144

COMPARATIVE EFFICIENCY IN CANADIAN TELECOMMUNICATIONS: AN ANALYSIS OF METHODS AND USES

Phase II: Productivity, Employment and Technical Change in Canadian Telecommunications

Michael Denny
Institute for Policy Analysis
University of Toronto

Alain de Fontenay
Department of Communications
Ottawa

Manuel Werner, Consultant

Draft Final Report for Department of Communications
(03SU. 36100-9-9527-DSS)

COMPARATIVE EFFICIENCY IN CANADIAN TELECOMMUNICATIONS: AN ANALYSIS OF METHODS AND USES

Phase II: Productivity, Employment and Technical Change in Canadian Telecommunications

Michael Denny
Institute for Policy Analysis
University of Toronto

Alain de Fontenay
Department of Communications
Ottawa

Manuel Werner, Consultant

Draft Final Report for Department of Communications
(03SU. 36100-9-9527-DSS)

COMPARATIVE EFFICIENCY IN CANADIAN TELECOMMUNICATIONS: AN ANALYSIS OF METHODS AND USES

PART I: Methods and Data

Phase II: Productivity, Employment and Technical Change in Canadian Telecommunications

Michael Denny
Institute for Policy Analysis
University of Toronto

Alain de Fontenay
Department of Communications
Ottawa

Manuel Werner, Consultant

Draft Final Report for Department of Communications
(03SU, 36100-9-9527-DSS)

UNEDITED

NOT FOR CITATION

Table of Contents

* These sections should be read by the non-specialist

* These sections should be read by the non-specialist

I. An Overview

Efficiency in production is a goal that is desirable for a nation, a firm and a regulatory agency. Productivity is a way of measuring efficiency. It is used to compare changes in efficiency through time and across firms. Without high levels and/or growing productivity, the real incomes of individuals in a nation and the wealth of shareholders in a firm will not be high and/or rising. Efficiency or productivity is often misinterpreted although its measurement and interpretation presents no difficulties which do not arise in measuring costs, revenues or profits.

Profits are what firms seek in order to pay dividends to shareholders and to increase the market value of their shares. Since revenues and costs are simply the components underlying the calculation of profits, the latter concept will suffice to indicate what we mean. High and growing levels of profits relative to the capital invested in the firm are goals for firms. Measured profits provide a reading on the success of the firm. Yet only a casual acquaintance with accounting coventions is required to realize the ad hoc and potentially misleading methodology underlying the standard measurement of profits. Moreover, any executive knows that the level of profits in any year or the change between any two years is the result of many planned and unplanned events. The precise contribution to the level (or change) in profits of any event is often not known or perhaps can not be estimated accurately. The difficulties of measuring profits and relating specific events to profits has not resulted in profits not being measured or used to evaluate performance. Instead managers, analysts and the general public use measured profits with discretion and attempt to supplement this single item with other information about performance. However, ultimately if the levels of profits does not indicate success then it is unlikely that other indicators will reverse this conclusion. The same type of interpretation is required for measured productivity.

The problems in measuring productivity are not more severe than those encountered in measuring profits. Comparing the profitability of firms is no easier than comparing their efficiency. Since efficiency is a component of profitability, the measurement of efficiency can assist in the interpretation of profits. However, any measure such as profits or productivity is an indicator. The reasons why either one is high, low, growing or falling is not part of that measure itself. Causations or explanations must be sought outside of the measuring rod.

Productivity and profits are closely linked. Profits depend on transforming resources, using capital and labour, into a finished product that can be sold. The required resources must be purchased with the least expense possible and the outputs sold for the largest possible revenues, However, the transformation or processing of the resources into finished products must be done efficiently if profits are to be large. All of the profits available at the existing market prices for inputs and outputs may be frittered away through inefficient production. It is often suggested that Japan was (and is) a low cost producer of manufactured goods. However there are many countries where the prices of inputs are equally low but few where production is as efficient.

Any organization that is interested in profits must be interested in productivity. However productivity is more basic than profits since the gains from productivity increases may be distributed in the form of lower prices to customers, higher incomes to workers and higher prices for other supplying industries in addition to their contribution to profits.

This report presents methods for measuring productivity and for comparing the productivity of firms. The application of these methods will yield indicators of the performance of firms relative to one another and relative to their own historical performance. As stated above the resulting measures do not tell why the observed results were obtained.

To measure productivity, the physical volumes of outputs produced and of inputs consumed must be measured. A substantial portion of this report discusses desired methods of measuring outputs and inputs. As in cost accounting, there are alternative possibilities which depend on the purpose, the feasibility and the cost involved in measurement. Recommendations are made concerning the preferred methods although alternatives are evaluated as possibilities under a variety of circumstances. The current practices in some of the telecommunications firms who are measuring productivity is critically discussed. This evaluation of practical measurement is extended to other regulated industries in transportation and public utilities.

Economists have always attempted to delve behind the cost and profit data collected by firms in an attempt to understand the choices about input uses and outputs produced by firms. As economists we have extended the statistical analysis of firms behavior to link it with productivity measurement. That is economic analysis can be used to understand some of the reasons why productivity is high or low. Our report may imply by omission that only economic analysis is relevant for understanding productivity measures. This is not correct. The economic analysis of costs, revenues and profits is believed by us to be useful. However, management, investment analysts and others use a wide variety of formal and informal methods to analyze profits. The same possibilities exist for attempting to understand productivity. Since our expertise is in economics, we have tried to investigate methods to analyze measured productivity using economic analysis. This does not preclude other methods with which we are less skilled. In fact we would encourage their use by others while defending the importance of economic analysis.

The telecommunication companies are familiar with analyzing levels and changes in costs and profits. Their experience with productivity is less extensive. To remedy this, we have devoted a portion of the report to explaining the possible management uses of productivity. In particular, we have discussed the uses made by AT&T and Teleglobe in the telecommunications industry. These uses include the integrated portrayal of profits and productivity and a method for measuring how the productivity gains arising from management efforts were divided between consumers and profits. Less extensive descriptions of other uses in planning and control are included. To the extent that companies can establish useful cost (or profit) centres they may also utilize productivity centers to further the drive for profits and efficiency.

The body of the report is predominantly written for the specialist. This is particularly true for the sections dealing with statistical analysis and the economic interpretation of productivity. The sections on uses should be accessible to a wider audience. In an attempt to provide some guidance, the table of contents is coded to indicate which sections are most useful for the non-specialist.

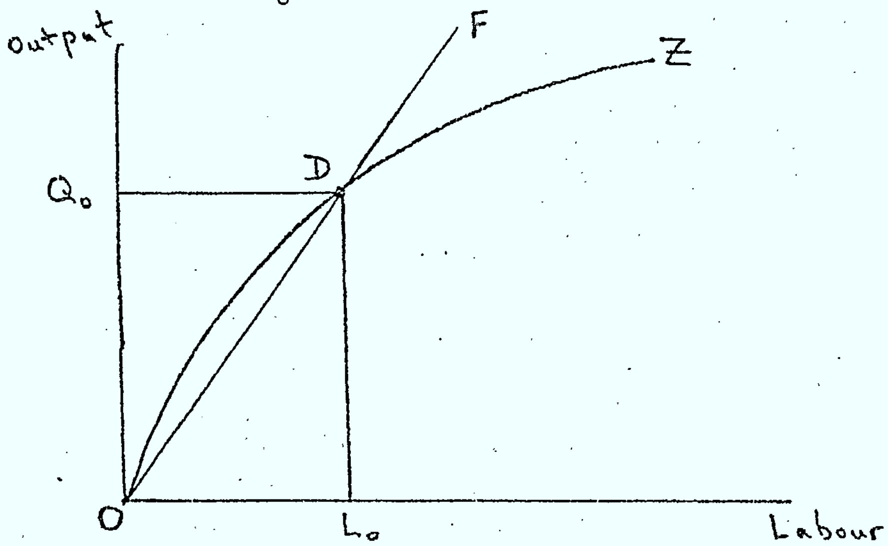
The regulator has an interest in productivity which does not necessarily conflict with the firm's profit goals. Whatever other goals the regulator may have, the efficiency of the firms under his regulation. must be a goal. Finding concrete ways in which regulation can encourage, enforce and reward efficient production should be part of the regulators task. Existing regulation may well encourage inefficiency of several types. The report attempts to assess the inefficiencies that might be generated by current regulation and to suggest methods for overcoming these problems. There is probably a tradeoff between efficiency and other regulatory goals. However regulators must evaluate the inefficiencies inherent in some policies they have currently encouraged. Otherwise the costs to society of pursuing certain goals will be unrecognized and improperly evaluated. The practical implementation of productivity regulation is unfortunately not studied here.

