alpha q}{(r_s+\alpha)T_c}$$

$$= q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \delta\right] + (\delta-\alpha)\frac{U_cq}{T_c}\left(1 - \frac{\alpha}{(r_s+\alpha)}\right)$$

$$= q\left[\theta r_B + \frac{(1-\theta)r_E}{T_c} + \delta\right] + \frac{r_c U_c q}{T_c(r_s+\alpha)}(\delta-\alpha)$$

The final expression for the user cost of capital services is

$$c = q\left[\theta r_B + (1-\theta)\frac{r_E}{T_c} + \delta\right] + \frac{q r_s U_c}{T_c(r_s+\alpha)} \cdot (\delta-\alpha) \tag{42}$$

and for the allowed gross return on capital services is

$$s = q\left[\theta r_B + (1-\theta)\frac{s_E}{T_c} + \delta\right] \tag{43}$$

There are several aspects of (42) and (43) worth noting. First, the cost of equity r_E and the allowed return on equity s_E are both after-tax percentages. They are "grossed up" through the division by $T_c = 1 - U_c$. The asymmetric treatment of r_B and r_E reflects the fact that the cost of bond financing is tax deductible whereas the cost of equity financing is not. Second, if accelerated depreciation is allowed for tax purposes then δ, the rate of economic depreciation, is less than α , the rate of depreciation for tax purposes, and the user cost of capital is correspondingly reduced. Third, if only economic depreciation is allowed for tax purposes, c and s exhibit a particularly simple relationship. They differ only by the difference between r_E and s_E . Fourth, the personal rate of taxation (U_p) does not appear in (42) and (43) and therefore does not have to be known. However the personal borrowing rate (r_s) does appear in (42) and is relevant as long as accelerated depreciation for tax purposes is allowed. Finally, suppose rate of return regulation is ineffective. This can occur if either the allowed rate of return is set sufficiently high so that constraint (13) is non-binding (and $\phi_3 = 0$), or if constraint (13) is ignored by the firm under the assumption that rate of return regulation will not be enforced. In both cases equation (42) still yields the appropriate cost of capital services. However now the allowed gross return on capital services (equation (43)) is no longer relevant to the utility's decision making processes.

B. The Cost of Capital Services and the Allowed Gross Return on Capital Services for Bell Canada

3.0 Introduction

In this section we implement empirically the theoretical model developed in section A in order to obtain estimates of the cost of capital services and the allowed gross return on capital services for Bell Canada for the period 1952-78. As part of the process of estimation we also obtain estimates of the cost of equity capital and the allowed rate of return on equity capital.

The two most difficult variables to measure are the cost of equity capital (r_E) and the real capital stock for tax purposes $(\hat{K})$ used in the computation of the accelerated depreciation rate (α). Hence most of the detailed explanation will be related to these two variables.

We begin with an explanation of the computation of the financial variables: the cost of debt capital (r_B), the cost of equity capital (r_E) and the allowed rate of return on equity capital (s_E).

3.1 The Cost of Debt Capital

Interest payments in year t were calculated as the total interest charges minus interest charges not related to capital (taken from [3])[5]. Total debt capital was calculated as the average of end of year t and end of year t-1 total debt capital (from [1]). The cost of debt capital is the ratio of interest charges to total debt capital. The calculated series for 1951-78 appears in column 1 of Table 3.1.[6]

3.2 The Cost of Equity Capital

(a) The Cost of Common Equity Capital

There are two main methods currently used by finance economists to measure the cost of common equity capital. The first is the intrinsic yield formula, in which the cost of capital is the discount rate (or yield) at which the stream of expected future dividends must be discounted in order that the present value of this stream equal the current share price. The second method for measuring the cost of equity capital is based on the capital asset pricing model derived from modern portfolio theory. In this method the cost of capital is determined by estimating the risk premium required by investors in order that the shares be held in a market portfolio, and adding this risk premium to the interest rate on a risk-free bond.

We have chosen to utilize the intrinsic formula method since adequate data were not available to estimate the parameters of the capital asset pricing model. In utilizing the intrinsic formula we have followed the procedure employed by Gordon and Pradham (1975).[7]

The intrinsic yield formula is

$$P_t = \sum_{t=1}^{\infty} \frac{D_t}{(1+k)^t} \qquad (1)$$

Table 3.1

Costs of Debt and Equity Capital (net of income tax)

T	Debt	Common Equity	Non-Convertible Preferred Equity	Expected Rate of Inflation
1951	0.034881	0.062045		-0.0039
1952	0.034368	0.062689		0.0024
1953	0.038517	0.065971		0.0045
1954	0.036394	0.067790		-0.0033
1955	0.036365	0.061069		-0.0021
1956	0.039144	0.060950		0.0076
1957	0.040502	0.065147		0.0156
1958	0.038453	0.068798		0.0047
1959	0.040882	0.064057		0.0194
1960	0.044522	0.068355		0.0152
1961	0.046382	0.062880		0.0137
1962	0.048039	0.057039		0.0160
1963	0.047001	0.062521		0.0148
1964	0.047252	0.062534		0.0172
1965	0.047665	0.060186		0.0190
1966	0.049620	0.065112		0.0255
1967	0.051688	0.081244		0.0264
1968	0.053337	0.081962		0.0368
1969	0.056020	0.079036		0.0466
1970	0.055673	0.080347		0.0410
1971	0.058850	0.077369		0.0255
1972	0.062364	0.081881		0.0326
1973	0.066057	0.086685	0.036889	0.0398
1974	0.067616	0.099817	0.074939	0.0512
1975	0.074427	0.099615	0.074908	0.0468
1976	0.075476	0.121205	0.059571	0.0531
1977	0.079496	0.112998	0.084003	0.0479
1978	0.079997	0.101566	0.084080	0.0600

where P_t = the current share price

D_t = the expected dividend in year t

k = the cost of common equity capital

If the current dividend is D and its expected growth rate is g , the cost of equity capital can be obtained as

$$k = \frac{D}{P} + g \tag{2}$$

Since we know D and P, the problem of estimating k reduces to the problem of estimating g , the expected growth rate of dividends. Following Gordon and Pradham, we measure g as the intrinsic growth rate defined as

$$g = br + v\sigma \tag{3}$$

where b = the fraction of its income the utility is expected to retain

r = the expected rate of return earned on common equity

σ = funds expected to be raised from the sale of stock, as a fraction of the existing common equity.

v = the fraction of funds raised which is expected to accrue to the existing shareholders

We can estimate b, r, and σ by using the actual historical data. The equity accretion rate can be shown to be equal to

$$v = 1 - E/P\lambda \tag{4}$$

where E = book value per share, P is the price per share, and λ is the proportion of the normal price accruing to the corporation from the stock issue, assumed to be 0.95.

Using the above results, the cost of common equity capital k_t is calculated as

$$k_t = \frac{D_t}{P_t} + b_t r_t + \sigma_t \left[1 - E_t/\lambda P_t\right] \qquad (5)$$

The sources of the variables used in (5) are as follows:

D_t = expected annual dividend rate per common share, based on the quarterly payment made in last quarter of year t. The source is [2].

P_t = share price at end of year t, taken from [2].

$b_t = \frac{NY_t - ND_t}{NY_t}$, where NY_t = net income applicable to common shares for year t, taken from [1].

and ND_t = total paid in dividends on common shares

$r_t = \frac{NY_t}{NE_t^{AVG}}$, where NE_t^{AVG} = average end of year (years t and $t-1$) total common equity capital, from [1].

(Note that we have assumed that actual and expected rates of return are equal.)

$\sigma_t = \alpha_t \frac{P_t^v}{E_{t-1}}$, where α_t = the stock issue rate,

P_t^v = the issue price of new shares,

E_{t-1} = common equity capital per share at end of year $t-1$, all from [1].

The series b_t , r_t , σ_t and v_t were smoothed in the manner suggested by Gordon and Pradham. Table 3.1, column 2 contains the cost of common equity capital k_t calculated for the years 1951-78.

(b) The Cost of Preferred Equity Capital

There are two kinds of preferred equity - convertible and non-convertible. We assumed that the cost of convertible preferred equity is the same as that of common equity. We calculated the return on non-convertible equity in the same manner as that used for debt equity. The return on non-convertible equity appears in column 3 of Table 3.1.

3.3 The Calculation of Real Rate of Return

Unless static expectations are assumed in the theoretical cost of capital services model, real rates of return should be used in calculating the actual gross cost of capital services and the allowed gross cost of capital services. The determination of real rates require a measure of the expected rate of inflation. The expected rate of inflation was estimated in the following way. It is generally believed that over long periods of time the real rate of return in the Canadian economy has averaged approximately 3% per year. We assume that a Government of Canada bond is a riskless asset whose yield (r_C) consists of the underlying real rate of return plus the expected rate of inflation (μ) . Hence we measure the expected rate of inflation as $\mu = r_C - .03$. The calculated values of μ appear in column 4 of Table 3.1. The bond rate used is the 3-5 year rate, chosen in order to estimate medium term inflationary expectations. Estimated real costs of debt and equity capital for Bell Canada are obtained by subtracting μ from the nominal rates.

3.4 A Comparison of the Cost of Equity Capital, the Actual Rate of Return on Equity, and the Allowed Rate of Return on Equity

The actual rate of return on equity was defined previously as

$$r_t = \frac{NY_t}{NE_t^{AVG}} \tag{6}$$

The allowed rate of return on common equity in year t is defined as

$$ALLR(t) = \frac{ALLEPS(t)}{EAVG(t)} \tag{7}$$

where ALLEPS(t) = allowed earnings per share for year t, taken from [4] and [5]. In those cases where an allowable range was indicated, the upper limit was used.

EAVG(t) = the average common equity capital per share during year t, taken from [1].

Table 3.2 presents estimates of nominal k_t, r_t, and ALLR(t). It is interesting to note that the allowed rate of return on common equity exceeds the cost of capital in all years except 1967, 1968 and 1976. By way of contrast, the cost of capital exceeds the actual rate of return in 1950, 1957, 1958, 1969, 1974, 1976 and 1977. Finally,

Table 3.2

A Comparison of the Cost of Common Equity,

The Allowed Return on Common Equity,

and the Actual Return on Common Equity (Post-tax)

T	Cost of Common Equity	Allowed Return on Common Equity	Actual Return on Common Equity
1950	0.066930	0.077404	0.046654
1951	0.062045	0.077389	0.066082
1952	0.062689	0.076650	0.081512
1953	0.065971	0.075487	0.078559
1954	0.067790	0.074287	0.074666
1955	0.061069	0.072737	0.075285
1956	0.060950	0.071031	0.070045
1957	0.065147	0.070262	0.062223
1958	0.068798	0.070105	0.062056
1959	0.064057	0.069776	0.073634
1960	0.068355	0.069296	0.072039
1961	0.062880	0.068528	0.072140
1962	0.057039	0.067306	0.072338
1963	0.062521	0.066308	0.070477
1964	0.062534	0.065466	0.076393
1965	0.060186	0.064251	0.079338
1966	0.065112	0.076852	0.076773
1967	0.081244	0.076016	0.082463
1968	0.081962	0.075818	0.084260
1969	0.079036	0.087974	0.078177
1970	0.080347	0.087888	0.084467
1971	0.077369	0.087894	0.088378
1972	0.081881	0.104654	0.094017
1973	0.086665	0.104686	0.099618
1974	0.099817	0.120511	0.096403
1975	0.099615	0.118096	0.147714
1976	0.121205	0.120479	0.100695
1977	0.112998	0.120154	0.090734
1978	0.101566	0.121290	0.110027

the actual rate of return exceeds the allowed rate of return in years 1952, 1953, 1954, 1955, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1967, 1968, 1971, and 1975. The exceeding of the allowed rate of return was commonplace during the 1950's and 1960's, but has occurred only twice during the 1970's. This fact suggests that rate of return regulation may have become effective only during the 1970's.

3.5 The Capital Stock for Tax Purposes and the Accelerated Rate of Depreciation

In order to calculate the rate of accelerated depreciation α_t, we require an estimated series of the capital stock for tax purposes, $\hat{K}_t$, and a series for the allowed depreciation for tax purposes A_t. The required series were obtained in the following way. The deferred income tax series was grossed up by the applicable income tax rate to provide an estimate of depreciation in excess of economic depreciation. Both these series were taken from [1]. A very detailed analysis of the deferred tax credit series was undertaken. Using information from regulatory hearings and Bell Canada annual reports retroactive adjustments to deferred credits were determined and included in the series. Adding economic depreciation to this excess provides an estimate of accelerated depreciation. The economic depreciation series was taken from [3]. The capital stock for tax purposes $\hat{K}_t$ was generated using the perpetual inventory difference equation,

Table 3.3

Economic and Accelerated Depreciation Rates

T	Economic Depreciation Rate (δ)	Accelerated Depreciation Rate (α)
1952	0.058824	0.058824
1953	0.058747	0.058745
1954	0.058197	0.093539
1955	0.055913	0.097441
1956	0.055667	0.096539
1957	0.058879	0.098359
1958	0.058824	0.065561
1959	0.060625	0.066808
1960	0.060101	0.065788
1961	0.060046	0.065207
1962	0.060621	0.065312
1963	0.062286	0.067538
1964	0.062958	0.067959
1965	0.064257	0.068968
1966	0.065493	0.070034
1967	0.066066	0.099516
1968	0.067290	0.099808
1969	0.069485	0.100191
1970	0.069682	0.108322
1971	0.070158	0.132605
1972	0.074389	0.138521
1973	0.077697	0.146611
1974	0.080010	0.145286
1975	0.084613	0.163208
1976	0.087318	0.159345
1977	0.087525	0.170261
1978	0.089158	0.156124

$$\hat{K}_t = \hat{K}_{t-1} + I_t - A_t \tag{8}$$

where I_t = real investment during year t

A_t = real allowed depreciation for tax purposes, year t

The initial value for $\hat{K}_t$ is the 1952 actual capital stock, K_{1951}. This is a correct procedure since accelerated depreciation for tax purposes was not introduced until 1954. The rate α_t is obtained as

$$\alpha_t = \frac{A_t}{\hat{K}_t} \tag{9}$$

Table 3.3 presents a comparison of the depreciation rate for tax purposes α_t and the rate of economic depreciation δ_t. As expected $\alpha_t \geq \delta_t$. Note that the gap between α_t and δ_t has widened considerably during the 1970's, reflecting the government liberalization of allowed depreciation and the accumulated effects of this liberalization.