This report provides the methods for implementing productivity measurement and its economic interpretation and uses within telecommunications. Recommendations on best practices and alternatives are included although not always explicitly summarized. References to further work is not made in the report. The implications for further work have been detailed in a separate summary prepared for the Program Manager of the joint DOC-CTCA project.

II. The Conceptual Basis for Measuring and Comparing Firms' Productivity

II.1 Introduction

This report is concerned with the measurement of productivity or efficiency. To illustrate what we are trying to measure a very simple example may help. In the figure below we have drawn a very simple production function represented by the curve OZ in the figure. For any amount of labour, the maximum amount of output that can be produced is given by the vertical coordinate of the point on OZ. Our example involves a very simple production process in which labour produces output. If production is efficient, then the producer is always obtaining the output quantity on the production function, OZ. That is, if the quantity L_0 of labour is used then the efficient quantity of output Q_0 is being produced.



Increases in productivity or efficiency imply that the production function, OZ, shifts up. More output is obtained per unit of input at (perhaps) all input levels. Productivity is often measured as the ratio of aggregate output to aggregate inputs. In our simple example this is equivalent to labour productivity, output per unit of labour. In the figure, productivity at output level Q_0 equals OQ_0/OL_0 , i.e., the slope of the line ODF .

If we measure productivity by the output per unit of input, two problems may arise. First notice that productivity falls as the quantity of labour rises with no shifting in the production function, OZ . This occurs because the curve ODZ is shaped like an upside down bowl, slightly tipped. The production function in our example exhibits decreasing returns to scale and consequently productivity falls as the level of output grows.

If the level of output and input, labour, is observed at two different time periods or for two firms at the same time period, we would like to be able to distinguish changes in productivity (output per unit of labour) that arise due to shifts upward in the production function, OZ , and movements along the curve. It will not always be easy to do this. One of our tasks is to understand the practical possibilities of making this distinction.

The shifts in the production function may arise for any number of reasons. It is not possible to know why a change occured from the measurement of productivity alone. The observation that productivity increases is encouraging. It is a complex task to sort out the reasons and we certainly do not do this. Over time the primary reasons for increased efficiency is the growth in our technical capabilities which are incorporated in the capital and used by skilled labour. However, changes in work organization, management structure or personnel policy may alter productivity. In a complex organization like a modern corporation measuring productivity in various disaggregated component activities of the firm will make it possible to identify specific reasons for productivity change in these activities. For the firm as a whole, the changes in efficiency in all portions of the operations result in the observed change in the firms productivity. Analyzing the reasons <u>why</u> productivity changes, for a firm, can be elusive and this is why it is similar to analyzing profits. Almost every facet of the firms operation effects both. To measure in detail the contribution of the firms components to either is difficult at best. Our analysis is concentrated at the level of the firm but we hope that the firms will pursue the more detailed internal analysis. We are willing to assist if that is useful but that effort is beyond the scope of this project.

II.2 Index Numbers and Aggregation

If firms produced a single output using a single input this section would not have to be written. In a concrete situation in which there are multiple inputs and multiple outputs, the problems of aggregation arise. Productivity has been defined as the ratio of aggregate output to aggregate input and we have to choose an aggregation formula. The theory of aggregation is replete with the problems of defining reasonable aggregate variables and we will mention some of these as we proceed. Index number theory may be considered to be a sub-field within aggregation theory or a separate but overlapping field of study. In section V, a lengthy and technical discussion of index number theory is presented. In a few pages here, we will attempt to summarize some of that material as it applies to the measurement of productivity. The specialist should consult section V.

Suppose that ten services are produced at a particular set of prices. We can form an aggregate output quantity by multiplying the prices and quantities of each service together and adding up the results. If we had the data on the quantities produced in several firms or time periods we can aggregate each data set using a common set of prices. This type of measure is a constant dollar quantity. Each output is valued at a fixed set of prices in all time periods or firms. The value of aggregate output depends on the set of fixed prices used in the aggregation. That is, this is a base weighted constant dollar measure of aggregate output. To form a quantity index, it is normal to divide the constant dollar quantity through by its value in the base year. This is a simple example of a Laspeyres index of aggregate output.

There are many alternative index number formulas that will convert a set of disaggregated observations into an aggregate index. There have been two streams of thought concerning which index number should be used. In both streams, the choice may depend on the purpose for which the index is to be used. An older and currently less important literature attempted to evaluate the algebraic properties of various index number formulae and to evaluate alternative index numbers according to the extent to which they possessed a number of algebraic properties. The desired properties were chosen arbitrarily due to their reasonableness rather than any economic reasoning.

The alternative stream, which certainly overlaps with the first and existed many years ago, has received increasing prominence. This approach might be called the economic theory of index numbers. In selecting an index number for a particular purpose one wishes to know what are the economic implications of the choice, not the algebraic properties. For example, in analyzing productivity the choice of an index number for aggregating inputs implies a choice of an underlying technology including the ease or difficulty with which inputs can be substituted for one another. Recent developments in this area have suggested that one use an index number which implies a relatively flexible underlying technology. By doing so one does not assume a restrictive form for the technology and consequently will not make erroroneous assumptions about productivity.

We have basically argued for the following conclusions concerning the choice of index numbers based on economic theory. First, fixed weights indexes should not be used. Of the variable weighted indexes, one should choose one that implies a non-restrictive underlying technology. Finally the Divisia index has many convenient features for making inter-firm comparisons and for linking, the index number approach to productivity measurement, with the econometric approach. Since the Divisia index (or a discrete approximation of it) satisfies the first two properties, this index number is preferred for our purposes.

There are at least three qualifications and one additional positive factor. First, in many instances the choice of one index number formula from amongst the variable weighted category, leads to empirical results which do not differ sharply from the results with other index number formula. However, one can not know whether this is true without doing the calculations. Second, it is cheaper and may be feasible to use fixed weight indexes in certain situations in which the data for the preferred indexes can not reasonably be developed. Third, there are certainly other indexes which will satisfy the first two criteria above. What is not clear is that some of our developments for making inter-firm comparisons can be implemented in their current form without using the Divisia index. There is certainly no requirement that a single type of index number be chosen before some empirical experiments are completed.

II.3 The Conventional Divisia Index of Total Factor Productivity

From a conceptual point of view, one of the most defensible methods of aggregation for use in productivity analysis is Divisia aggregation. This fact has become well established through the research of Jorgenson and Griliches (1967), Richter (1966), Hulten (1973), and Diewert (1976), among others.

The conventionally measured Divisia index of total factor productivity is obtained in the following way. First we define total factor productivity (TFP) as the ratio of aggregate output (Q) to aggregate input (F). Aggregate output (input) is an index of disaggregated outputs (inputs). The Divisia indices for aggregate output (Q) and input (F) are defined in terms of proportionate rates of growth ($\dot{Q}$ and $\dot{F}$) as

$$\dot{Q} = \sum_j \frac{P_j Q_j}{R} \cdot \dot{Q}_j \tag{1.1}$$

where

P_j = price of output j

Q_j = quantity of output j

$\dot{Q}_j$ = proportion rate of growth of output j

$R = \sum_j P_j Q_j$ = total revenue

and

$$\dot{F} = \sum_i \frac{w_i X_i}{C} \cdot \dot{X}_i \tag{1.2}$$

where

w_i = price of input i

X_i = quantity of input i

$\dot{X}_i$ = proportionate rate of growth of input i

$C = \Sigma_i w_i X_i$ = total cost .

Since TFP = Q/F , the proportionate rate of growth of total factor productivity $(\dot{TFP})$ is defined by

$$\dot{TFP} = \dot{Q} - \dot{F} \qquad (1.3)$$

The formulas (1.1 - 1.3) are in terms of instantaneous changes. For data obtainable at yearly intervals, the most commonly used discrete approximation to the continuous formulae (1.1) and (1.2) is given by the Tornqvist approximations:

$$\Delta \log Q = \log (Q_t/Q_{t-1}) = \tfrac{1}{2} \sum_j (r_{jt} + r_{j,t-1}) \log (Q_{jt}/Q_{j,t-1}) \qquad (1.4)$$

where

Q_{jt} = quantity of output Q_j produced in period t

$r_{jt} = \dfrac{P_{jt}Q_{jt}}{\sum_j P_{jt}Q_{jt}}$ = revenue share of output Q_j in total revenue during period t

and

$$\Delta \log F = \log (F_t/F_{t-1}) = \tfrac{1}{2} \sum_i (s_{it} + s_{i,t-1}) \log (X_{it}/X_{i,t-1}) \qquad (1.5)$$

where

X_{it} = quantity of input X_i used in period t

$s_{it} = (w_i X_i)/\sum_i w_i X_i$, the cost share of input X_i in the total cost during period t.