3.6 The User Cost of Capital (c_t) and the Allowed Gross Return to Capital (s_t)

We are now in a position to calculate c_t and s_t using equations (42) and (43) respectively, from the theoretical section of this chapter. The only variables which appear in these equation that have not been discussed are the personal borrowing rate r_s and the asset price q. There exists no information which would allow us to compute a representative borrowing rate faced by owners of Bell Canada common stock.

For want of a better alternative, we assume $r_s = r_B$. To the extent that stockholders can use stock as collateral for loans, this may not be an unreasonable assumption. For q we have used the telephone plant price index obtained from [3].

The user cost of capital services and the allowed gross return on capital services are presented in Table 3.4. Note that in all cases the allowed gross return is greater than the gross cost of capital services. This includes the three cases noted previously where the net rate of return (r) was less than the cost of capital (k). In those cases the benefits accruing to accelerated depreciation have reversed the inequality. The fact that $s_t > c_t$ opens up the possibility that an Averch-Johnson capital accumulation bias is present.

Table 3.4

User Cost of Capital Services and the Allowed Gross Return on Capital Services

T	User Cost of Capital Services	Allowed Gross Return on Capital Services
1952	0.120438	0.134806
1953	0.114053	0.122791
1954	0.115551	0.127859
1955	0.107200	0.124868
1956	0.097666	0.114483
1957	0.097376	0.109826
1958	0.118085	0.120859
1959	0.102700	0.110261
1960	0.111709	0.114393
1961	0.111522	0.119001
1962	0.104208	0.116318
1963	0.112289	0.117793
1964	0.110686	0.115274
1965	0.108169	0.113768
1966	0.110361	0.123941
1967	0.123933	0.127940
1968	0.116140	0.119588
1969	0.105996	0.126893
1970	0.123060	0.149247
1971	0.148356	0.177441
1972	0.152384	0.193994
1973	0.157424	0.199091
1974	0.171400	0.214610
1975	0.206320	0.245758
1976	0.245085	0.270057
1977	0.261302	0.305436
1978	0.236086	0.303185

Footnotes to Chapter Three

1. For an analysis of the variable profit function, see Diewert (1974) or McFadden (1978), where it is called the restricted profit function.
2. The cost of equity capital is calculated as $(1-\theta)r_E \cdot A$. This is only an approximation for two reasons. First, only part of the equity cost is in the form of dividends, the remainder being in the form of expected capital gains. However, for public utilities, dividends form the largest part of equity costs. Second, the expression $r_E \cdot A$ should really be $r_E\bar{A}$ where $\bar{A}$ = A - value of shares owned by the shareholder who receives V . Since shares of public utilities tend to be widely held, a reasonable approximation is that $A = \bar{A}$.
3. On the importance of a constraint on liabilities accumulation similar to (11), see Boadway and Bruce (1979).
4. Boadway and Bruce (1979) have shown that constraint (11) must hold with equality for an optimal solution. In addition, we assume that rate of return regulation is effective so that constraint (13) holds with equality.
5. The data sources appear at the end of this section.
6. This cost of debt capital is an average cost. From a theoretical point of view, it is the marginal cost which should determine investment expenditures. However the correct marginal cost is the <u>expected</u> marginal cost, which is unknown at the time the investment is planned. This distinction is important in telecommunications, where major investment decisions require considerable lead-time. We have used the average cost of debt capital as a proxy for the expected marginal cost, which implies a relatively simple autoregressive expectations formation mechanism.
7. For additional details concerning the methodology, see also Gordon (1974).

Data Sources

[1] Bell Canada Annual Charts 1935-77.

[2] Toronto Stock Exchange Monthly Review/TSE Review, December issues.

[3] Confidential data set supplied by Bell Canada. This data set revises and updates to 1978 the data set contained in the Memorandum on Productivity and Bell Canada's Productivity.

[4] Bell Canada Exhibit to the CRTC No. B-78-649 (volume 22, May 7, 1978).

[5] Bell Canada Annual Reports.

CHAPTER FOUR

The Cost Structure of Bell Canada - Tests of the Natural Monopoly Hypothesis

4.1 Introduction

In this chapter we present an econometric model of the cost structure for Bell Canada. The model is developed under the assumption that rate-of-return regulation does not constrain the production activities of the firm. (The effect of rate-of-return regulation is analysed in Chapter 5.) The activities of the firm are, however, assumed to be constrained by tariff regulation. In particular, we assume that the regulatory commission prevents Bell Canada from charging the profit maximizing price for basic local service.

The main purpose of estimating the cost structure is to provide a means of testing the hypothesis that Bell Canada is a natural monopoly in some or all of its service offerings. If Bell Canada is a natural monopoly over some range of outputs its cost function will be subadditive over that range. While subadditivity is very difficult to test per se, as noted in Chapter 1, there exist sufficient conditions for subadditivity - economies of scale and economies of scope, and these conditions are more amenable to the formulation of testable hypotheses. We next outline in general terms the relevant tests. In subsequent sections we will develop specific tests in terms of the particular econometric model estimated.

4.2 Tests of Overall Economies of Scale and Overall Economies of Scope

The starting point for testing the natural monopoly hypothesis is the construction of a test for overall economies of scale. Overall economies of scale exist if an increase in all outputs of $\lambda\%$ leads to a cost increase of less than $\lambda\%$. As shown by Panzar and Willig (1979) and Fuss and Waverman (1977), local overall economies of scale are measured by the scale elasticity

$$S = \frac{1}{\sum_{j=1}^{N} \varepsilon_{CQ_j}} \quad (1)$$

where ε_{CQ_j} is the cost-output elasticity of the j-th output. If $S > 1$, economies of scale prevail locally; if $S < 1$ diseconomies of scale prevail and if $S = 1$, constant returns to scale prevail.

Economies of scope can be tested in the following way. Suppose an N output production process can be represented by the joint cost function

$$C = C(Q_1, Q_2, \ldots Q_N) \quad (2)$$

where factor prices and any other arguments of the cost function have been suppressed for simplicity. Overall economies of scope can be determined by comparing the cost of producing each output separately (the "stand alone" cost) with the actual joint cost. The relevant expression is

$$SC = \sum_{j=1}^{N} C_j(Q_j) - C(Q_1, Q_2, \ldots Q_N) \quad (3)$$

If $SC > 0$, economies of scope exist; if $SC < 0$, diseconomies of scope exist and independent production is cost-minimizing. If $SC = 0$, joint production neither yields cost savings nor causes cost increases.

It should be noted that to compute overall economies of scope requires that one be able to compute stand-alone costs. In telecommunications this would require observations on independent production of outputs such as message toll, competitive and local services. We do not have the required set of observations and will not attempt to estimate overall economies of scope.

4.3 Product-Specific Economies of Scope and Economies of Scale

One particular public policy issue of considerable importance is the question of whether competition in the provision of certain services should be encouraged. We can shed light on this issue by attempting to estimate the extent of product specific economies of scope and economies of scale in the provision of private line services. One requirement for computing product-specific economies of scope is that we observe a production process in which a zero amount of the product under consideration is produced. For private line services this requirement is approximately met, since Bell Canada produced a very small output of this service in the early 1950's, which is part of our sample. Unfortunately, a second requirement for computing private line-specific economies of scope is that we observe independent production of this output, so that stand-alone costs can be estimated. The cost function which we will present in the next section allows for the estimation of stand-alone costs. However this estimation requires extrapolation of the cost function well outside the observed data points for toll and local services, and thus considerable caution must be exercised in interpreting the results.

We now present the test for product-specific economies of scope. Suppose private line service is the j-th service output. Product-specific economies of scope with respect to private line service exist if

$$C(Q_1,Q_2,\ldots Q_{j-1},0,Q_{j+1},\ldots Q_N) + C(0,\ldots 0,Q_j,0,\ldots 0) - C(Q_1,Q_2,\ldots Q_N) > 0 \qquad (4)$$

Panzar and Willig (1979) have defined the degree of product specific economies of scope as

$$SC_j = \frac{C(Q_1,Q_2,\ldots Q_{j-1},0,Q_{j+1},\ldots Q_N) + C(0,\ldots 0,Q_j,0,\ldots 0) - C(Q_1,\ldots Q_N)}{C(Q_1,\ldots Q_N)} \qquad (5)$$

If $SC_j > 0$, SC_j measures the proportionate increase in cost from separating private line services from the production of other services. If $SC_j < 0$, it measures the proportionate cost decrease from independent production of private line services.

Panzar and Willig (1979) have also proposed a measure of product specific economies of scale. They define the degree of product j specific economies of scale as

$$S_j = \frac{IC_j}{Q_j \frac{\partial C}{\partial Q_j}} \qquad (6)$$

where $IC_j = C(Q_1,Q_2,\ldots Q_N) - C(Q_1,\ldots Q_{j-1},0,Q_{j+1},\ldots Q_N)$ is the incremental cost of producing product j . It can be shown that (6) can be written in the form

$$S_j = \frac{IC_j}{C} \Big/ \varepsilon_{CQ_j} \qquad (7)$$

If $S_j > 1$, there exists product j specific economies of scale (locally). If $S_j < 1$, there exists diseconomies of scale and if $S_j = 1$, there exists constant returns to scale.

4.4 The Econometric Model

We assume that the production process is one in which three outputs are produced using three inputs. The three outputs chosen are (i) monopoly toll (message toll and WATS) (Q_1), (ii) competitive toll-private line services (plus TWX), (Q_2) and (iii) local service plus miscellaneous (Q_3). Inputs into the production process are labour (L), capital services (K) and materials (M). We utilize two technical change indicators[1]: A, the proportion of telephones with access to direct distance dialing facilities, and S the percentage of telephones connected to central offices with modern switching facilities[2], in order to model the shift in the cost function due to technical change. Hence the cost function can be written as

$$C = g(P_L, P_M, P_K; Q_1, Q_2, Q_3; A, S) \tag{8}$$

where P_L, P_M, P_K are the input prices of labour, capital services and materials respectively.

The behavioural specification employed is one in which the regulated telecommunications firm chooses the profit maximizing levels of toll services (Q_1 and Q_2), but is constrained by the regulatory authorities to charge a price for local services (Q_3) below the profit-maximizing price[3]. We assume that the cost function (8) can be written in the "output-augmenting" form

$$C = C[P_L, P_K, P_M, Q_1 \cdot h_1(A), Q_2 \cdot h_2(A), Q_3 \cdot h_3(S)] \tag{9}$$

where A and S are the technical change indicators defined previously. The h_i functions are augmentation functions such that for any given Q_1, Q_2 and Q_3, an increase in A and/or S will lead to a decline in costs, but an increase in A will have as its major impact a decline in the marginal cost of toll services and an increase in S will have its major impact on the marginal cost of local service. Define the "augmented" outputs by

$$Q_1^* = Q_1 \cdot h_1(A) = Q_1 e^{\lambda_1 A} \tag{10}$$

$$Q_2^* = Q_2 \cdot h_2(A) = Q_2 e^{\lambda_2 A} \tag{11}$$

$$Q_3^* = Q_3 \cdot h_3(S) = Q_3 e^{\lambda_3 S} \tag{12}$$

Then the cost function (9) becomes

$$C = C[P_L, P_K, P_M, Q_1^*, Q_2^*, Q_3^*] \tag{13}$$

In previous analyses of Canadian telecommunications, the production structure of Bell Canada was estimated using Cobb-Douglas (Dobell et al (1972)) or Translog (Fuss and Waverman (1977), Denny et al (1979), Breslaw and Smith (1980)) functional forms. For our purposes, the major defect of these functional forms is that they are undefined whenever one of the outputs is zero. As we have seen above, necessary and sufficient tests of economies of scope and tests of product-specific economies of scale require a cost function which is defined at zero levels of output[4].