Finally, the corresponding discrete approximation to (1.3) is provided by

$$\Delta TFP = \Delta \log Q - \Delta \log F \qquad (1.6)$$

Choosing the index to equal 100.0 in a particular year, and accumulating the measure in accordance with (1.6) provides estimates of what we call the conventional index of total factor productivity.

II.4 Total Factor Productivity and the Theory of Production

A brief summary will be presented of the major links between the measurement of total factor productivity and the theory of production. More detailed treatment will be found in Denny, Fuss and Waverman (1979).

The firms technology is described by the production function

$$Q = f(x_1, x_2, \ldots x_n, t) \tag{1.7}$$

For any variable Z, define $\dot{Z}$ as the proportional rate of change of Z with respect to time. Totally differentiating the production function with respect to time and assuming cost minimization, we obtain

$$\dot{Q} = \sum_i \varepsilon_{CQ}^{-1} s_i \dot{X}_i + \dot{A} \quad , \quad i = 1, \ldots, n \; , \tag{1.8}$$

where

$\varepsilon_{CQ} = \frac{\partial C}{\partial Q} \cdot \frac{Q}{C}$, the elasticity of total cost with respect to output

s_i = cost share of input i

$\dot{A} = \frac{\partial f}{\partial t} \cdot \frac{1}{f}$, the rate of technical change.

Using the definition of aggregate input F, given above, we may write

$$\dot{A} = \dot{Q} - \varepsilon_{CQ}^{-1} \dot{F}$$

or

$$\dot{TFP} = \dot{A} + (\varepsilon_{CQ}^{-1} - 1)\dot{F} \tag{1.9}$$

With constant returns to scale $(\varepsilon_{CQ} = 1)$, the conventional Divisia measure of productivity growth will be equal to the rate of technical change. With increasing returns to scale, the conventional measure will overestimate technical change. The overestimate will be larger, for any given level of scale economies, the faster that aggregate inputs are growing.

Total factor productivity can be divided into two components using equation (1.9). As inputs grow, the presence of non-constant returns to scale leads to productivity changes whose magnitude depends on the exact nature of the scale factor. Shifts in the production function $(\dot{A})$ contribute the other component of productivity.

The same type of analysis can be carried out using the cost function. Suppose we represent the cost function by the equation

$$C = g(w_1, w_2, \ldots w_n, Q, t) \tag{1.10}$$

Applying the same procedures we find that the proportionate shifting of the cost function, $\dot{B}$, may be written

$$-\dot{B} = \varepsilon_{CQ}\,\dot{Q} - \dot{F}$$

and

$$-\dot{B} = \varepsilon_{CQ}\,\dot{A} \tag{1.11}$$

This can be directly related to the conventional productivity measures by writing

$$\dot{TFP} = -\dot{B} + (1 - \varepsilon_{CQ})\dot{Q} \qquad (1.12)$$

With constant returns to scale, conventional measures of the rate of growth of total factor productivity provide estimates of the rate of technical change measured from either the cost function or the production function.

It is important to recognize that productivity may be both measured and thought about in relation to costs as well as production. The shifting of the production function $(\dot{A})$ will not be identical to shifting of the cost function $(\dot{B})$ but they may be related to each other (1.11) and to the rate of growth of total factor productivity.

Extending the analysis to the multiple output case, we find that

$$\dot{TFP} = -\dot{B} + (1 - \Theta)\dot{Q}_c + (\dot{Q}_p - \dot{Q}_c)$$

where

$$\Theta = \sum_j \varepsilon_{CQ_j} \text{ , the sum of the cost elasticities for all outputs.}$$

The two measures of the growth in aggregate output $\dot{Q}_p$ and $\dot{Q}_c$ differ in the weights used to aggregate the component outputs. The conventional aggregate output, $\dot{Q}_p$, uses revenue shares as weights and is defined in equation (1.1). The alternative, $\dot{Q}_c$, uses weights that are the share of the cost elasticity for output j in the sum of the cost elasticities. If there are constant returns to scale $(\Theta = 1)$ then the second term drops out. The third term represents departures from marginal cost pricing.

If prices are equal to either marginal costs or a uniform proportion of marginal costs for all outputs then the third term is zero. Otherwise, the third term shows the contribution of non-marginal cost pricing to conventional measures of productivity growth.

The procedures outlined above provide an interpretation of conventional measures of productivity. Since data on cost elasticities and differences between prices and marginal costs will not always be available, calculations of conventional efficiency differences over time or space must be viewed as a compendium of particular effects. Resolution of the efficiency differences into its components is only possible when sufficient data is available to estimate cost elasticities. When that data is not available the efficiency differentials must not be narrowly interpreted as a reflection of differences in production technique.

II.5 Alternative Specifications of Productivity

There are an unlimited potential number of definitions of productivity. Most of these are certainly less useful than total factor productivity although in particular applications they may be informative and perhaps the best attainable. Our preferred specification has been given in the previous section and in this section we will outline the alternatives.

Each alternative specification of a productivity measurement system implies at least three choices:

1) A functional form; whereby we exploit the notion of exact and superlative indexes (Diewert, 1976) in order to choose that index number which most closely approximates the hypothesized underlying functional form. For example, if we choose the translog as our functional form, then a Divisia index, in its discrete form, as outlined by Tornqvist, is the best approximation (Diewert 1979).

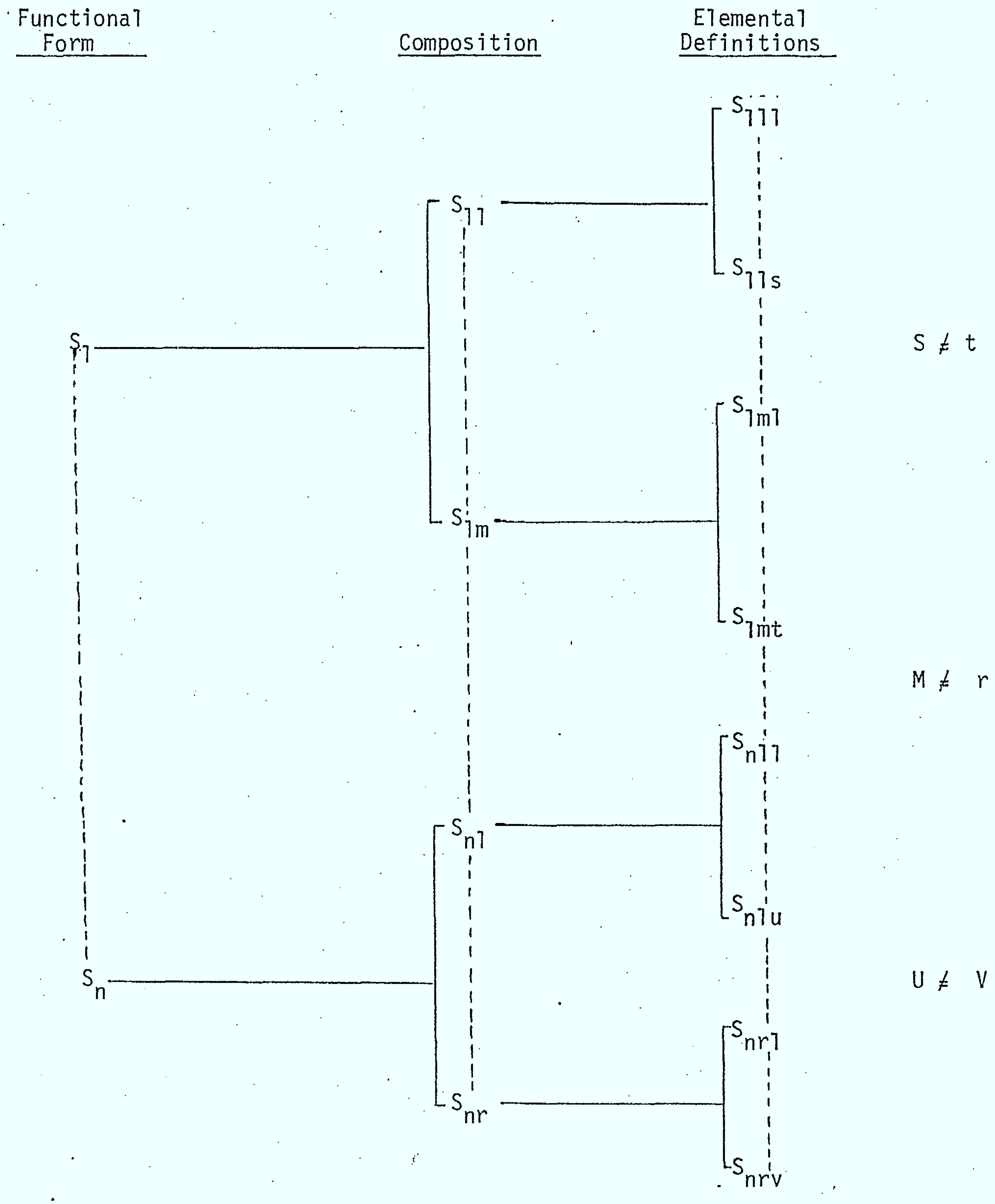
2) Composition; whereby we choose the exact composition of aggregate inputs and outputs in terms of their "elemental components". This family of alternatives comprises a set of specifications such as the various forms of value added, total outputs, partial inputs, etc. Thus, we have a specification of productivity as,

$$\text{Productivity Index} = \frac{F(O_1 \ldots\ldots O_n,\ I_1 \ldots\ldots\ldots I_m)}{G(I_1 \ldots\ldots\ldots\ldots\ldots I_m)}$$

where $I_j \geq 0$ with at least one $I_j \geq 0$.

3) <u>Elemental Definitions</u>; whereby the components, or elements of F and G above are defined at the appropriate points in the input-output continuum. This consists of ensuring that the final product as perceived by the consumer is in fact identical to the supply of finished goods as perceived by the producer. Thus, a steel mill that produced 10% more X-forms would be considered more productive in only a very narrow sense if the consuming public required more Y-forms and refused to accept anything else.

These choices can be represented schematically as:



Since the issues concerning choices (1) and (3) are given very thorough coverage elsewhere in this report, we will restrict the following discussion to the question of composition. There are two broad classes, total and partial productivity measurement. Within the former we can distinguish between two important sub-classes, each in turn being sub-divided into two types. Thus, we are looking at four specifications pertaining to a total type measure. Partial measures can also be categorized into two broad classes, each of which comprises a of very large number of sub-possibilities. We may begin by tabulating all the relevant total and partial possibilities.

1) TOTAL MEASURES:

1.1) Value Added:

1.1.1) Net Value Added:

$$= \frac{\text{Gross Output - Materials - Depreciation}}{\text{Net Capital Input + Labour Input}}$$

1.1.2) Gross Value Added:

$$= \frac{\text{Gross Output - Materials}}{\text{Gross Capital Input + Labour Input}}$$

1.2) Total Output:

1.2.1) Net Total Output:

$$= \frac{\text{Gross Output - Depreciation}}{\text{Net Capital Input + Labour Input + Materials Input}}$$

1.2.2) Gross Total Output:

$$= \frac{\text{Gross Output}}{\text{Gross Capital Input + Labour Input + Materials Input}}$$

2) PARTIAL MEASURES:

2.1) Total-Partial:

2.1.1) Value Added Total-Partial:

2.1.1.1) Net Value Added Total-Partial:

$$= \frac{\text{Gross Output - Depreciation - Materials}}{\text{the } i^{th} \text{ input}}$$

i = Net Capital or Labour

2.1.1.2) Gross Value Added Total-Partial:

$$= \frac{\text{Gross Output - Materials}}{\text{the } i^{th} \text{ input}}$$

i = Gross Capital or Labour

2.1.2) Total Total-Partial:

2.1.2.1) Net Total Total-Partial:

$$= \frac{\text{Gross Output - Depreciation}}{\text{the } i^{th} \text{ input}}$$

i = Net Capital or Labour

2.1.2.2) Gross Total Total-Partial:

$$= \frac{\text{Gross Output}}{\text{the } i^{th} \text{ input}}$$

i = Gross Capital or Labour

2.2) Partial-Partial:

$$= \frac{\text{The } i^{th} \text{ Output}}{\text{The } j^{th} \text{ Input}} \qquad i \neq j$$

Although we have listed a number of alternative specifications, we will concentrate on those comparisons which imply important differences in results. These include the bilateral distinctions:

1) Value Added vs Total
2) Total vs Total-Partial
3) Total-Partial vs Partial-Partial

There are two important considerations when choosing between Value Added and Total. First of all, without any separability requirements, it can be shown that the rate of growth of the productivity index for total output will always be less than that for real value added (Denny & May 1977). From

$$O_t = TFP_{it} F(K_t, L_t, M_t)$$

where TFP_{it}; $i = g, v$ is an index of total or value added productivity, the proportional rate of change of the TFP index for total output equals

$$\dot{TFP}_{gt} = \dot{O}_t - \Sigma_i S_{it}\dot{X}_{it} \; ; \quad \text{where } X_i = K, L, M \text{ for } i = K, L, M$$

$$\text{and} \quad S_{it} = \frac{r_{it}X_{it}}{P_t O_t} \; ; \; \Sigma S_{it} = 1$$

$$\text{where } r_i = \text{the price of the i-th input}$$

and for a real value added model of:

$$O_t = F\left[TFP_{vt}\, G(K_t, L_t), M_t\right]$$

the following relationship holds

$$T\dot{F}P_{vt} = \frac{(\dot{O}_t - \Sigma_i S_{it}X_{it})}{S_{vt}}$$; where S_{vt} = share of value added in the total value of output

Then, $T\dot{F}P_{gt} = S_{vt}\, T\dot{F}P_{vt}$

which demonstrates the differential rates of productivity growth. A simple example, (Vincent), can be used to illustrate this point. Assuming labour and materials as the only inputs we have, in real terms,

TYPE \ YEAR	0	1
Output	100	140
Labour	85	75
Materials	15	25

Then:

Total Measure	Value Added
$\left(\dfrac{140}{75+25}\right) \Big/ \left(\dfrac{100}{85+15}\right)$	$\left(\dfrac{140-25}{75}\right) \Big/ \left(\dfrac{100-15}{85}\right)$
$= 1.40$	$= 1.53$

and $TFP_{gt} < TFP_{vt}$

The second consideration includes the question of separability (Denny & May). The standard procedure for measuring real value added is called double deflation. Outputs and materials are deflated separately. If this were not done, then it would be impossible to apply revenue share weights (in the absence of cost elasticity information) with the linear homogenity property whereby $\Sigma S_{it} = 1$. The difference between these independantly deflated series measures is real value added. In general terms, unless the production technology is additively separable,

$$O = f(K, L) + g(M)$$

then real value added measurement will result in errors. This possibility of error can be illustrated as follows: suppose there was no technical change and no change in productivity during a given time period, however, the use of materials grew faster than output, then the real value added measurement will record a non-existent decline in productivity. That is, if the separability hypothesis is rejected then ascribing factor specific technological changing to materials, as does real value added measurement, leads to errors.

Another consideration in the value-added vs. total debate is that of disequilibrium (Treadway), unobserved in the basic accounting identities of the firm. If disequilibrium is allowed, it is virtually impossible to say what kinds of distortions value-added measures imply. The notion that internal resources contribute in an "essential" fashion while purchases do so only "inessentially" is based upon considerations that have no relationship to the relevant technical or organizational structure. The use of (O-M) as the definition of real output implies a loss of information("O-M is consistent with an infinite number of O, M pairs"). "We could never do worse by treating O as output and M as an input and could often do better."

The <u>total vs. total-partial</u> issue is more straightforward. First of all every total measure is a weighted average of total-partials (assuming that all respective outputs and inputs are identically defined and composed, i.e., when the total measure is net value-added then the total partial measure must have output defined to exclude depreciation and materials, and capital, when it is the relevant input, must also be net of depreciation). That is, it can be shown that, if product is exhausted, ensuring O/I = R/P then,

$$TFP = \sum_{i=1}^{3} W_i \, O/X_i \qquad X_i = K, L, M \text{ as } i = 1.......3$$

where the $W_i = (P_i X_i)^2 / P_o O \Sigma P_i X_i$

In general, "partial-partial" measures, unless the j^{th} input is in fact the only significant factor ineracting with the i^{th} output (where this is not necessarily a finished good), suffer the same drawback as the total-partail measure which offers indications of change which can be misleading. For example, if the number of operator handled calls per circuit increased, do we attribute the growth to better operators, better management of operator time, better circuits, etc...?