In order to resolve this problem we introduce the "hybrid" Translog cost function. The hybrid translog approximation to (13) takes the form

$$\log C = \alpha_0 + \sum_i \alpha_i \log P_i + \sum_k \beta_k \left[\frac{Q_k^{*\theta} - 1}{\theta}\right]$$

$$+ \tfrac{1}{2} \sum_i \gamma_{ii} (\log P_i)^2 + \sum_i \sum_{\substack{j \\ i \neq j}} \gamma_{ij} \log P_i \log P_j$$

$$+ \frac{1}{2} \sum_{k} \delta_{kk} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right]^2 + \sum_{k} \sum_{\substack{\ell \\ k \neq \ell}} \delta_{k\ell} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] \left[\frac{Q_\ell^{*\theta} - 1}{\theta}\right]$$

$$+ \sum_{i} \sum_{k} \rho_{ik} \log P_i \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] \qquad (14)$$

where $i,j = L,K,M$

$k,\ell = 1,2,3$

The hybrid translog function contains as special cases the ordinary translog function and the Cobb-Douglas function. When $\theta = 0$, (14) reduces to the translog function since

$$\lim_{\theta \to 0} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] = \log Q_k^* \qquad (15)$$

If in addition, $\gamma_{ii} = \gamma_{ij} = \delta_{kk} = \delta_{k\ell} = \rho_{ik} = 0$, (14) reduces to the Cobb-Douglas cost function. Note that (14) is not degenerate when $Q_k^* = 0$ as would be the case with the translog and Cobb-Douglas functions. The input cost share equations can be obtained from Shephard's Lemma (e.g. $\frac{\partial C}{\partial P_L} = L$) as

$$S_i = \alpha_i + \sum_{i} \gamma_{ij} \log P_j + \sum_{k} \rho_{ik} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] \qquad (16)$$

$i = L,K,M \quad ; \quad k = 1,2,3$

The fact that $\sum S_i = 1$ implies the constraints

$$\sum \alpha_i = 1 , \quad \sum_{i} \gamma_{ij} = 0 , \quad \sum_{i} \rho_{ik} = 0 \qquad (17)$$

The second order approximation property of the cost function implies the additional constraints

$$\gamma_{ij} = \gamma_{ji} \quad i \neq j ; \qquad \delta_{k\ell} = \delta_{\ell k} \quad \ell \neq k \tag{18}$$

Profit-maximizing behaviour with respect to toll services involves the setting of marginal costs equal to marginal revenues, and yields two additional equations which we label the revenue 'share' equations[5].

$$\frac{P_1 Q_1}{C} = \left[\frac{1}{1 + 1/\varepsilon_1}\right]^{-1} \cdot \left[Q_1^{*\theta}\right] \cdot \left\{\beta_1 + \sum_{\ell} \delta_{1\ell} \left[\frac{Q_\ell^{*\theta} - 1}{\theta}\right] + \sum_{i} \rho_{i1} \log P_i\right\} \tag{19}$$

$$\frac{P_2 Q_2}{C} = \left[\frac{1}{1 + 1/\varepsilon_2}\right]^{-1} \cdot \left[Q_2^{*\theta}\right] \cdot \left\{\beta_2 + \sum_{\ell} \delta_{2\ell} \left[\frac{Q_\ell^{*\theta} - 1}{\theta}\right] + \sum_{i} \rho_{i2} \log P_i\right\} \tag{20}$$

$i = L,K,M$

$\ell = 1,2,3$

where ε_1 and ε_2 are the own price elasticities of demand for monopoly toll and competitive toll services respectively.[6]

The formulas for factor price elasticities of demand are the same as those obtained for the ordinary translog cost function. These can be shown to be (Berndt and Wood (1975))

$$\varepsilon_{ii} = \frac{\gamma_{ii} + S_i^2 - S_i}{S_i} \qquad i = L,K,M$$

$$\varepsilon_{ij} = \frac{\gamma_{ij} + S_i S_j}{S_i} \quad (i \neq j) \qquad i,j = L,K,M \tag{21}$$

where ε_{ij} is the elasticity of demand for factor i with respect to the price of factor j. For the hybrid translog model the cost-output elasticities $\varepsilon_{CQ_\ell} = \frac{\partial \log C}{\partial \log Q_\ell}$ can be calculated as

$$\varepsilon_{CQ_\ell} = \left[Q_\ell^{*\theta}\right] \cdot \left\{\beta_\ell + \sum_k \delta_{\ell k} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] + \sum_i \rho_{i\ell} \log P_i\right\} \qquad (22)$$

$k,\ell = 1,2,3$

$i = L,K,M$

4.5 Hypothesis Tests for the Hybrid Translog Cost Function

In this section we demonstrate the way in which the existence of economies of scale and scope can be tested at a point (i.e. locally) using the hybrid translog model. We begin by noting that, like the ordinary translog function, the hybrid translog function's factor price and cost elasticities are invariant to any multiplicative scaling of P_i and Q_k^*. The same phenomenon is true for likelihood ratio, F and t tests of hypotheses. Hence we can scale the data so that $P_i = 1$ and $Q_k^* = 1$ at the point at which an hypothesis is being tested. This transformation greatly simplifies the formulas on which the hypotheses are based. For example, consider the test of overall economies of scale. At $P_i = Q_k^* = 1$, the cost elasticity ε_{CQ_ℓ} reduces to the parameter β_ℓ and the total cost elasticity becomes $\sum_\ell \beta_\ell$. The overall scale elasticity is reduced to

$$S = \frac{1}{\sum_\ell \beta_\ell} \qquad (23)$$

A local test of overall constant returns to scale reduces to a test of the null hypothesis $\sum_{\ell} \beta_{\ell} = 0$.

To obtain a local test of product-specific economies of scope with respect to competitive toll services, we note that when $P_i = Q_k^* = 1$,

$$C(Q_1,Q_2,Q_3) = C(1,1,1) = \exp[\alpha_0] \tag{24}$$

$$C(Q_1,0,Q_3) = C(1,0,1) = \exp\left[\alpha_0 - \frac{\beta_2}{\theta} + \frac{\delta_{22}}{2\theta^2}\right] \tag{25}$$

$$C(0,Q_2,0) = C(0,1,0) = \alpha_0 - \frac{1}{\theta}(\beta_1 + \beta_3) + \frac{1}{2\theta^2}(\delta_{11} + \delta_{33} + 2\delta_{13}) \tag{26}$$

Hence from (5)

$$SC_2 = \frac{\exp\left[\alpha_0 - \frac{\beta_2}{\theta} + \frac{\delta_{22}}{2\theta}\right] + \exp\left[\alpha_0 - \frac{1}{\theta}(\beta_1 + \beta_3) + \frac{1}{2\theta^2}(\delta_{11} + \delta_{33} + 2\delta_{13})\right] - \exp[\alpha_0]}{\exp[\alpha_0]} \tag{27}$$

We can now test the null hypothesis of an absence of product-specific economies of scope $(SC_2 = 0)$ by forming a test statistic based on the right-hand side of (27).

In a similar way we can test the null hypothesis of an absence of product-specific economies of scale with respect to competitive toll services (Q_2). At the point $P_i = Q_k^* = 1$, the incremental cost of producing Q_2 is

$$IC_2 = C(1,1,1) - C(1,0,1) = \exp[\alpha_0] - \exp\left[\alpha_0 - \frac{\beta_2}{\theta} + \frac{\delta_{22}}{2\theta^2}\right] \tag{28}$$

Hence the degree of product 2 specific returns to scale (equation (7)) is given by

$$S_2 = \frac{IC_2}{C} \Big/ \varepsilon_{CQ_2}$$

$$= \frac{\exp[\alpha_0] - \exp\left[\alpha_0 - \frac{\beta_2}{\theta} + \frac{\delta_{22}}{2\theta^2}\right]}{\alpha_2 \cdot \exp[\alpha_0]} \qquad (29)$$

To test the null hypothesis that competitive toll services are produced under constant returns to scale $(S_2 = 0)$ by the multiproduct firm, we form a test statistic based on the right-hand side of (29).

One final test of product-specific returns to scale will prove of interest. Suppose private line services are produced by two firms in the amounts Q_2^1 and Q_2^2, so that industry output is $Q_2 = Q_2^1 + Q_2^2$. We may be interested in whether the takeover of firm 2's output by firm 1 would allow firm 1 to produce the additional output under increasing returns to scale (declining average incremental cost). It can be shown that the degree of returns to scale associated with this takeover can be computed as

$$\tilde{S}_2 = \left[\frac{IC_2}{C} \Big/ \varepsilon_{CQ_2}\right] \cdot \left[1 - \frac{Q_2^1}{Q_2}\right] \qquad (30)$$

where IC_2, C, and ε_{CQ_2} are all evaluated at Q_2. If $\tilde{S}_2 > 1$, then the additional production is subject to increasing returns to scale. The appropriate test statistic at the point $P_i = Q_k^* = 1$ and $Q_2^1 + Q_2^2 = 1$ is

$$\tilde{S}_2 = \frac{\exp[\alpha_0] - \exp\left[\alpha_0 - \frac{\beta_2}{\theta} + \frac{\delta_{22}}{2\theta^2}\right]}{\alpha_2 \cdot \exp[\alpha_0]} \cdot Q_2^2 \qquad (31)$$

4.6 Estimation of the Econometric Model of the Cost Structure

4.6.1 The Data

Data to estimate the model were taken from an updated version of the Memorandum on Productivity and Bell Canada Productivity, Bell Canada, April 3, 1978; updated as of February 8, 1980. These data were kindly made available to us by Bell Canada for the years 1952 to 1978. In addition, Bell Canada provided us with a decomposition of "other toll" data into series for WATS and competitive toll sufficient to compute price and quantity indices for the separate outputs.

The output measures used in the econometric estimation were constant dollar measures of: (1) monopoly toll revenue (message toll - within Bell Canada territory, within and outside Canda plus WATS), (2) competitive toll (private line plus TWX) and (3) local revenue (including miscellaneous).

Input measures used were constant dollar materials, labour manhours (adjusted for changing skill levels), and a measure of the real capital stock. A materials price index, net of indirect taxes, was constructed from the basic data base. An implicit wage deflator was used as the price of labour services. Details of the data construction for the above variables are given in Denny, Fuss and Waverman (1979).

The price of the capital input used is a user cost of capital services constructed to take into account the following influences: (1) the existence of a corporate income tax on profits; (2) the fact that financial capital is obtained from multiple sources with varying tax treatments, i.e., retained earnings, bond issues, common and preferred stock issues; and (3) the existence of depreciation for tax purposes in excess of economic depreciation. For details of the construction of this user cost of capital series, see Chapter 3.

4.6.2 The Estimated Cost Structure

The cost structure which was estimated consists of the cost function (14), two of the three cost share equations (16) and the two "revenue share" equations (19, 20). As we noted in Chapter 2, these 5 equations were estimated simulataneously with the demand equations for monopoly toll and competitive toll services, yielding a simultaneous system of 7 equations. Simultaneous estimation is the preferred procedure since the elasticities of demand parameters appear in both the demand and cost behavioural models. As well, the error terms in the two systems are likely to be contemporaneously correlated.

Since service outputs (Q_1,Q_2,Q_3) and non-local service prices (P_1,P_2) are endogenous in the cost/demand model, a simultaneous equations estimation procedure should be employed. The method chosen was iterative three stage least squares. The required instrumental variables consisted of variables exogenous to the demand and cost structures - factor prices, local service output price, the technology indicators A and S, real income, real domestic product, the number of households, population and a price index of all non-telecommunications goods and services. Iterative three stage least squares estimates of the parameters of the cost structure are presented in Table 4.1. The estimated demand structure has already been presented in Chapter 2. Summary statistics appear in Table 4.2. Table 4.3 contains the factor price elasticities matrix calculated at the means of the exogenous variables, while Table 4.4 presents the cost-output elasticities, also calculated at the mean observations.

The factor price elasticities (outputs held constant) indicate inelastic response to prices. The own price elasticity of demand for capital was slightly positive in an unconstrained regression. A test of the hypothesis that this price elasticity is zero can be performed at the mean by imposing the constraint $\gamma_{KK} = \alpha_K(1 - \alpha_K)$. A likelihood ratio test of this hypothesis yielded the Chi-squared test statistic 1.08. The critical value of this statistic is 3.84 (6.64) at the 5% (1%) level. We cannot reject the hypothesis and hence have imposed it as a maintained hypothesis. The fact that capital and labour are substitutes in production while capital and materials are complements is consistent with the very inelastic demand for capital. Note also that labour and materials are substitutes in production.

4.7 Empirical Tests of the Natural Monopoly Hypothesis

4.7.1 Overall Economies of Scale

With the calculation of cost-output elasticities in Table 4.4 we are ready to begin an examination of the estimated economies of scale and economies of scope. From Table 4.4 we note that the overall cost elasticity at the mean is 1.061 with a standard error of (0.072). A test of the hypothesis of constant returns to scale $(\varepsilon_{CQ} = \sum_j \varepsilon_{CQ_j} = 1)$ yields the test statistics 0.85. The critical value for a normally distributed variable is 1.96 (1.64) at the 5% (1%) significance level. We cannot reject the hypothesis of constant returns to scale (at the mean). Table 4.5 presents the time series estimates of the overall cost elasticity and $S = \varepsilon_{CQ}^{-1}$, the overall scale elasticity. A ninety-five percent confidence interval around $\varepsilon_{CQ} = 1$, using the standard error calculated at the mean is [1.14, 0.85] which contains

Table 4.1

Parameter Estimates
(Standard Errors in Brackets)

α_0	6.530 (0.007)	δ_{33}	-0.492 (0.161)
α_L	0.352 (0.003)	δ_{12}	-0.0233 (0.0047)
α_K	0.475 (0.004)	δ_{13}	0.0101 (0.0125)
α_M	0.173 (0.002)	δ_{23}	0.0190 (0.0126)
β_1	0.0943 (0.0148)	ρ_{L1}	0.0479 (0.0065)
β_2	0.0248 (0.0046)	ρ_{L2}	0.0308 (0.0058)
β_3	0.942 (0.079)	ρ_{L3}	-0.260 (0.024)
γ_{LL}	0.0744 (0.0119)	ρ_{K1}	-0.0585 (0.0080)
γ_{KK}	0.249 (0.001)	ρ_{K2}	-0.0252 (0.0057)
γ_{MM}	0.0790 (0.0108)	ρ_{K3}	0.310 (0.026)
γ_{LK}	-0.122 (0.006)	ρ_{M1}	0.0106 (0.0047)
γ_{LM}	0.0480 (0.0097)	ρ_{M2}	-0.00558 (0.00549)
γ_{KM}	-0.127 (0.006)	ρ_{M3}	-0.0499 (0.0117)
δ_{11}	-0.0356 (0.0083)	λ_1	-1.592 (0.276)
δ_{22}	-0.00210 (0.00605)	λ_2	-2.396 (0.399)
ε_1	-1.386 (0.081)	λ_3	-1.076 (0.093)
ε_2	-2.047 (0.391)	θ	0.473 (0.041)

Table 4.2

Summary Statistics

Equation	R^2	D.W. Statistic
Cost Function	.998	1.25
Labour Share	.977	1.43
Capital Share	.942	1.40
Message Toll "Share"	.957	1.50
Other Toll "Share"	.953	1.00

Table 4.3

Factor Price Elasticities*
(Evaluated at the Mean Observations)

	Labour	Capital	Materials
Labour	-0.437 (0.033)	0.127 (0.017)	0.310 (0.027)
Capital	0.0942 (0.0126)	0**	-0.0942 (0.0134)
Materials	0.629 (0.059)	-0.258 (0.034)	-0.371 (0.060)

* The first row presents the elasticity of the demand for labour, capital and materials respectively with respect to the price of labour. The other rows are interpreted in an analogous manner.

** The own price elasticity of capital was constrained to 0 at the mean by setting $\gamma_{KK} = \alpha_K(1-\alpha_K)$. Unconstrained estimation yielded a positive price elasticity which was insignificantly different from zero at conventional significance levels.

Table 4.4

Cost-Output Elasticities
(Evaluated at Mean Observations)

ε_{CQ_1}	ε_{CQ_2}	ε_{CQ_3}	$\sum_{j=1}^{3} \varepsilon_{CQ_j}$
0.0943 (0.0148)	0.0248 (0.0046)	0.942 (0.079)	1.061 (0.072)

Table 4.5

Overall Cost Elasticity and Scale Elasticity Estimates

Time	Cost Elasticity ε_{CQ}	Scale Elasticity S
1952	1.03477	.966398
1953	1.03933	.962158
1954	1.05515	.947732
1955	1.05778	.945378
1956	1.05537	.947536
1957	1.06246	.941213
1958	1.09673	.911802
1959	1.06903	.935429
1960	1.07633	.929087
1961	1.06832	.936050
1962	1.05488	.947979
1963	1.06630	.937819
1964	1.05725	.945853
1965	1.04863	.953628
1966	1.04083	.960770
1967	1.05674	.946305
1968	1.02340	.977132
1969	.984839	1.01539
1970	.995097	1.00493
1971	1.01682	.983458
1972	.994301	1.00573
1973	.986092	1.01410
1974	.975422	1.02520
1975	.978821	1.02164
1976	.993188	1.00686
1977	.965528	1.03570
1978	.911519	1.09707

all estimates (1952-78) of the overall cost elasticity. Based on the results contained in Table 4.5 there is little evidence with which to reject the hypothesis of constant overall returns-to-scale.

The above conclusion must be tempered by the realization that estimates of returns-to-scale are extremely sensitive to the specification of the cost model. To illustrate this point we present, in Table 4.6, estimates of the overall scale elasticity at the mean for some recent studies of Bell Canada's production technology with which the current authors have been associated. In Table 4.7 we present a description of the relevant features of the various models used. There are several important features of Tables 4.6 and 4.7 which should be noted. First as one reads from left to right, one encounters increased generality both in model specification and estimation technique. This is what one would expect from ongoing basic research aimed at providing improvements in methodology. However one also encounters considerable variation in scale elasticity estimates. This creates a real dilemma for policy decision-makers who wish to base their decisions, at least in part, on empirical estimates of scale economies. While there are a number of differences between the various studies, it is our belief that the major reason for the variation in scale elasticity estimates which appear in Table 4.6 can be found in functional form differences. Consider the last three columns of Tables 4.6 and 4.7. First, recall that the hybrid translog function approaches the ordinary translog function as θ approaches 0 . However at $\theta = 0$, the likelihood function becomes degenerate and hence this value cannot be imposed in estimating the hybrid function. Neverhteless the translog function can be approximated as closely as desired by choosing θ close to 0 . We have chosen $\theta = 0.01$. At that point, $\frac{Q_j^{*\theta} - 1}{\theta}$ is virtually identical to $\log Q_j^*$. Hence the column marked "this study, $\theta = 0.01$" is essentially

Table 4.6

Estimates of Mean Scale Elasticity

Dobell et al (1972)	Fuss-Waverman (1978)	Denny et al (1979)	This Study $\theta = 0.01$	This Study $\theta = 0.473$
1.11	1.05	1.47	1.43	0.94

Table 4.7

A Comparison of Basic Features of Studies of Bell Canada's Technology

Features	Dobell et al (1972)	Fuss-Waverman (1978)	Denny et al (1979)	This Study $\theta = 0.01$	This Study $\theta = 0.473$
Time Period	1952 - 67	1952 - 75	1952 - 76	1952 - 78	
Outputs	single aggregate output	Q_1 = message toll Q_2 = private line + WATS + TWX + miscellaneous Q_3 = local	Q_1 = message toll Q_2 = private line + WATS + TWX Q_3 = local + miscellaneous	Q_1 = message toll + WATS Q_2 = private line + TWX Q_3 = local + miscellaneous	
Functional Form	Cobb-Douglas Production Function	Translog Cost Function	Translog Cost Function	Hybrid Translog Cost Function, θ constrained to 0.01	Hybrid Translog Cost Function
Technical Change Specification	Hicks neutral - used technological indicator (direct distance dialing)	Capital augmenting - used time trend	Output-augmenting - used two technological indicators (A,S)		
Method of Estimation	Ordinary Least Squares estimation of the production function	Iterative 3SLS - cost and demand estimated separately	Iterative 3SLS - cost and demand estimated separately	Iterative 3SLS - cost and demand estimated simultaneously	

the result of estimating a translog model. The scale elasticity is almost the same as that estimated by Denny et al (1979), suggesting that the scale elasticity estimates were insensitive to the differences in data and estimation techniques employed. A comparison of the last two columns illustrates the sensitivity of the scale elasticity to variations in functional form, since the only difference is the difference between the hybrid translog and the ordinary translog-data and estimation procedures being identical. A likelihood ratio test of the null hypothesis $\theta = 0.01$ yields the test statistic 11.92. The Chi-squared critical value is 3.84 (6.64) at the 5% (1%) significance level. At any reasonable significance level the null hypothesis is rejected, which implies rejection of the ordinary translog model and its associated estimates of substantial overall scale economies.

In summary, the most general model and estimation procedure that has been utilized to date suggests that any overall economies of scale which exist are modest at best. We believe that this conclusion provides a reasonable working hypothesis for policy decision makers at the present time. However, given the demonstrated lack of robustness of this important measure to the evolution of research methodology, we cannot guarantee that the constant returns to scale hypotheses will not be overturned in the future by the results flowing from some new, improved methodology.

4.7.2 Product-Specific Economies of Scope

Having estimated the parameters of the model we are in a position to calculate product-specific economies of scope with respect to competitive toll services, and test for its significance. We begin by presenting, in Table 4.8, estimates of the degree of competitive toll specific

Table 4.8

The Degree of Competitive Toll Specific

Economies of Scope

Time	Degree of Economies of Scope SC_2
1952	.613248E-01
1953	.538024E-01
1954	.470743E-01
1955	.392027E-01
1956	.299881E-01
1957	.223822E-01
1958	.162975E-01
1959	.923726E-02
1960	.494543E-02
1961	.283012E-04
1962	-.657222E-02
1963	-.101956E-01
1964	-.132549E-01
1965	-.162239E-01
1966	-.188336E-01
1967	-.218277E-01
1968	-.280182E-01
1969	-.340089E-01
1970	-.378999E-01
1971	-.377575E-01
1972	-.419550E-01
1973	-.421173E-01
1974	-.395739E-01
1975	-.396333E-01
1976	-.380180E-01
1977	-.413711E-01
1978	-.467246E-01

economies of scope for the period 1952-78 calculated using equation (5). From the estimates it would appear there were mild economies of scope in the early part of the sample ($SC_2 > 0$) and mild diseconomies of scale in the latter part of the sample ($SC_2 < 0$). The degree of economies of scope at the mean is -0.017 with a standard error of 0.021. At the mean, we cannot reject the hypothesis that $SC_2 = 0$. Using the mean estimate of the standard error, SC_2 is insignificantly different from zero for 26 of the 27 data points. There would appear to be little evidence of economies of scope with respect to competitive toll services. In evaluating this statement, the reader should recall that the estimation of economies of scope requires extrapolation of the cost function into the region of zero outputs, a region where no actual observations exist.

4.7.3 Product-Specific Economies of Scale

In an earlier section of this chapter we discussed two aspects of competitive toll-specific economies of scale. First, we may be interested in whether Bell Canada produces its private line toll services subject to increasing returns to scale (declining average incremental cost). Second, and more importantly, we are interested in whether Bell would produce its competitor's (CNCP) private line services subject to increasing returns to scale if this output were added to its own output.

The first column of Table 4.9 presents estimates of the degree of product-specific scale economies (S_2) for Bell's own output of private line services. These estimates are calculated using equation (7). As can be noted, the degree of private-line specific scale economies is substantial. The value of S_2 at the mean is 2.24 (calculated using equation

Table 4.9

Competitive Toll Specific Economies of Scale

Time	Bell's Output	Market Output
1952	2.37323	
1953	2.32080	
1954	2.33623	
1955	2.35294	
1956	2.33253	
1957	2.32155	
1958	2.26586	
1959	2.19595	
1960	2.16454	1.24373
1961	2.14466	1.21851
1962	2.14757	1.19689
1963	2.15040	1.17836
1964	2.14193	1.15943
1965	2.13932	1.16076
1966	2.13945	1.16878
1967	2.15472	1.16771
1968	2.13408	1.15347
1969	2.12038	1.14204
1970	2.12791	1.13866
1971	2.13766	1.13777
1972	2.13266	1.12487
1973	2.14739	1.11894
1974	2.16433	1.11963
1975	2.18841	1.10942
1976	2.22477	
1977	2.21917	
1978	2.18741	

(29)) with a standard error of 0.63. A test of the hypotheses $S_2 = 1$ is rejected at the 5% and 1% significance levels. Bell Canada appears to produce private line services subject to increasing returns to scale.

In evaluating the above conclusion, two points should be noted. First, as indicated previously, all estimates of economies of scope and product specific economies of scale must be treated cautiously, since they are the result of extrapolating the cost function into the region of zero outputs. Second, the high estimates of S_2 probably are due to the spreading out of estimated "unavoidable" costs of production. While strictly speaking in the theoretical long-run "unavoidable" costs should be zero, so that $C(0,0,0) = 0$, for telecommunications this long-run may be very long indeed. Our estimates of unavoidable costs ($C(0,0,0)$) range from 16% of total costs in 1952 to 1% of total costs in 1978, with the mean value being 4%. The large relative value of $C(0,0,0)$ in the early years of the sample (the years of small outputs) may be a reflection of the very long-lived nature of capital equipment in telecommunications. However it may also be a warning about the problems associated with trying to extrapolate the cost function into the region of zero outputs.

The more interesting computation of private line scale economies relates to the nature of average incremental costs were Bell to become a monopoly supplier of private line services. Estimates of the degree of returns to scale ($\tilde{S}_2$) associated with the addition of CNCP's output to Bell's output are contained in column 2 of Table 4.9. These are computed using equation (30).[7] The estimates are much lower than those contained in column 1, supporting the "spreading out of unavoidable costs" hypothesis suggested earlier as an explanation of the large numbers in column 1. The estimates suggest increasing returns to scale with the degree of returns to scale falling over time. However the hypothesis of constant returns to scale cannot be rejected. We computed the degree of returns to scale and its approximate standard error for an observation in which the normalized market output was unity, using equation (31). At this observation Bell and CNCP had approximately equal shares of the market. The value of $\tilde{S}_2$ was 1.17 with a standard error of 0.15. In fact, a 95% confidence interval around $\tilde{S}_2 = 1$ contain all actual estimates of $\tilde{S}_2$. Subject to the usual caveats regarding estimates of product-specific economies of scale, there does not appear to be strong evidence of static efficiency gains with respect to scale economies from monopoly production of private line services.

Footnotes to Chapter Four

1. For a detailed discussion of the use of technical changes indicators in cost function specification see Denny, Fuss, Everson and Waverman (1979).

2. These include No. 5 Cross-bar, Electronic and SD1.

3. For further discussion of this behavioural model see Fuss and Waverman (1977). Recall that in this chapter we are not imposing rate of return regulation.

4. The Translog function can be used to provide a sufficient, but not necessary, (local) test for economies of scope (see Fuss and Waverman (1977)). However, because of the degeneracy of the cost function at zero outputs, it is a relatively weak test. This fact has been noted by Fuss and Waverman (1978) and Baumol, Fischer and Nadiri (1978).

5. For a derivation of these equation in the case of the ordinary trans-log cost function, see Fuss and Waverman (1977).

6. Note that the demand elasticities are denoted b_{11} and b_{21}, respectively, in Chapter 2 (Demand Estimation).

7. Estimates of CNCP's private line output can be obtained from testimony presented in the Interconnection case regarding trends in CNCP's market share over time.

Chapter Five

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

5.1 The Behavioural Model[1]

Bell Canada has been subject to regulation which limits the maximum rate-of-return which can be earned on invested capital since 1966. Before that year, Bell was limited in its surplus earnings per share. Maximum earnings per share do imply a ceiling on the rate of return on common equity.[2] It is well-known that rate of return regulation can bias the choice of inputs away from the cost-minimizing mix. Recently this hypothesis (known as the Averch-Johnson 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 due to misspecification of the behavioural 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. Suppose the product transformation function is given by

$$F(Y_1 \ldots Y_m; K, X_2, \ldots X_n) \leq 0 \tag{1}$$

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

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

subject to (1) and

$$\Sigma\, q_i Y_i - \Sigma\, p_j X_j \leq sK \tag{3}$$

where q_i, i=1, ... m are endogenous output prices, and s is the allowed rate of return.