A final alternative measure offers some interesting possibilties for detailed application by management of firms. The latter have become familiar with the usefulness of defining cost centers or profit centers. One can define, with no greater difficulty, productivity centers. These can be at the level of broad functions, e.g., transmission or switching, or at a more detailed level of a particular working group. Companies often used informal productivity indicators in many segments of their activities, e.g. calls handled per operator hour. It is possible to develop an information system which uses a variety of partial productivity indicators and links these to the overall productivity performance of the firm. A detailed model is not developed here but it can be done and some efforts in this direction will be forthcoming during the next phase. The problems are of the same nature as those that arise in the creation of profit or cost centers. Information is already collected on many detailed activities and the major task is the coordination of the disaggregated measures into a useful information system for management.

II.6 Inter-Firm Comparisons: Some Methodological Issues

In earlier sections, we have discussed the task of evaluating the efficiency of a given firm. In this section, the problems that confound the comparison of firms will be analyzed. Recall that most measures of productivity at the firm or industry level are index numbers with a base year equal to one. If information about productivity indexes is available for several firms, we can certainly compare the rates of productivity growth for the firms. However, if we are to compare the levels of efficiency in addition to the rates of efficiency change then the level comparison requires something beyond the information available in the productivity indexes.

This problem is very old. It has been discussed in the context of comparing the welfare of individuals for decades. Suppose that there are two individuals, Smith and Jones, and two vectors of commodities, $\bar{X}_1$ and $\bar{X}_2$. Which commodity bundle is preferred is not independent of whose utility function one uses to evaluate the bundles. If Smith and Jones have the 'same' utility function then the question of who is better off may easily be answered. One bundle will be preferred to the other by both individuals. The individual with the preferred bundle is better off. In all other cases the comparison of inter-personal welfare may founder on differences in preferences. Bundle $\bar{X}_1$ may be preferred to $\bar{X}_2$ by Smith and the reverse may be true for Jones. Some additional structures must be added to compare welfare in these contexts.

Similar problems will occur in the comparison of the production or cost efficiency between firms. Consider a vector of inputs, X_0 which will produce an output level Q_1 in one company and $Q_2 < Q_1$ in another firm. It is tempting to argue that firm one is more efficient than firm two and that the relative efficiency may be measured by the relative output levels. It is necessary to explore the context in which this is a sensible conclusion as well as to develop concrete methods for measuring inter-firm comparisons.

There were two special features in our example. Only one output was produced and only one particular input bundle was considered. If two or more outputs are produced then the comparison is more complex. If all the components of the output vector for one firm are larger than the other the comparison is straightforward. However if this is not true it is necessary to decide on how to define aggregate output so that one can determine which firm is more efficient given that each is using a given input vector. Even with a single output, once the input vector is altered, $X \neq X_0$, the relative efficiency rankings of the firm may change. If at different input vectors, the rankings change what are we to conclude? From the economic theory of production, we might conclude that the production function for the two firms was different. Since the comparisons are being made at identical input vectors, any reversals of the efficiency rankings suggest (a) that the isoquants for one firm lie inside the comparable isoquants over some output ranges and outside for other ranges and/or (b) the isoquants intersect. The second case would arise if we restricted the observations to any series of input vectors that produce the same output in one firm. If the second firm produces more output at some of the input vectors and less at others then the isoquants must cross. When this occurs, the technology is different in a more fundamental sense than in case (a). In both cases we need to clarify the interpretation of the comparative efficiency of the two firms.

Most practical situations are more complex. Firms produce different output vectors using different input vectors and the prices for outputs and

and inputs are not the same for each firm. In order to compare the firms in this context, a method for standardizing the outputs and inputs or prices must be chosen. This is certainly the case with which we must contend and it will be considered below.

To illustrate a simple methodology, consider taking a particular vector of output prices $\bar{p}_o$ and input prices $\bar{w}_o$. Using the actual observations on outputs and inputs, calculate for each firm k the ratio

$$PR_k = \sum_i p_{io} \cdot Q_{ik} / \sum_j w_{jo} X_{jk} \qquad k = 1, \ldots$$

The firms can be ranked by the value of PR_k . The ranking will depend on the particular set of output, p_{io} , and input w_{io} , prices chosen and on the observed quantity vectors for ourputs and inputs. A firm that does well in a comparison under one set of prices and quantities may do badly using another set. How might we choose a set of prices and quantities at which to make a comparison. This problem and extensions of it will concern us throughout this section.

The firms may be compared under a number of alternative assumptions about their production or cost functions. Crucial to any analytic foundations for a comparative efficiency measure are the explicit or implicit assumptions about the differences in technology. For example, suppose we assume that each firm has a cost function

$$C_k = g_k(w,Q)$$

where $w = (w_1,\ldots w_n)$ and $Q = (Q_1,\ldots Q_s)$ are vectors of input prices and output quantities. For any given input price and output quantity vector $(w_o Q_o)$, one could rank the firms according to their total cost. The problems discussed earlier in regard to the ranking changing with w and Q

hold in this case also. For this type of comparison we need to know the cost function for each firm j. Econometric estimates of the cost function could be made but this would require substantial data for each firm. If the cost functions were available then we could compare firms at any w and Q. The rankings may change for different values of w and Q but this is a true aspect of the comparison. Some firms are relatively more efficient at certain input-output combinations than at others. There is no reason a priori to expect the ranking to be independent of w and Q. It would have been entirely equivalent to begin with knowledge of the production functions for each firm. If these were known then rankings can be made at any input-output quantities.

We generally observe, the prices and quantities of inputs and outputs used by the firm. Information is not generally available on the cost or production function and sufficient data may not be available for estimation of these functions. What we are seeking is a method for comparing firms without requiring that we know the cost or production function in detail. That is, we would like a method for comparing the efficiency of firms using only limited price and quantity data without estimation of the cost and production function. The simple arbitrary formula for PR_k, introduced earlier, is an example. Observed input and output quantities are weighted with identical input and output prices for each firm. How do we select the price vector? The traditional possibilities in a two firm comparison have been the actual price vectors faced by the two firms or some sort of average of these price vectors. If more than one price vector is used we can have different rankings. If the number of firms is larger than two then the problems of choosing a single price vector are expanded. An average price vector for all firms can be chosen but this is nothing but an arbitrary solution.

In a recent article, Jorgenson and Nishimizu (1978) compared the relative efficiency of the Japanese and U.S. economy during the last twenty-five years. The method they used is interesting and will provide a starting point for some new developments suggested here.

The basis for their development is the assumption that the firms have relatively similar production functions. Instead of permitting the function for the production or cost function to vary across firms, these functions are assumed to be identical across firms but these are firm specific arguments of the common function. This permits the output level to be different for firms using an identical input vector but it restricts the differences in the technology across firms. Firms' production or cost functions cannot be completely different although the precise nature of the restriction depends on the functional form. It will be illustrated below for the Jorgenson and Nishimizu case.

Assume that each of two countries or firms produces one output using capital and labour. The production technology may be written

$$Q = f(K,L,t,D) \tag{1.13}$$

where Q is output, K is capital and L is labour. Output depends not only on the inputs but on an index of technical change through time t and on a dummy variable D which has a value one for one firm and zero for the other. This model assumes that the firms have the same production technology except for a shift parameter (D) at any moment of time (t). Since time and the shift parameter may interact there is no assumption that technical change affects the two firms in an identical form.

Assume that the production function may be approximated by a translog function in the four arguments. Then,

$$\log Q = \alpha_0 + \alpha_D D + \alpha_K \ln K + \gamma_{KD} D\cdot\ln K + \alpha_L \ln L + \gamma_{LD} D\cdot\ln L$$
$$+ \alpha_t t + \gamma_{tD} t\cdot D + \tfrac{1}{2}\gamma_{KK}(\ln K)^2 + \tfrac{1}{2}\gamma_{LL}(\ln L)^2 + \tfrac{1}{2}\gamma_{DD} D^2 \quad (1.14)$$
$$+ \tfrac{1}{2}\gamma_{tt} t^2 + \gamma_{KL} \ln K\cdot\ln L + \gamma_{Kt} t \ln K + \gamma_{Lt} t \ln L \quad .$$

This is a particular second order approximation to the production function (1.13). For a binary comparison, the variable D can be thought of as a dummy variable identifying the firm. In the translog example, the first order coefficients all are firm specific while the second order coefficients are common to all firms. At least for all approximations which are expansions of the original function, the highest order parameters are common to all firms. That is in a third order approximation, the first and second order parameters are firm specific while the third order are common to all firms.

Equation (1.14) provides an example of specifying a production function with some differences permitted across firms. The link between this specification and practical measurement can now be examined.