The appropriate Lagrangian expression is

$$L = \sum_i q_i Y_i - \sum_j p_j X_j - p_k K + \lambda_1 [sK - \sum_i q_i Y_i + \sum_j p_j X_j]$$

$$+ \lambda_2 [- F(Y_1 \ldots Y_m; K, X_2, \ldots X_n)] \qquad (4)$$

If we assume that production is technologically efficient and the firm earns exactly the allowed rate of return (1) and (3) become equalities. Further, if we assume that the optimal solution results in non-zero Y_i and X_j for all i, j then the first order Kuhn-Tucker conditions for a maximum of (2) subject to (1) and (3) will involve no inequalities. These conditions are

$$\frac{\partial L}{\partial X_j} = - p_j(1-\lambda_1) - \lambda_2 \frac{\partial F}{\partial X_j} = 0 \qquad j = 2, \ldots n \qquad (5)$$

$$\frac{\partial L}{\partial K} = -(p_k - \lambda_1 s) - \lambda_2 \frac{\partial F}{\partial K} = 0 \qquad (6)$$

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

or

$$MR_i (1-\lambda_1) - \lambda_2 \frac{\partial F}{\partial Y_i} = 0 \qquad (7)$$

where MR_i is the marginal revenue of the ith output.

$$\frac{\partial L}{\partial \lambda_1} = sK - \sum_i q_i Y_i - \sum_j p_j X_j = 0 \qquad (8)$$

$$\frac{\partial L}{\partial \lambda_2} = - F(Y_1 \ldots Y_m; K, X_2, \ldots X_n) = 0 \qquad (9)$$

From (45) and (46) we obtain

$$\frac{\partial F}{\partial X_g} \Big| \frac{\partial F}{\partial X_\ell} = \frac{p_g(1-\lambda_1)}{p_\ell(1-\lambda_1)} = \frac{p_g^*}{p_\ell^*} \qquad g, \ell = 1, \ldots n \qquad (10)$$

and

$$\frac{\partial F}{\partial X_g} \Big| \frac{\partial F}{\partial K} = \frac{p_g(1-\lambda_1)}{p_k - \lambda_1 s} = \frac{p_g^*}{p_k^*} \qquad g = 1, \ldots n \qquad (11)$$

where p_g^* , p_ℓ^* , p_k^* are <u>shadow</u> prices of the inputs. Equations (10) and (11) 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 <u>as if</u> it minimized cost subject to the shadow prices. Therefore solving equations (8)-(11) we can obtain the producer's constrained multiproduct cost function

$$C^* = C^*(p_1^* \ldots p_n^*, Y_1, \ldots Y_m) \qquad (12)$$

Alternatively, utilizing the theory of duality between production and cost, we can start with the cost function (12) and assume that the producer acts <u>as if</u> he minimizes cost subject to the outputs and shadow prices appearing in (12). We know from the marginal conditions (10) and (11) 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 below. Utilizing Shephard's

Lemma once again we have

$$\frac{\partial C^*}{\partial p_j^*} = X_j \qquad j = 1, \ldots n \qquad (13)$$

Equations (13) 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 (8), (9), (10) and (11) determine the cost minimization solution subject to the production technology and the rate of return constraints. We will now show that equation (7), 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

$$\sum_{i=1}^{m} \frac{\partial F}{\partial Y_i} dY_i + \frac{\partial F}{\partial K} dK + \sum_{j=2}^{n} \frac{\partial F}{\partial X_j} dX_j = 0 \qquad (14)$$

Using (5) and (6), (7) 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

$$\lambda_2 \sum_{i=1}^{m} \frac{\partial F}{\partial Y_i} dY_i - [(p_k dK + \sum_{j=2}^{n} p_j dX_j) - \lambda_1(sdK + \sum_{j=2}^{n} p_j dX_j)] = 0 \quad (15)$$

Since

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

$$dC = p_k dK + \sum_{j=2}^{n} p_j dX_j \qquad (16)$$

In addition, from (8),

$$sdK + \Sigma\, p_j dX_j = d(\Sigma\, q_i Y_i) = dR \qquad (17)$$

Now suppose only Y_i changes, so that $dY_k = 0$, $k \neq i$. Then equation (15) becomes (using (16) and (17)),

$$\lambda_2 \frac{\partial F}{\partial Y_i} dY_i - [dC - \lambda_1 dR] = 0 \tag{18}$$

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

We can write equation (18) in the form

$$\lambda_2 \frac{\partial F}{\partial Y_i} = \frac{dC}{dY_i} - \lambda_1 \frac{dR}{dY_i} = MC_i - \lambda_1 MR_i \tag{19}$$

Substituting for $\lambda_2 \frac{\partial F}{\partial Y_i}$ in equation (7) we obtain

$$MR_i(1-\lambda_1) - [MC_i - \lambda_1 MR_i] = 0$$

or

$$MR_i = MC_i \tag{20}$$

Thus equation (7) is just the marginal revenue = marginal cost condition in somewhat disguised form.

The above interpretation of the first order conditions suggest 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 (5), (6), (8) and (9)). Second, conditional on the optimal input proportions choose outputs so as to equate marginal revenue to marginal cost (from equation (7)).

Because a sequential analysis can be applied in the case of rate of return regulation, the approach used in Chapter Four is applicable. That is, we shall 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

$$C^* = p_k^* K + \sum_{j=2}^{n} p_j^* X_j \tag{21}$$

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

$$\begin{aligned} C^* &= (p_k - \lambda_1 s)\cdot K + \sum_{j=2}^{n} p_j(1-\lambda_1)X_j \\ &= p_k \cdot K + \sum_{j=2}^{n} p_j X_j - \lambda_1[sK + \sum_{j=2}^{n} p_j X_j] \\ &= C(p_k, p_2, \ldots p_n, s, Y_1, \ldots Y_m) - \lambda_1 \Sigma q_i Y_i \end{aligned} \tag{22}$$

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

Now

$$\frac{\partial C}{\partial p_\ell} = \sum_j \frac{\partial C^*}{\partial p_j^*} \cdot \frac{\partial p_j^*}{\partial p_\ell} + \frac{\partial C^*}{\partial p_k^*} \cdot \frac{\partial p_k^*}{\partial p_\ell} + (\Sigma q_i Y_i)\frac{\partial \lambda_1}{\partial p_\ell} \quad j = 2, \ldots n \tag{23}$$

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

$$\begin{aligned} \frac{\partial C}{\partial p_\ell} &= X_\ell(1-\lambda_1) - (\sum_j p_j X_j)\frac{\partial \lambda_1}{\partial p_\ell} - sK\frac{\partial \lambda_1}{\partial p_\ell} + (\Sigma q_i Y_i)\frac{\partial \lambda_1}{\partial p_\ell} \\ &= X_\ell(1-\lambda_1) - \frac{\partial \lambda_1}{\partial p_\ell}[sK + \sum_j p_j X_j - \sum_i q_i Y_i] \\ &= X_\ell(1-\lambda_1) \text{ , using equation (7).} \end{aligned}$$

Thus we have a modified Shephard's Lemma:

$$\frac{\partial C}{\partial p_j} = X_j(1-\lambda_1) \qquad j = 2, \ldots n \tag{24}$$

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

$$\frac{\partial C}{\partial p_k} = \sum_j \frac{\partial C^*}{\partial p_j^*} \cdot \frac{\partial p_j^*}{\partial p_k} + \frac{\partial C^*}{\partial p_k^*} \cdot \frac{\partial p_k^*}{\partial p_k} + (\Sigma\, q_i Y_i)\frac{\partial \lambda_1}{\partial p_k}$$

$$= -\frac{\partial \lambda_1}{\partial p_k}(\sum_j p_j X_j) + K - \frac{\partial \lambda_1}{\partial p_k}(sK) + (\Sigma\, q_i Y_i)\frac{\partial \lambda_1}{\partial p_k}$$

$$= K \qquad (25)$$

$$\frac{\partial C}{\partial s} = \sum_j \frac{\partial C^*}{\partial p_j^*} \cdot \frac{\partial p_j^*}{\partial s} + \frac{\partial C^*}{\partial p_k^*} \cdot \frac{\partial p_k^*}{\partial s} + (\sum_i q_i Y_i)\frac{\partial \lambda_1}{\partial s}$$

$$= -\frac{\partial \lambda_1}{\partial s}(\Sigma\, p_j X_j) - \lambda_1 K - \frac{\partial \lambda_1}{\partial s}(sK) + (\Sigma\, q_i Y_i)\frac{\partial \lambda_1}{\partial s}$$

$$= -\lambda_1 K \qquad (26)$$

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

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

$$\frac{\partial C}{\partial p_k} = K \qquad (27)$$

$$\frac{\partial C}{\partial s} = -\lambda_1 K$$

Actual estimating equations can be formed by noting that

$$X_j/X_\ell = \frac{\partial C}{\partial p_j} \Big/ \frac{\partial C}{\partial p_\ell} \tag{28}$$

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

$$\lambda_1 = -\frac{\partial C}{\partial s} \Big/ \frac{\partial C}{\partial p_k} \tag{29}$$

Finally, the remaining equations in the profit maximizing model can be obtained from the equations $\frac{\partial R}{\partial Y_i} = \frac{\partial C}{\partial Y_i}$, where C is defined as in (22).

5.2 A Long-Run Translog Econometric Model Under Rate of Return Constraint

The cost function can be written in the form

$$C = C(p_k, p_2, \ldots p_n, s, Q_1^*, \ldots Q_m^*) \tag{30}$$

For ease of notation, let $p_k = p_1$ and $s = p_{n+1}$. Then the translog approximation to the cost function (30) is

$$\log C = \alpha_0 + \sum_{i=1}^{m} \alpha_i \log Q_i^* + \sum_{j=1}^{n+1} \beta_j \log p_j$$

$$+ \frac{1}{2} \sum_{i=1}^{m} \sum_{k=1}^{m} \delta_{ik} \log Q_i^* \log Y_k + \frac{1}{2} \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \log p_j \log p_k$$

$$+ \sum_{i=1}^{m} \sum_{j=1}^{n+1} \rho_{ij} \log Q_i \log p_j \quad + \tag{31}$$

The cost share equations become

$$\frac{\partial \log C}{\partial \log p_j} = \frac{p_j(1-\lambda_1)X_j}{C} = (1-\lambda_1)M_j = \beta_j + \sum_{k=1}^{n+1} \gamma_{jk} \log p_k$$

$$+ \sum_{i=1}^{m} \rho_{ij} \log Q_i^* \qquad j = 2, \ldots n \qquad (32)$$

$$\frac{\partial \log C}{\partial \log p_1} = \frac{p_1X_1}{C} = M_1 = \beta_1 + \sum_{k=1}^{n+1} \gamma_{1k} \log p_k + \sum_{i=1}^{m} \rho_{i1} \log Q_i^* \qquad (33)$$

$$\frac{\partial \log C}{\partial \log p_{n+1}} = \frac{p_{n+1}(-\lambda_1X_1)}{C} = -\lambda_1M_{n+1} = \beta_{n+1} + \sum_{k=1}^{n+1} \gamma_{n+1,k} \log p_k$$

$$+ \sum_{i=1}^{m} \rho_{i,n+1} \log Q_i^* \qquad (34)$$

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

The cost system to be estimated consists of equations (31), (33) and equations of the form:

$$\frac{M_j}{M_2} = \frac{\beta_j + \sum_{k=1}^{n+1} \gamma_{jk} \log p_k + \sum_{i=1}^{m} \rho_{ij} \log Q_i^*}{\beta_2 + \sum_{k=1}^{n+1} \gamma_{2k} \log p_k + \sum_{i=1}^{m} \rho_{i2} \log Q_i^*} \qquad j = 3, \ldots n \qquad (35)$$

One the parameters have been estimated, λ_1 can be obtained from the ratio of (33) and (34). In addition to the cost share equations, we have, as before the revenue "share" equations obtained from the $MC_i = MR_i$ optimally conditions:

$$R_i = \frac{q_iQ_i}{C} = \left[\alpha_i + \sum_{i=1}^{m} \delta_{ik} \log Q_k^* + \sum_{j=1}^{n+1} \rho_{ij} \log p_j\right] (1+\frac{1}{\varepsilon_i})^{-1} \qquad (36)$$

where ε_i is the own price elasticity of demand.

The model represented by equations (34), (35) and (36) was estimated for the period 1952-76. The results proved disappointing. Table 1 provides estimates of $\frac{\partial \log C}{\partial \log p_k}$, $\frac{\partial \log C}{\partial \log s}$ and λ. It should be the case that $\frac{\partial \log C}{\partial \log p_k} > 0$, $\frac{\partial \log C}{\partial \log s} > 0$ and $0 \leq \lambda \leq \frac{p_k}{s}$. The unreasonable values of $\frac{\partial \log C}{\partial \log s}$ and λ for much of the period illustrates the unsatisfactory nature of the current set of results.

Table 5.1

Summary Results for Long-Run Rate of Return Model

	$\frac{\partial \log C}{\partial \log p_k}$	$\frac{\partial \log C}{\partial \log s}$	λ
~~1953~~	~~.437672~~	~~-1.16552~~	~~2.78804~~
1954	.433955	-1.14975	2.69083
1955	.423774	-.968585	2.20834
1956	.374664	.245160	-.628705
~~1957~~	~~.394192~~	~~.256583~~	~~-.651704~~
1958	.444178	-.302159	.747821
1959	.435095	-.232232E-02	.560619E-02
1960	.458506	-.147777	.354337
~~1961~~	~~.457845~~	~~-.205339~~	~~.473939~~
1962	.448276	-.471653E-01	.106475
1963	.466220	-.515371E-01	.118768
1964	.468512	.621306E-01	-.143432
~~1965~~	~~.466465~~	~~.152085~~	~~-.349488~~
1966	.457313	.351241E-01	-.773059E-01
1967	.482187	.320752E-01	-.726974E-01
1968	.483001	.383255	-.869740
~~1969~~	~~.454932~~	~~.351134~~	~~-.735433~~
1970	.463015	.633070E-01	-.128428
1971	.479829	-.312575	.619983
1972	.474564	-.337963	.638737
~~1973~~	~~.476712~~	~~-.341724~~	~~.648988~~
1974	.477012	-.370188	.708941
1975	.494198	-.485561	.937343
1976	.514800	-.522256	1.05467

5.3 A Conditional Translog Econometric Model Under Rate of Return Constraint

In this section we derive a rate-of-return constraint analysis under the assumption that capital can be considered a pre-determined variable for the purposes of econometric estimation. In effect we are eliminating from consideration any attempt to explain capital accumulation. We begin by assuming that the utility minimizes the cost of producting the vector of outputs Q_1, Q_2, Q_3 conditional on the capital stock at the beginning of the period (K_{-1}). In that case there exists a (short-run) variable cost function.

$$VC = VC(P_L, P_M, K_{-1}, Q_1, Q_2, Q_3, A, S) \tag{37}$$

with the following properties:

(a) VC is concave in P_L, P_M

(b) VC is linear homogeneous in P_L, P_M

(c) VC is increasing in P_L, P_M, Q_1, Q_2, Q_3 and decreasing in K_{-1} (monotonicity)

(d) $\frac{\partial VC}{\partial p_L} = L$, $\frac{\partial VC}{\partial p_M} = M$ (Shephard's Lemma)

The conditional profit maximizing problem for the rate-of-return regulated utility can now be written as: choose outputs Q_i and output prices p_1, p_2, so as to maximize:

$$\text{Profit} = \sum_{i=1}^{3} p_i Q_i - VC(P_L, P_M, K_{-1}, Q_1, Q_2, Q_3, A, S) - P_K \cdot K_{-1} \tag{38}$$

subject to

$$\Sigma p_i Q_i - VC(P_L, P_M, K_{-1}, Q_1, Q_2, Q_3, A, S) \leq sK_{-1} \tag{39}$$

K_{-1} predetermined, and p_3 fixed by the regulatory commission.

The Lagrangian expression for the above problem is

$$\mathcal{L} = \Sigma p_i Q_i - VC - P_K K_{-1} + \mu[sK_{-1} - \Sigma P_i Q_i + VC] \tag{40}$$

Optimizing with respect to Q_i yields

$$\frac{\partial \mathcal{L}}{\partial Q_i} = p_i + Q_i \frac{\partial p_i}{\partial Q_i} - \frac{\partial VC}{\partial Q_i} - \mu(p_i + Q_i \frac{\partial p_i}{\partial Q_i}) + \mu \frac{\partial VC}{\partial Q_i} = 0$$

$$MR_i(1-\mu) - MC_i(1-\mu) = 0$$

or

$$MR_i = MC_i \qquad i = 1,2 \qquad (41)$$

where MR_i = marginal revenue and MC_i = short-run marginal cost .

The estimating equations are obtained from Shephard's Lemma and the marginal conditions $MR_i = MC_i$, i = 1,2,3 . Suppose we assume that the variable cost function takes the translog form

$$\begin{aligned} \log VC = {} & \alpha_0 + \sum_{i \neq K} \alpha_i \log P_i + \alpha_K \log K_{-1} + \sum_k \beta_k \log Q_k^* \\ & + \tfrac{1}{2} \sum_{i=K} \gamma_{ii} (\log P_i)^2 + \gamma_{KK} (\log K_{-1})^2 \\ & + \sum_{\substack{i \neq j \\ i,j \neq K}}\sum \gamma_{ij} \log P_i \log P_j + \sum_{i \neq K} \gamma_{iK} \log P_i \log K_{-1} \qquad (42) \\ & + \tfrac{1}{2} \sum_k \delta_{KK} (\log Q_k^*)^2 + \sum_{k\ell}\sum \delta_{k\ell} (\log Q_k^* \log Q_\ell^*) \\ & + \sum_{i \neq K} \sum_k \rho_{ik} \log P_i \log Q_k^* + \sum_k \rho_{Kk} \log Q_k^* \log K_{-1} \end{aligned}$$

where $i,j = L,M$

$k,\ell = 1,2,3$

The cost share equations are

$$S_i = \alpha_i + \sum_i \gamma_{ij} \log P_j + \gamma_{ik} \log K_{-1} + \sum_k \rho_{ik} \log Q_k^* \qquad (43)$$

$i,j = L,M$

The revenue "share" equations are

$$\frac{P_1 Q_1}{VC} = \left[\frac{1}{1+\frac{1}{\varepsilon_1}}\right]^{-1} \cdot \left[\beta_1 + \sum_{\ell} \delta_{1\ell} \log Q^*_{\ell} + \sum_{i} \rho_{i1} \log P_i + \rho_{K1} \log K_{-1}\right. \qquad (44)$$

$$\frac{P_2 Q_2}{VC} = \left[\frac{1}{1+\frac{1}{\varepsilon_2}}\right]^{-1} \cdot \left[\beta_2 + \sum_{\ell} \delta_{2\ell} \log Q^*_{\ell} + \sum_{i} \rho_{i2} \log P_i + \rho_{K2} \log K_{-1}\right. \qquad (45)$$

The equations to be estimated are: equation (42), one of the two equations (43) and equations (44) and (45).

The value of the Lagrangian multiplier can be obtained from a consideration of the long-run choice of capital K . Maximizing profit with respect to K yields the first order condition

$$\frac{\partial \mathcal{L}}{\partial K} = -\frac{\partial VC}{\partial K} - P_K + \mu s + \mu \frac{\partial VC}{\partial K} = 0$$

or

$$-\frac{\partial VC}{\partial K} = \frac{P_k - \mu s}{1 - \mu} = P^*_K \qquad (46)$$

The right hand side of (46) is the shadow price of capital P^*_K (relative to P_M, P_L) . The left hand side can be computed once the parameters of VC are known. It can be shown that (46) can be solved for μ in the form

$$\mu = \frac{M_1 + \frac{\partial \log VC}{\partial \log K}}{M_2 + \frac{\partial \log VC}{\partial \log K}} \qquad (47)$$

where $M_1 = \frac{P_K \cdot K}{VC}$

$M_2 = \frac{sK}{VC}$

Table 5.2 presents estimates of the parameters of the translog variable cost function VC . Table 5.3 presents a time series of shadow prices P_K^*, Lagrangian multipliers μ, and the theoretical upper limit of μ : P_K/s . From the period 1952-59. $P_K^* < 0$ which violates the monotonicity requirements for VC . While the remainder of the P_K^*'s satisfy the monotonicity requirement they imply values of μ which appear to be unreasonably close to the theoretical upper limits.

We also attempted to estimate, for the period 1952-78, the variable hybrid translog cost function which takes the form

$$\begin{aligned}
\log C = {} & \alpha_0 + \sum_{i \neq K} \alpha_i \log P_i + \sum_k \beta_k \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] + \alpha_K \log K_{-1} \\
& + \tfrac{1}{2} \sum_{i \neq K} \gamma_{ii} (\log P_i)^2 + \sum_{\substack{i \\ i \neq j \\ i,j \neq K}} \sum_j \gamma_{ij} \log P_i \log P_j + \gamma_{KK} (\log K_{-1})^2 \\
& + \tfrac{1}{2} \sum_k \delta_{kk} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right]^2 + \sum_{\substack{k \\ k \neq \ell}} \sum_\ell \delta_{k\ell} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] \left[\frac{Q_\ell^{*\theta} - 1}{\theta}\right] \\
& + \sum_{\substack{i \\ i \neq K}} \sum_k \rho_{ik} \log P_i \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] + \sum_k \rho_{Kk} \log K_{-1} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] \qquad (48)
\end{aligned}$$

where $i,j = L,M$

$k,\ell = 1,2,3$

The remainder of the estimating equations for the hybrid translog model take the form:

$$S_i = \alpha_i + \sum_i \gamma_{ij} \log P_j + \sum_k \rho_{ik} \left[\frac{Q_k^{*\theta} - 1}{\theta}\right] + \gamma_{iK} \log K_{-1} \qquad (49)$$

$i = L, M \quad ; \quad k = 1,2,3$

$$\frac{P_1 Q_1}{C} = \left[\frac{1}{1 + 1/\varepsilon_1}\right]^{-1} \cdot \left[Q_1^{*\theta}\right] \cdot \left\{\beta_1 + \sum_\ell \delta_{1\ell}\left[\frac{Q_\ell^{*\theta} - 1}{\theta}\right] + \sum_i \rho_{i1} \log P_k + \rho_{K1} \log K_{-1}\right\} \qquad (50)$$

$$\frac{P_2 Q_2}{C} = \left[\frac{1}{1 + 1/\varepsilon_2}\right]^{-1} \cdot \left[Q_2^{*\theta}\right] \cdot \left\{\beta_2 + \sum_\ell \delta_{2\ell}\left[\frac{Q_\ell^{*\theta} - 1}{\theta}\right] + \sum_i \rho_{i2} \log P_i + \rho_{K2} \log K_{-1}\right\} \qquad (51)$$

$i = L,M$

$\ell = 1,2,3$

Estimates of the parameters of the above model also proved disappointing. For all sample points the estimated shadow price of capital was negative, violating the monotonicity regularity conditions. From an analysis of the results it seems clear that the major difficulty in estimating the parameters of the variable cost function both in its ordinary and hybrid translog forms is multicollinearity. The fact that capital is a right-hand side variable adds another strongly trended variable to the list of regressors and appears to render estimation of the variable cost function with the current set of data extremely problematic.

Table 5.2

Parameter estimates - Three Output Variable Ordinary Translog Cost Function
(Standard Errors in Brackets)

α_o	-3.041 (13.704)	δ_{33}	-0.198 (0.408)
α_L	1.340 (0.358)	δ_{12}	0.0142 (0.0258)
α_K	2.874 (3.803)	δ_{13}	-0.0298 (0.0336)
α_M	0.340 (0.358)	δ_{23}	-0.00224 (0.0137)
β_1	0.226 (0.735)	ρ_{L1}	0.00464 (0.0161)
β_2	1.229 (0.541)	ρ_{L2}	0.00841 (0.0132)
β_3	2.020 (1.182)	ρ_{L3}	-0.0159 (0.020)
γ_{LL}	-0.0134 (0.0298)	ρ_{K1}	-0.0144 (0.0948)
γ_{KK}	-0.465 (0.527)	ρ_{K2}	-0.158 (0.0692)
γ_{MM}	0.0134 (0.0298)	ρ_{K3}	-0.325 (0.219)
γ_{LK}	-0.0920 (0.0454)	ρ_{M1}	-0.00464 (0.0161)
γ_{LM}	0.0134 (0.0298)	ρ_{M2}	0.00841 (0.0132)
γ_{KM}	-0.0920 (0.0454)	ρ_{M3}	-0.0159 (0.020)
δ_{11}	-0.284 (0.0203)	λ_1	-1.112 (0.516)
δ_{22}	0.0596 (0.0165)	λ_2	1.597 (0.674)
		λ_3	-7.009 (6.282)

Table 5.3

Shadow Prices, Lagrangian Multipliers and Upper Limits to Lagrangian Multipliers

	Shadow Price p_k^*	Lagrangian Multiplier μ	Upper Limits to Lagrangian Multipliers
1953	-.126232	.964592	.931191
1954	-.106326	.945762	.903621
1955	-.806523E-01	.912705	.855645
1956	-.569402E-01	.901229	.85956[illegible]
1957	-.388490E-01	.918046	.890816
1958	-.174070E-01	.950739	.970082
1959	-.322453E-02	.936309	.934517
1960	.115110E-01	.975741	.978165
1961	.230502E-01	.926440	.940222
1962	.269915E-01	.871736	.900195
1963	.306666E-01	.941184	.955435
1964	.360722E-01	.945829	.962315
1965	.394126E-01	.930405	.953722
1966	.426968E-01	.843252	.895525
1967	.431180E-01	.959248	.972350
1968	.464793E-01	.960670	.975230
1969	.548447E-01	.742244	.847761
1970	.580705E-01	.738407	.835714
1971	.645962E-01	.764924	.846776
1972	.691380E-01	.693734	.798004
1973	.742071E-01	.702517	.805518
1974	.758359E-01	.723077	.812781
1975	.837233E-01	.779410	.848814
1976	.894076E-01	.876016	

Footnotes to Chapter Five

1. This section is derived from Fuss and Waverman (1977).

2. However, note that in Chapter 3 we showed that the actual rate of return on common equity exceeded the allowed rate-of-return in 1952, 1953, 1954, 1955, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1967, 1968, 1971 and 1975.

CHAPTER SIX

The Rate Structure of Bell Canada - Some Considerations of Efficiency and Equity

6.1 Introduction

In this chapter we will use the estimated demand and cost structures (chapters two and four) in order to discuss efficiency and equity aspects of Bell Canada's rate structure. We begin with some introductory remarks.