The difference in efficiency between two firms is defined by the equation

$$\ln Q_1 - \ln Q_2 = \hat{s}_K[\ln K_1 - \ln K_2] + \hat{s}_L[\ln L_1 - \ln L_2] + \hat{s}_D \qquad (1.15)$$

where the numeric subscript indicates the firm,

$$\hat{s}_K = \tfrac{1}{2}(s_{K1} + s_{K2}) \quad , \qquad \hat{s}_D = \tfrac{1}{2}(s_{D1} + s_{D2})$$

$\hat{s}_L = \tfrac{1}{2}(s_{L1} + s_{L2})$ and s_{ji} is the cost share of input j in firm i .

The logarithmic differences in the input uses are weighted by the average share of the input in each firm. The average difference in the efficiency of the firms $\hat{s}_D$ is equal to the difference in the logarithms of the output levels minus the sum of the weighted difference in the logarithms of the input quantities used. An estimate of the average efficiency difference, $\hat{s}_D$, can be calculated from observed prices and quantities of inputs and outputs. To understand what is implied by this particular measure of the difference in firms' efficiency, we can relate the measure to the production function in equation (1.14). The input shares s_{Ki}, s_{Li}, $i = 1,2$ can be related to the translog production function. For example,

$$\frac{\partial \log Q}{\partial \log K} = \alpha_K + \gamma_{KK} \ln K + \gamma_{KL} \ln L + \gamma_{KD} D + \gamma_{Kt} t$$

and a similar expression exists for all other inputs. In competitive equilibrium real factor prices equal the marginal product of each factor. Equivalently, the logarithmic marginal product $(\partial \log Q / \partial \log K)$ equals the input share (s_K).

The difference in the firms' technologies is measured as $\hat{s}_D = \frac{1}{2}(s_{D1} + s_{D2})$. This variable is unobservable. In terms of the translog production function,

$$s_{Di} = \frac{\partial \log Q}{\partial D} = \alpha_D + \gamma_{KD} \ln K_i + \gamma_{LD} \ln L_i + \gamma_{DD} D + \gamma_{Dt} \cdot t \,, \quad i = 1,2$$

The variable s_{Di} measures the logarithmic difference in the output of the two firms holding input levels and technical change constant. This is the Jorgenson-Nishimizu definition of the efficiency difference between the firms. Notice that the value of the efficiency difference depends on the level of the inputs and technical change (t). Consequently at any moment of time unless the two firms use the same input quantitites the efficiency differences between them depends on the input levels. For this reason, in equation (1.15), the average efficiency difference, $\hat{s}_D$, evaluated at the input levels of each firm is used. The choice of the average of the two values of the efficiency differences is arbitrary since other weights are possible. We will return to this point later.

To summarize the material to this point, rewrite equation (1.15),

$$\hat{s}_D = \log Q_1 - \log Q_2 - \{\hat{s}_K[\log K_1 - \log K_2] + \hat{s}_L[\log L_1 - \log L_2]\} . \quad (1.16)$$

The terms on the right hand side are all measurable from observations on the prices and quantities of outputs and inputs. Consequently, this equation can be used to evaluate efficiency differences between two firms. It is a discrete approximation to the instantaneous efficiency difference $s_D = \partial \log Q/\partial D$ evaluated at any K, L, t. The actual weights, $\hat{s}_i$, are compromises since observations are made at different input levels for each firm.

Within this model, the measurement of the rate of growth of productivity for any firm (D, constant) is derived as an approximation to

$$s_t = \partial \log Q/\partial t$$

the rate of growth of output holding all inputs constant. The approximation is

$$s_t = \log Q_t - \log Q_{t-1} - \{\bar{s}_{Kt}[\log K_t - \log K_{t-1}] - \bar{s}_{Lt}[\log L_t - \log L_{t-1}]\} , \quad (1.17)$$

where

$$\bar{s}_{Kt} = \tfrac{1}{2}(s_{Kt} + s_{Kt-1})$$

$$\bar{s}_{Lt} = \tfrac{1}{2}(s_{Lt} + s_{Lt-1}) \quad .$$

The rate of growth of productivity is measured using only the observable prices and quantities of outputs and inputs from the right hand side of equation (1.17). This expression is the usual approximation to the Divisia index used by this author and many others to measure productivity for a firm or industry. Consequently the methodology is consistent with recent productivity studies.

The methodology discussed above needs to be revised in several directions. This can be accomplished by using some current results by Denny and Fuss (1980). Define a quadratic function, $f(x)$,

$$Q = f(x) = \alpha_0 + \sum_i \alpha_i x_i + \tfrac{1}{2}\sum_i\sum_j \alpha_{ij} x_i x_j \quad .$$

Diewert (1976) has proved the following theorem. Suppose we consider any two vectors, $x_0 (= (x_{io}))$, and $x_1 (= (x_{i1}))$, then

$$Q_1 - Q_0 = f(x_1) - f(x_0)$$

$$= \tfrac{1}{2}\sum_i \{(\frac{\partial f}{\partial x_i}(x_1) - \frac{\partial f}{\partial x_i}(x_o))(x_{i1} - x_{io})\} \quad . \qquad (1.18)$$

Consider the production function used by Jorgenson and Nishimizu, $Q = f(K,L,D,t)$ in the particular translog form they selected. This function is a quadratic in the logarithms of capital, labour, the time variable and the shift parameter. The theorem on quadratic functions can be applied to this case. The theorem implies,

$$
\begin{aligned}
\ln Q_1 = \ln Q_0 = \tfrac{1}{2}\left(\frac{\partial \ln Q_1}{\partial \ln K_1} + \frac{\partial \ln Q_0}{\partial \ln K_0}\right)(\ln K_1 - \ln K_0) \\
+ \tfrac{1}{2}\left(\frac{\partial \ln Q_1}{\partial \ln L_1} + \frac{\partial \ln Q_0}{\partial \ln L_0}\right)(\ln L_1 - \ln L_0) \\
+ \tfrac{1}{2}\left(\frac{\partial \ln Q_1}{\partial t_1} + \frac{\partial \ln Q_0}{\partial t_0}\right)(t_1 - t_0) \\
+ \tfrac{1}{2}\left(\frac{\partial \ln Q_1}{\partial D_1} + \frac{\partial \ln Q_0}{\partial D_0}\right)(D_1 - D_0) \quad .
\end{aligned}
\tag{1.19}
$$

The application of this theorem provides a convenient and insightful method for interpreting differences in efficiency. The logarithm of the output ratio between any two firms depends on the weighted sum of (a) the logarithm of the input ratios, (b) the differences in productivity due to time and (c) the differences in efficiency at a moment of time.

The right hand side of equation (1.19) might appear to be difficult to evaluate. Recall that

$$s_K = \partial \ln Q/\partial \ln K \ , \qquad s_L = \partial \ln Q/\partial \ln L$$

$$s_t = \partial \ln Q/\partial \ln t \ , \qquad s_D = \partial \ln Q/\partial D$$

and that under competition s_K and s_L are the input shares. Rewrite (1.19) as

$$\begin{aligned}\ln Q_1 - \ln Q_0 = {} & \tfrac{1}{2}(s_{K1} + s_{K0})(\ln K_1 - \ln K_0) \\ & + \tfrac{1}{2}(s_{L1} + s_{L0})(\ln L_1 - \ln L_0) \\ & + \tfrac{1}{2}(s_{t1} + s_{t0})(t_1 - t_0) \\ & + \tfrac{1}{2}(s_{D1} + s_{D0})(D_1 - D_0)\end{aligned} \tag{1.20}$$

Equation (1.20) integrates the conventional Divisia index of productivity with the Jorgenson-Nishimizu inter-firm comparison of productivity. Moreover its implementation requires only data that are observations on prices and quantities. Some confusion may arise with the interpretation of the subscripts, $i = 1,2$. In general, the subscripts refer to the two sets of observations on prices and quantities. However, the particular origins of these two data sets are not specified by the equation. That is the two data sets may be observations on (a) two firms in the same time period (b) one firm in two time periods or (c) two firms in two different time periods. In case (a), since the time period is the same $(t_1 = t_0)$ the third term on the RHS drops out. The remaining terms are the Jorgenson-Nishimizu (see (1.16)) measure of the inter-firm efficiency differential. In case (b), there is only one firm which necessarily implies that $D_1 = D_0$ and the fourth term drops out. The remaining expression is the conventional Divisia index of productivity growth. The observations on prices and quantities are for the same firm in different time periods. The final case, (c), involves two firms in different time periods. In this case, we can not distinguish between efficiency differences that are due to differences in the firms at a moment in time and differences across time unless we apply econometric techniques. We can still measure the differences between firms but it will reflect a combination of the last two terms in equation (1.20). This is still useful information about inter-firm differentials.

In defining the production function, (1.13), we have implicitly interpreted technical change as shifts in the production technology through time. This is not required and if specific measures of technical change are available they may be used. However time passing is the most comprehensive measure that is likely to be available.

It will be useful to extend this analysis to incorporate cost efficiency in the multiple output case. Define a multiple output cost function,

$$C = g(w_L, w_K, t, D, Q_A, Q_B) \tag{1.21}$$

where w_i is the price of input i and Q_A and Q_B are the two outputs. Approximate this function with a translog cost function,

$$\begin{aligned} \ln C = {} & \alpha_o + \alpha_K \ln w_K + \alpha_L \ln w_L + \alpha_D D + \alpha_t t \\ & + \tfrac{1}{2}\gamma_{KK}(\ln w_K)^2 + \tfrac{1}{2}\gamma_{LL}(\ln w_L)^2 + \tfrac{1}{2}\gamma_{DD}D^2 \\ & + \tfrac{1}{2}\gamma_{tt}t^2 + \gamma_{KL} \ln w_K \ln w_L + \gamma_{KD} D \ln w_K \\ & + \gamma_{Kt} t \ln w_K + \gamma_{LD} D \ln w_L + \gamma_{Lt} t \ln w_L \\ & + \alpha_{QA} \ln Q_A + \alpha_{QB} \ln Q_B + \gamma_{LA} \ln w_L \cdot \ln Q_A \\ & + \gamma_{LB} \ln w_L \ln Q_B + \gamma_{KA} \ln w_K \ln Q_A + \gamma_{KB} \ln w_K \ln Q_B \end{aligned} \tag{1.22}$$

$$+ \gamma_{AB} \ln Q_A \ln Q_B + \tfrac{1}{2}\gamma_{AA}(\ln Q_A)^2 + \tfrac{1}{2}\gamma_{BB}(\ln Q_B)^2$$

$$+ \gamma_{DA} D \ln Q_A + \gamma_{DB} D \ln Q_B + \gamma_{tA} t \ln Q_A$$

$$+ \gamma_{tB} t \ln Q_B + \gamma_{tD} t \cdot D \quad .$$

If there are two firms, one and zero, and we wish to explain the difference in their costs, the quadratic theorem can be applied

$$\begin{aligned} \ln C_1 - \ln C_0 = {} & \tfrac{1}{2}\left[\frac{\partial \ln C_1}{\partial \ln w_{L1}} + \frac{\partial \ln C_0}{\partial \ln w_{L0}}\right](\ln w_{L1} - \ln w_{L0}) \\ & + \tfrac{1}{2}\left[\frac{\partial \ln C_1}{\partial \ln w_{K1}} + \frac{\partial \ln C_0}{\partial \ln w_{K0}}\right](\ln w_{K1} - \ln w_{K0}) \\ & + \tfrac{1}{2}\left[\frac{\partial \ln C_1}{\partial \ln Q_{A1}} + \frac{\partial \ln C_0}{\partial \ln Q_{A0}}\right](\ln Q_{A1} - \ln Q_{A0}) \\ & + \tfrac{1}{2}\left[\frac{\partial \ln C_1}{\partial \ln Q_{B1}} + \frac{\partial \ln C_0}{\partial \ln Q_{B0}}\right](\ln Q_{B1} - \ln Q_{B0}) \\ & + \tfrac{1}{2}\left[\frac{\partial \ln C_1}{\partial t_1} + \frac{\partial \ln C_0}{\partial \ln t_0}\right](t_1 - t_0) \\ & + \tfrac{1}{2}\left[\frac{\partial \ln C_1}{\partial D_1} + \frac{\partial \ln C_0}{\partial \ln D_0}\right](D_1 - D_0) \ . \end{aligned} \tag{1.23}$$

Differences in total cost between the firms is explained by differences in the input prices facing the two firms, differences in the output mix being produced and differences in the rate of technical change through time as well as the efficiency at a moment of time.

In the first two terms on the RHS of (1.23), the bracketed expression is the average cost share for labour or capital. In the third and fourth term, it is the average cost elasticity. If there is no information about the cost elasticity then an assumption has to be made about these terms. As we discussed earlier, the conventional treatment of productivity assumes that there is constant returns to scale, e.g. Jorgenson and Nishimizu, and we can do the same. However, this will imply that the measured efficiency differences between firms will include scale effects. This is acceptable and is simply an example of the difficulties of dividing up productivity differentials into scale and technical change. Productivity measures for a single firm require the same division and there are no new difficulties for inter-firm comparisons.

The major advantage of this procedure is the explicit development of the methodology that should alert the user of both its possibilities and limitations. The methodology developed for comparing the efficiency of firms implies that the technologies of the firms is similar. This is not in fact a limitation of this particular methodology. All non-econometric comparisons will have to assume some degree of similarity although it could be different from that assumed here. What is assumed here is that we can approximate the true cost function for two firms by a Translog second order approximation. The cost functions for the firms differ by only the shift parameter, D, which enters into many terms of the Translog cost function. If one chooses another functional form, it may not be easy to apply the quadratic approximation theorem and obtain results that have an easy interpretation. In particular, one may need to use econometric evidence to a greater extent than with the form we have chosen.

In the following large section, we will stress the necessity of careful measurement of outputs and inputs. This is required because productivity growth should not include the errors in measuring these variables. Our procedure accounts for differences in the levels of inputs and outputs but can only do so accurately if these are properly measured.

II.7 Technology and Economics in Telecommunications

One of the most difficult questions to answer is the approximate usefulness of the economists' abstract notion of a production function when one is confronted with a concrete and complicated telecommunications network. The abstraction from technical details is absolutely essential to permit a unified economic theory of production that is not encumbered by specific technical detail. However, it may result in problems in any specific application, e.g. telecommunications, if no consideration of the broad scientific underpinnings are undertaken. With that in mind this section provides a beginning for the investigation of the major aspects of the technology that are of some economic interest.

The term, scientific underpinnings, was chosen deliberately. There is no suggestion that detailed engineering studies are a substitute for economic analysis. The history of engineers attempting to do economics is miserable as the recent energy crisis has illustrated. What is required is a cooperative effort with the engineer or scientist providing technical expertise that can be amalgamated with the economic analysis. It is recommended that this be done in telecommunications during the next phases of this project. The following discussion is illustrative rather than definitive although it is partially based on the engineering training of one of the authors.

The provision of telecommunications services can be divided up into two broad areas. They are not distinctly separate in practice but for the purposes of our discussion there will be no serious flaws in our procedure. The first area is the network aspects of the system and the second is the characteristics of an individual link.

To illustrate the network aspects, let us construct an example. In Figure 1, seven locations A, B, C, D, E, F and G represent particular point sources for sending or receiving information. The problems of sending messages between any particular two sources, e.g. A and F will be our second concern. First we will consider the network relationships between all the points. Consider the simple seven location network in Figure 1. It would of course be possible to provide a direct and separate link between each of the locations and the remaining six. This would require forty-two separate links joining the locations and is very unlikely to be observed in practice. To investigate this, consider some simple economics. Suppose that the cost C of sending messages over any link depends on the distance travelled t over any link and on the volume of messages Q.

$$C = g(t,Q)$$

The initial network layout might involve costs for any link $C_{ij} = g(t_{ij}, Q_{ij})$ where the subscripts reference A, B,...,G and $i \neq j$. Why would we consider any other layout? We want to determine the lowest cost network for providing any given message volume, (Q_{ij}), between fixed locations. In our example, there is no growth and no indivisible and irreversible investment which should signal the simplicity of our example. First, we would like to know how costs varied with distance and message output level. Assume initially that there are constant returns to scale with respect to both distance and message output separately. Costs increase with both variables and multiplying <u>either</u> variable by a constant λ