A multiproduct firm in a perfectly competitive industry will be characterized, in long-run equilibrium, by overall constant returns to scale (locally). In that case it is well known that the efficient (Pareto Optimal) prices are those which equal marginal costs. For any firm subject to constant returns to scale, marginal cost pricing will yield revenues which cover total costs and is the efficient pricing scheme in the sense that consumers' marginal willingness to pay matches the marginal costs of production. For a firm subject to increasing returns to scale, marginal cost pricing will not yield revenues which cover total costs, so that a different pricing rule, one in which at least one price exceeds marginal cost, is required. The pricing rule most often advocated in the case of increasing returns to scale is the Ramsey Rule. When demand for the multiple outputs is characterized by zero cross price elasticities, the Ramsey Rule states that prices should be raised above marginal costs in inverse proportion to demand elasticities. Hence the rule is often called the Inverse Elasticity Rule. In particular the prices of products i and j should be set so that

$$R_{i,j} = \frac{\dfrac{p_i - MC_i}{p_i}}{\dfrac{p_j - MC_j}{p_j}} = \frac{\varepsilon_j}{\varepsilon_i} \qquad (1)$$

where p_i, p_j are prices of outputs i and j

MC_i, MC_j are marginal costs of outputs i and j

ε_i, ε_j are the own price elasticities of demand for outputs i and j

The Ramsey Rule can be shown to be "quasi-efficient" (constrained Pareto Optimal) when the firm is subject to the constraint that it earn a level of profit which is independent of the particular level of services provided.[1]

The Ramsey Rule can be compared with the profit-maximizing pricing rule chosen by an unconstrained monopolist. The monopolist sets marginal revenues (MR_i) equal to marginal costs (MC_i). Since $MR_i = p_i(1 + \frac{1}{\varepsilon_i})$ when demands are independent, the monopolist solves the equations

$$p_i \left(1 + \frac{1}{\varepsilon_i}\right) = MC_i \qquad i = 1, \ldots N \qquad (2)$$

If we form a ratio of the ith and jth equations of (2), and rearrange this ratio we obtain

$$\frac{\dfrac{p_i - MC_i}{p_i}}{\dfrac{p_j - MC_j}{p_j}} = \frac{\dfrac{1}{\varepsilon_i}}{\dfrac{1}{\varepsilon_j}} \qquad (3)$$

which is equivalent to (1). If demands are independent an unconstrained profit maximizing monopolist will choose the Inverse Elasticity Rule. The difference between the monopolist's pricing policy and the Ramsey Rule is that the monopolist's profit level is unconstrained. Hence while an unconstrained monopolist will follow the Ramsey Rule in setting relative prices, absolute prices can be expected to be higher in the case of the unconstrained monopolist. One interesting aspect of the above analysis is that a monopolist with an unconstrained choice of rate structure, but constrained to earn a constant level of profit (perhaps zero) and wishing to deter entry, will choose the rate structure dictated by the Ramsey Rule.[2] The same phenomenon will be true for a monopolist subject to rate of return regulation if the outputs it produces have the same capital intensities of production. If capital intensities differ, there may be a lowering of the relative prices of the capital intensive services below that suggested by the Ramsey Rule.

Many services which might be labelled "necessities" are characterized by inelastic demand (e.g. basic local service). The Inverse Elasticity Rule implies that these services should be priced proportionately high relative to marginal costs, when compared with more elastically demanded services (e.g. long-distance message toll). If the expenditure on basic local service implied by the Ramsey Rule were to constitute a significant proportion of the income of low income consumers, and compensating lump sum income transfers are not possible, consideration of equity might suggest a modification of the Rule. One possible modification is that suggested by Feldstein (1972). He introduced an income distributional characteristic aspect into the Ramsey Rule formulation. The effect of Feldstein's formulation is to lower the price of any service whose consumption is concentrated in low income families relative to the level dictated by the Ramsey Rule.

6.2 A Comparison of Actual Prices and Marginal Costs for Bell Canada

We begin our quantitative analysis of Bell Canada's rate structure by comparing the actual prices and marginal costs over the 1952-78 period. The marginal costs are obtained as

$$MC_i = \frac{C}{Q_i} \cdot \varepsilon_{CQ_i} \quad (4)$$

where C is total cost and ε_{CQ_i} is calculated from equation (22) of Chapter 4. Table 6.1 presents prices and marginal costs for the three services. Both prices and marginal costs are normalized by the 1967 prices so that $p_i = 1$ in 1967 in order to facilitate comparison.

From Table 6.1 it can be seen that prices exceed marginal costs for monopoly and competitive toll services and fall short of marginal costs for local services. Since the efficiency rules discussed above would dictate that prices be no less than marginal costs, the current rate structure is inefficient. The current rate structure would also not be chosen by a profit-maximizing monopolist. The behaviourial assumption, used in formulating the cost model, that local rates are constrained by the regulatory commission would appear to be well-founded. While there exists <u>some</u> welfare function which would produce the results depicted in

Table 6.1

Prices and Marginal Costs of Services

Time	Monopoly Toll		Private Line		Local	
	P_1	MC_1	P_2	MC_2	P_3	MC_3
1952	1.00785	.289930	.999682	.724654	.907587	1.18059
1953	1.00867	.302334	1.02559	.755338	.915369	1.19548
1954	1.00991	.313949	1.04126	.783526	.915738	1.20144
1955	1.00946	.309827	1.04126	.705065	.919688	1.23744
1956	1.00855	.306104	1.04126	.636229	.921537	1.26298
1957	1.00353	.306382	1.04126	.620192	.922599	1.27662
1958	1.01175	.326079	1.04126	.705122	.929794	1.26183
1959	1.05497	.325455	1.04126	.675813	.990319	1.38576
1960	1.06775	.331377	1.04126	.680059	.992828	1.37637
1961	1.05488	.330977	1.04126	.662026	.991218	1.36366
1962	.996501	.317139	1.04249	.620637	.992393	1.39065
1963	.994841	.321076	1.04382	.599029	.998952	1.39319
1964	.993399	.306828	1.04259	.521236	.999002	1.42974
1965	1.03736	.307328	1.01774	.525294	.999852	1.47071
1966	1.00991	.308930	1.00133	.552406	1.00008	1.45807
1967	1.00000	.305335	1.00000	.552579	1.00000	1.47285
1968	.991137	.306894	.999487	.540959	1.00095	1.46469
1969	.994927	.308316	1.02436	.546890	1.00780	1.48144
1970	1.07338	.320209	1.02448	.572959	1.02144	1.52980
1971	1.08791	.350308	1.05826	.660914	1.06097	1.50439
1972	1.10443	.355200	1.07194	.674576	1.09859	1.55802
1973	1.11928	.342537	1.09419	.648950	1.12427	1.65312
1974	1.13510	.349455	1.12777	.676733	1.15454	1.68258
1975	1.17539	.356120	1.18937	.695589	1.21409	1.76589
1976	1.23939	.374181	1.30529	.723887	1.29163	1.85926
1977	1.27175	.417602	1.38926	.855020	1.37316	1.82198
1978	1.33947	.422143	1.59089	.858166	1.50058	2.05763

Table 6.1 using Feldstein 's equity formulation, this function would have to be extremely skewed towards users of local services.

Table 6.2 compares the actual price-marginal cost markups with those implied by the Inverse Elasticity Rule. In the next section of this chapter we will consider the appropriateness of this Rule. At this point the reader should note that the actual relative markups are very different from those implied by the Inverse Elasticity Rule, except for the monopoly toll and private line relative markups. The fact that the monopoly toll/private line relative markup obeys the Inverse Elasticity Rule should not be surprising since the behavioural model assumes profit maximizing behaviour with respect to the choice of the prices of these services.

6.3 An Analysis of Alternative Rate Structures

In this section we consider the effects of several alternative rate structures. These effects are obtained by imposing a particular rate structure and then solving the simultaneous demand and cost equations estimated in Chapters 2 and 4 for prices, outputs and marginal costs. Since the demand equations contain lagged endogenous variables to the second degree, both dynamic and static simulated solutions are possible. In the dynamic case, solutions for years after 1954 reflect the cumulative effects of the changes in prior years. In the static case, solutions reflect only the effect of the change in the particular year in question, since lagged values are taken to be the actual values. While we are able to obtain both sets of solutions, we present only the static solutions since the "comparative statics" results appear to be the most relevant.

Table 6.2

Comparison of Actual Markups and the Inverse Elasticity Rule

Time	Monopoly Toll/Private Line		Monopoly Toll/Local		Local/Private Line	
	R_{12}	$\varepsilon_2/\varepsilon_1$	R_{13}	$\varepsilon_3/\varepsilon_1$	R_{32}	$\varepsilon_2/\varepsilon_3$
1952	2.58921		-2.36806		-1.09338	
1953	2.65744		-2.28838		-1.16128	
1954	2.78412		-2.20885		-1.26044	
1955	2.14659		-2.00603		-1.07006	
1956	1.79055		-1.87979		-.952526	
1957	1.71791		-1.81041		-.948909	
1958	2.09934		-1.89777		-1.10621	
1959	1.97028		-1.73177		-1.13772	
1960	1.98809		-1.78521		-1.11364	
1961	1.88420		-1.82146		-1.03445	
1962	1.68474		-1.69881		-.991716	
1963	1.58936		-1.71609		-.926155	
1964	1.38211		-1.60291		-.862248	
1965	1.45442	1.48	-1.49436	0.20	-.973276	7.33
1966	1.54820		-1.51566		-1.02147	
1967	1.55260		-1.46910		-1.05684	
1968	1.50483		-1.49011		-1.00988	
1969	1.48056		-1.46883		-1.00799	
1970	1.59209		-1.40988		-1.12924	
1971	1.80597		-1.62248		-1.11309	
1972	1.83003		-1.62216		-1.12814	
1973	1.70543		-1.47529		-1.15600	
1974	1.73062		-1.51334		-1.14357	
1975	1.67891		-1.53358		-1.09476	
1976	1.56726		-1.58850		-.986627	
1977	1.74654		-2.05481		-.849975	
1978	1.48694		-1.84481		-.806008	

From the point of veiw of economic efficiency, the most interesting simulations would be those associated with marginal cost pricing and Ramsey quasi-optimal pricing. As can be seen from Tables 6.1 and 6.2, both pricing schemes yield rates very different from the actual rates. Preliminary simulations indicated that marginal cost pricing would result in an increase of 70% in the local service price and declines of 70% and 50% in monopoly toll and competitive toll prices respectively. Under this pricing scheme both toll services' outputs would have doubled. Such large movements in prices and outputs place the new equilibrium data well outside the data used to estimate the model. This result suggests that the estimated parameter values of both the cost and demand models are likely to be inappropriate at the new equilibrium. In particular it is difficult to believe that the demand elasticities would remain constant at their estimated values. Hence we will not present detailed results of the effects of marginal cost pricing, since they are probably unreliable.

Even more extreme price and output changes occur if one attempts to impose Ramsey Rule pricing. The relative prices between toll and local services would have to change by about 800% (local increasing, toll decreasing) if the parameters of the model did not change. Clearly, the parameters of the model would change. Suppose we contemplated a four-fold increase in the local service price. We could expect reductions in local service demand considerably in excess of those predicted from a demand elasticity of -0.27.

Both of the above efficiency solutions cannot be simulated reliably using models estimated from existing data. But they are also unlikely to be realistic policy options since they imply large abrupt changes from current (and historical) practice. To bring the analysis back into the region of realistic policy options, we consider a case where the perturbation from the status quo is relatively small. It seems clear from a comparison of actual prices and marginal costs that welfare can be improved by reducing the monopoly toll price down towards its marginal cost and raising the local service price up towards its marginal cost. We proceed to ask the following question. Suppose we reduce the monopoly toll rate by 20%, leave the private line rate unchanged, and allow Bell to earn the same level of profits as before the rate change. What is the required increase in local service rates which will accomplish these objectives? Tables 6.3 and 6.4 present the static results of such a policy simulation. In Table 6.3 we present the new prices, marginal costs, and quantities for monopoly toll and local service (selected years). Private line prices and quantities remain unchanged. Table 6.4 presents the percentage changes in prices and quantities which result from the new rate structure. A 20% reduction in the monopoly toll rate requires about a 15% increase in the local service rate to maintain a constant profit level for Bell. Monopoly toll output increases about 35% whereas local service output decreases about 3%. That an increase in welfare is likely from this change in rate structure can be seen from the fact that a large increase in long distance output requires only a small decrease in local service output.[3]

Table 6.3

Simulated Results From a 20% Reduction in the Monopoly Toll Rate

Time	Monopoly Toll			Local		
	P_1	MC_1	Q_1[a]	P_3	MC_3	Q_3[a]
1955	0.81	0.25	0.38	1.01	1.27	0.41
1960	0.85	0.27	0.58	1.10	1.42	0.63
1965	0.83	0.24	0.95	1.10	1.52	0.85
1967	0.80	0.24	1.18	1.12	1.52	1.00
1970	0.86	0.25	1.64	1.17	1.59	1.23
1975	0.94	0.27	2.93	1.41	1.87	1.68
1978	1.07	0.31	3.88	1.75	2.20	2.00

[a] The mean value is 1.0 before reduction in toll rate.

Table 6.4

Percentage Changes in Prices and Quantities

Resulting From a 20% Reduction in the Monopoly Toll Rate

Time	Monopoly Toll		Local	
	P_1	Q_1	P_3	Q_3
1955	-0.20	0.28	0.09	-0.02
1960	-0.20	0.33	0.11	-0.01
1965	-0.20	0.37	0.10	-0.03
1967	-0.20	0.34	0.12	-0.02
1970	-0.20	0.37	0.14	-0.02
1975	-0.20	0.34	0.16	-0.03
1978	-0.20	0.35	0.17	-0.03

Footnotes to Chapter Six

1. Candidates for the constrained level of profit would be those associated with (i) the actual rate-of-return on capital, (ii) the allowed rate-of-return and (iii) the actual cost of capital services.
2. For an analysis of this case see Baumol, Bailey and Willig (1977).
3. Real resource expenditures remain approximately constant in this simulation, since total cost increases by about 1%.

CHAPTER 7: CONCLUSIONS AND POLICY RECOMMENDATIONS

7.1 Introduction

A primary concern of public policy decision-makers is the extent to which the telecommunications sector is a natural monopoly, hence perhaps rendering competition undesirable. A secondary concern relates to the structure of rates charged for the various services of telecommunications firms. Are these rates efficient and equitable? If not, what would be a desirable change in the rate structure?

This study has considered both of the above telecommunications issues using data drawn from Bell Canada's production and financial accounts. We have conducted empirical tests of the natural monopoly hypothesis and investigated the welfare aspects of the current rate structure and several potential alternative structures. Two main conclusions flow from our study:

(1) Hypothesis tests based on the most general model of the telecommunications sector's production technology estimated to date show that there is little evidence to suggest that Bell Canada is a natural monopoly with respect to all its principal service offerings. In particular, tests of overall economies of scale and tests of economies of scale and scope with respect to private line services fail to reject the hypothesis that private line services can be provided on a competitive basis without efficiency loss. Our hypothesis tests <u>cannot prove</u> that Bell Canada is <u>not</u> a natural monopoly, <u>but</u> the competitive alternative is <u>not</u> rejected by the data.

(2) The current structure of Bell Canada's rates is inefficient. By the efficiency criterion, long distance message toll rates are too high and local service rates are too low. For example, a decrease in long distance rates of 20% accompanied by a 15% increase in local service rates leaves Bell Canada's profit level unchanged and results in a welfare improvement.

7.2 The Natural Monopoly Issue

In Chapter One we surveyed the evidence, prior to this study, concerning the natural monopoly characteristics of telecommunications. The majority of studies surveyed purport to show the existence of economies of scale in telecommunications production. However these studies are sufficiently flawed to render the past accumulation of evidence very weak. Our more general formulation provides results which do not support the existence of scale economies. However, conclusions regarding existence or non-existence of scale economies appear to be very sensitive to functional form specification. Hence any policy recommendations which rely on empirical estimates of scale economies must, at best, be tentative. We prefer to take the approach which is probably advocated by most economists - that competition should be encouraged unless there is strong evidence that competition would result in inefficient production. This approach is consistent with recent interpretations of the Railway Act. Until a more extensive data base than we had access to becomes available, the scale economies issue will remain ultimately unresolved. Nevertheless, at this time we can conclude there exists no evidence on scale economies, which is robust to reasonable alternative specifications, that would contradict the recent U.S. Federal Communications

Commission and CRTC Interconnection decisions concerning the desirability of competition in "competitive" service offerings.

Another aspect of the natural monopoly issue is the extent to which cost advantages accrue to multiple output production. We concentrated on the question of whether Bell Canada's costs of producing local and message toll services were reduced because of its production of competitive services (product specific economies of scope). We found no significant evidence of competitive services specific economies of scope. Hence our results would not contradict the statement that Bell Canada could provide competitive services through a separate subsidiary without a significant cost disadvantage.

The final aspect of the natural monopoly issue which we investigated concerned the question of monopoly supply of private line services. Does there exist private line specific economies of scale of sufficient significance to render monopoly provision efficient. Our empirical results suggest the answer to that question is no. There is no statistically significant evidence that the existence of CNCP Telecommunications as a competitor to Bell Canada has resulted in an increase in the costs of providing private line services.

7.3 Rate of Return Regulation and the Averch-Johnson Effect

Since investor-owned telecommunications firms are subject to rate of return regulation, it is possible that these firms will use more capital-intensive techniques than would unregulated cost-minimizing firms. This phenomenon is known as the Averch-Johnson (A.-J.) capital accumulation bias.

In order to investigate the A.-J. effect, a careful study of the firm's user cost of capital services and allowed gross return on capital services must be undertaken. In Chapter Three we built a dynamic model of the capital accumulation process for a regulated firm which yields a user cost of capital services and an allowed return on capital services. Calculation of the user cost showed that it was less than the allowed gross return for all years in the 1952-78 period. This is the relationship between cost and return which must occur if the A.-J. effect is to be present. In Chapter Five we built a model of the telecommunications production process which incorporated the rate of return regulatory constraint. In that chapter we attempted a test of the hypothesis that an A.-J. capital accumulation bias is present in Bell Canada's production process. The model performed poorly and no evidence of an A.-J. bias was found. If Bell Canada is typical of investor-owned telecommunications firms in Canada, input use inefficiency due to rate of return regulation is a minor problem.

7.4 The Telecommunications Rate Structure

In Chapter Six we analysed Bell Canada's rate structure, bringing together estimates of the demand characteristics of Bell Canada services (Chapter Two) and production characteristics (Chapter Four). Where production is characterized by constant returns to scale, economic efficiency (meaning the equating of consumer's marginal willingness to pay with the marginal costs of production) requires equating price and marginal costs. Where production is characterized by increasing rather than constant returns to scale, an efficient pricing scheme, the Ramsey rule, requires raising prices above marginal costs in inverse proportion to demand elasticities (where the services are independently demanded).

Our analyses indicated that the prices of toll and competitive services exceed marginal costs while the price of local service falls far short of its marginal cost. Thus the actual pricing structure is far from either efficient pricing scheme - marginal cost pricing or the Ransey rule. Equity considerations can dictate departures from efficient pricing schemes. However, the present structure of Bell Canada's prices is so inefficient that equity would have to imply social welfare weights which were strongly dominated by local service users in order to rationalize this structure.

We do not attempt to provide estimates of 'efficient' prices because our anaysis indicates that such pricing schemes would entail substantial departures from present prices; departures, in fact, outside the range of data over which the models were estimated. In addition, it is likely that the underlying parameters of the models (the elasticities) would change given such extreme changes in prices. Instead, we examined the following question - What is the required increase in local rates needed to maintain Bell's profit given a 20% decrease in toll rates? Our answer - a 15% increase in local rates. The effects of a 20% decrease in toll rates and a 15% increase in local rates would be a 35% increase in toll service demand but only a 3% decrease in local service demand. This reconfiguration of outputs is most likely welfare-improving because the large output increase in toll as compared to the relatively small decrease in local service demand moves the output mix closer to that suggested by the efficiency criteria.

7.5 Limitations of Empirical Analysis

At the risk of being accused of kicking a dead horse, we must reiterate the problems of attempting to answer policy questions with econometric methodology. The interesting questions require precise estimates of overall and product-specific economies of scale, economies of scope and demand elasticities. It may be possible to accumulate such evidence, but not without substantial increases in the quantity of data (further disaggregation, more firms) than is present in the public arena. The problem is not unique to telecommunications; generally, it is difficult to sort out the effects due to output changes, technical change and factor price changes using highly trended time series data. This difficulty was illustrated in Chapter Four where we demonstrated that estimates of overall economies of scale are very sensitive to functional form specification (even within the class of flexible functional forms). In addition, particular problems plague empirical estimates of the characteristics of telecommunications service demand and production - the inability to control for capacity utilization, the inability to obtain disaggregated business and residential demand elasticities, and the inability to obtain cross price elasticities of demand between competitive and monopoly toll services. Furthermore, many of the interesting policy questions require an extrapolation of cost functions to estimate "stand alone" costs. This extrapolation, needed for the tests of economies of scope and product specific economies of scale, will always be required since independent production of any particular service offering cannot be observed directly.

7.6 Policy Conclusions

Despite the limitations of empirical estimation enunciated above, we do find a number of conclusions of policy relevance:

1) There is no evidence that all telecommunications services should be supplied by Bell Canada, in particular:

   there is no evidence that competition should not be encouraged in private line services.

2) There is evidence that revisions in the rate structure would improve social welfare. Our study indicates that monopoly message toll rates are too high while local service rates are too low. Moderate changes in the rate structure would lead to large increases in toll output and only small reductions in local demand. Given equity considerations and the low price and income elasticities of the demand for local service, substantial rate structure revisions might require 'life-line' provisions.

References

Arrow, K.J., "Economic Welfare and the Allocation of Resources for Invention" in The Rate and Direction of Inventive Activity, NBER (Princeton, Princeton University Press, 1962).

Baseman, K.C., "Open Entry and Cross-Subsidy in Regulated Markets", paper presented at the National Bureau of Economic Research Conference on Public Regulation, Washington, D.C., December 15-17, 1977.

Baumol, W.J. (1971), "Rate Making: Incremental Costing and Equity Considerations", in H.M. Trebbing (ed.), Essays on Public Utility Pricing and Regulation.

Baumol, W.J., E.E. Bailey, and R.D. Willig (1977), "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly", American Economic Review, June, pp. 350-365.

Baumol, W.J., D. Fischer and M.I. Nadiri (1978), "Forms for Empirical Cost Functions to Evaluate Efficiency of Industry Structure", Paper No. 30, Centre for the Study of Business Regulation, Graduate School of Business Administration, Duke University, Durham, N.C.

Beckman, M.T. and R. Sato (1969), "Aggregate Production Functions and Types of Technical Progress: A Statistical Analysis", American Economic Review, vol. 59, no. 1, pp. 88-101.

Berndt, E.R. and D.W. Wood (1975), "Technology, Prices and the Derived Demand for Energy", The Review of Economics and Statistics 57, August, pp. 259-68.

Boadway, R.W. and N. Bruce (1979), "Depreciation and Interest Deductions and the Effect of the Corporate Income Tax on Investment", Journal of Public Economics, 11, pp. 93-105.

Breslaw, J. and J.B. Smith (1979, 1980), "Efficiency, Equity and Regulation: An Econometric Model of Bell Canada", Report to the Department of Communications, Interim Report, December 1979; Final Report, March 1980.

Carr, J.L. (1972), "A Suggestion for the Treatment of Serial Correlation: A Case in Point", Canadian Journal of Economics, May, pp. 301-306.

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 Publishing Company.

Denny, M., C. Everson, M. Fuss and L. Waverman (1979), "Estimating the Effects of Diffusion of Technological Innovations in Telecommunications: The Production Structure of Bell Canada", paper presented at the Seventh Annual Telecommunications Policy Research Conference April 29 - May 1, 1979, forthcoming in the Canadian Journal of Economics.

Denny, M., M. Fuss and L. Waverman (1979), "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications", paper presented at the National Science Foundation Conference on Productivity Measurement in Regulated Industries, Madison, Wisconsin, April 30 - May 1, 1979; forthcoming in T. Cowing and R. Stevenson (eds.) Productivity Measurement in Regulated Industries, (New York: Academic Press).

Diewert, W.E. (1974), "Applications of Duality Theory", in M.D. Intrilligator and D.A. Kendrick (eds.), Frontiers of Quantitative Economics, Vol. II, Amsterdam: North-Holland Publishing Company.

Dobell, A.R., L.D. Taylor, L. Waverman, T.H. Liu, and M.D.G. Copeland (1972), "Communications in Canada", The Bell Journal of Economics and Management Science, Spring, pp. 179-219.

Feldstein, M. (1972), "Distributional Equity and the Optimal Structure of Public Prices", American Economic Review, March, pp. 32-36.

Fuss, M. and L. Waverman (1977), "Regulation and the Multiproduct Firm: The Case of Telecommunications in Canada", paper presented at the NBER Conference on Public Regulation, Washington, September.

Fuss, M. and L. Waverman (1978), "Multi-product, Multi-input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada", Institute for Policy Analysis, Working Paper No. 7810, revision of Fuss and Waverman (1977), forthcoming under the title "Regulation and the Multiproduct Firm: The Case of Telecommunications in Canada", in G. Fromm (ed.) Studies in Public Regulation (Cambridge: M.I.T. Press, 1981).

Gordon, M.J. (1974), The Cost of Capital to a Public Utility (Michigan State University Public Utilities Studies).

Gordon, M.J. and Pradham, K. (1975), "A Study of the Cost of Capital for a Telecommunications Utility", mimeo, University of Toronto.

Littlechild, S.C. and J.J. Rousseau (1975), "The Pricing Policy of a U.S. Telephone Company", The Journal of Public Economics.

McFadden, D. (1978), "Cost, Revenue and Profit Functions", Chapter I.1 in M. Fuss and D. McFadden (eds.), Production Economics: A Dual Approach to Theory and Applications, Amsterdam: North-Holland Publishing Company.

Nadiri, M.I. and M. Shankerman (1979), "The Structure of Production, Technological Change and the Rate of Growth of Total Factor Productivity in the Bell System", paper presented at the N.S.F. Conference on Productivity in Regulated Industries, University of Wisconsin, April 30 - May 1, 1979.

Panzar, J.C. and R.D. Willig (1977), "Free Entry and the Sustainability of Natural Monopoly", The Bell Journal of Economics, vol. 8, no. 1, pp. 1-22.

Panzer, J.C. and R.D. Willig (1979), "Economies of Scope, Product Specific Economies of Scale, and the Multiproduct Competitive Firm", Bell Laboratories Economics Discussion Paper #152, August.

Peterson, H.C. (1975), "An Empirical Test of Regulatory Effects", The Bell Journal of Economics, Spring, pp. 111-126.

Schumpeter, J. (1962), Capitalism, Socialism and Democracy, (New York: Harper & Row).

Smith, J.B. and V. Corbo (1979), "Economies of Scale and Economies of Scope in Bell Canada", Final Report to the Department of Communications, March, Ottawa.

Spann, R. (1974), "Rate of Return Regulation and Efficiency in Production: An Empirical Test of the Averch-Johnson Thesis", The Bell Journal of Economics and Management Science, Spring, pp. 38-52.

Ullah, A. (1978), "Ridge Regression Revisited", mimeo, University of Western Ontario.

Westfield, F.M. (1971), "Innovation and Monopoly Regulation", in W.M. Capron (ed.) Technological Change in Regulated Industries (Washington: The Brookings Institute).

Mar 9/92