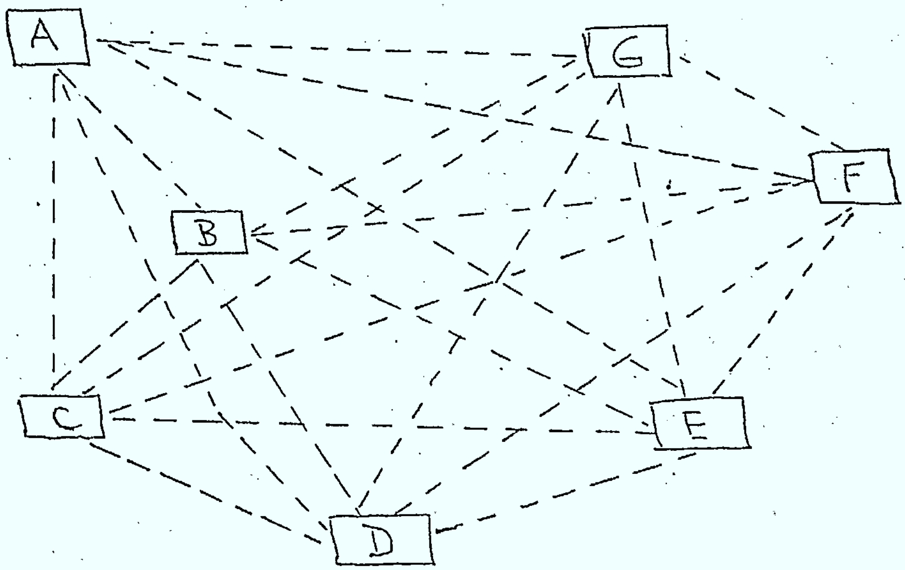


Fig. 1. A Simple Network

increases costs by that same proportion λ . For example suppose output was increased by a factor λ_Q and distance by a factor λ_T then under these assumptions, costs must rise by a factor $\lambda_Q \cdot \lambda_T$. With fixed locations A,...,G , there is no reason to have any other network layout. The separate links minimize the total cost since any other layout will increase total distance for a fixed total message output. Without introducing some additional consideration or altering our assumption of constant returns to scale all links are chosen.

For example, if there were increasing returns to scale in the production of messages over any feasible output range in our example then 'trunk' lines that carried more messages would be cost efficient even if the total distance travelled increased. The savings on message transmission would on average outweigh the increased distance costs and on the margin for the optimal network these costs and benefits would be equal. It is likely that there are increasing returns to scale with respect to messages due to construction activity. On the other hand it is difficult to conceive of increasing returns to scale with respect to distance. After relatively short distances, it is likely that there are constant or decreasing returns to distance.

Even in our simple network, there are choices to be made about the most efficient network design and consequently inefficient choices are possible. Implicit in our description is a timeless measure of output, messages. The network design had to carry the given volume of messages at any moment of time. This simplification is not meaningless since actual design concentrates on the capability of carrying 'busy hour' message volume. This is also a timeless output measure. Rather oddly, the telephone companies have planned the network to meet any timeless demand at the busy hour and then attempted to find a pricing system that would pay for the network. In our example and in practice this leaves potentially large message carrying capacity underutilized outside the 'busy hour'.

It would be relatively easy to add on the stochastic nature of demand. The demand for service, even during the 'busy hour', is not a constant. The probability of a line being demanded during any small interval may be described by a probability distribution like the Poisson. Similarly the length of time during which the line is held may be described by a proabability distribution such as the Geometric or perhaps Lognormal. While these facets of a telecommunications network are of vital practical importance they do not provide a wide array of new economic issues. In passing one might mention that the major cost implication of stochastic demand is the choice about the quality level. That is what is the probability of not being able to obtain a line. There may be sharp cost differences associated with choosing different levels of this design parameter.

Three aspects of most telecommunications systems will not be introduced. First, we have said nothing about switching. Literally with no switching we might imagine that our network had six entirely separate private lines at each location. To make our example concrete, suppose that the telecommunications network is for voice transmission. At any location, we might imagine (a) one of more telephone-type instruments with six jacks representing the six private lines to the other locations, (b) six or more telephone-type instruments permanently attached to the lines, (c) one or more telephone-type instruments each with the capability of connecting to any of the six lines. Option (a) could be conceived of as manual switching. The user must plug his instrument into the jack to make the connection desired. With option (b) there is no switching since a premanent connection exists for each line. Option (c) implies that the telephone instrument has built into it the capability of selecting the desired connection. This is a form of built-in switching.

For the moment we will ignore the possibility of having more than one instrument in option (a) and (c) and more than six instruments in option (b). This brings in the question of multiple voice channels which we will introduce in a moment.

Assuming that only one voice transmission is possible on any link, there are obvious cost differences involved in the simplied switching alternatives, (a), (b) and (c). For example, we would like to know the relative cost of providing a telephone instrument with automatic switching such as that considered in (c) compared to the costs of user switching in (a) and multiple permanent connections in (b). Presumably (a) is cheaper than (b) in terms of equipment but user time costs in making connections will offset the equipment saving. We will return to switching after introducing a number of very crucial aspects of a telecommunications network.

Most telecommunications networks provide service on demand. That is the user does not order a particular link and quantity of message transmission and wait for its production. The user is able to assume over a wide variety of services that the capacity is available on demand. This is not a necessary part of a telecommunications network. One could have a network in which users ordered services and paid a price based on how soon they wanted a transmission and on the message quantity to be transmitted and distance of transmission. If service is to be available on demand then the existing capacity will be underutilized much of the time. The existence of excess capacity is not necessarily inefficient. Rather it is a characteristic and a real cost of providing service on demand.

Could an alternative type of service not available on demand be provided. Certainly the general answer is yes but some attributes of providing telecommunications services may suggest that there are reasons for providing service on demand.

Most telecommunications links have required high fixed costs relative to variable costs. Whether this has changed through time with the newer types of links is something we might investigate in a later phase of this project. To the extent that this is a true characteristic of the links then the price for usage in the short-run should be low. Having service available on demand is not the same thing as a relatively low short-run demand price. However, the large fixed investment incorporated into any link implies that producing only after demand appears is not particularly sensible. That is, the basic links can not be put in place and removed depending on demand. Once the transmission link is in place some type of service on demand is quite feasible.

There is one further problem that requires a brief discussion before we try to bring the various arguments together. On any link it may not be technically feasible to produce a link whose capacity is as small as the smallest demand. The most obvious examples relate to the local telephone lines. The local loop between a subscriber and his local Central Office is reserved for the subscriber. Not only does the subscriber not use the line most of the time but when he is on the line, the line is capable of carrying more than one voice channel but it never does. This should be kept distinct from the previous issues of service on demand and underutilization. The idea in this instance relates to the technical question about the minimum capacity on any link. This may exceed the maximum demand.

Let us try to draw these various issues together and relate it back to switching. There is a great deal of potential in a telecommunications network for underutilization of the links based on both purely technological considerations as well as the combination of time varying demand imposed on a link with high relative fixed costs. This may also result in underutilization of the switching.

Switching has almost never been done at the sending and receiving locations such as A, B,...G in our simple network. Rather in order to reduce total switching equipment costs through higher utilization of equipment, local exchange switching centers have always existed. Since large segments of the network linkages will not be used at most if not all moments of time, switching utilization will be increased by linking each location to a switching center at which all switching is done. This will increase the total distance travelled by a fixed quantity of messages. However, the total costs will be less due to the savings in switching equipment expenses.

The economic tradeoff is between the increasing cost of longer lines in the local loops versus the savings in switching capacity and in using trunk lines as the size of the local Central Office grows.

A switching center for only seven locations is tiny and at least today could easily be installed at each of our locations. However, once one thinks of thousands or millions of locations the situation is different. The physical space alone, required to provide switching equipment at every location for every other location would not be justifiable. Switching via cross-bar or step by step systems usually permits local exchanges with 10^4 phones for example. The probability of very many of the possible links in a network of even 10^4 locations being used at one time is very small. For perhaps a majority of the links the probability of any use at any time is also small.

It is the switching function that dominates the design of a local network, but other considerations may enter into the longer distance transmission. Over a longer distance the costs per unit distance probably rise more than proportionately with mileage. Ever since WWII developments have been made that have substantially reduced the costs of long range transmission. Most of these are associated with the use of trunk lines that package many voice channels together in order to lower the cost of sending a given quantity of information.

In a larger more realistic network, one of the important features is the possibility of choosing alternative routes between any two points. On the local loop, a single connection is permanently wired into a particular Central Office. However once a call is to proceed beyond the Central Office of the originating party there are a variety of options available. Trunk lines between local central offices provide voice channels that can be used for any call. If necessary, routes that are not direct can be chosen as alternatives to the most direct routes. Route selection is of great importance for long distance calls and the alternatives are much larger than with local calls.

The volume of long distance calls is quite small compared to the volume of local calls. Since constructing lines over long distance is expensive, there are strong incentives to collect calls and send them over a limited number of high capacity lines. The stochastic nature of the demand for any particular link permits the total capacity of the system to be reduced through the availability of a variety of alternative routes. There is a limit to the alternative segments for any connection since the quality of transmission falls as additional separate line paths are added. The process can be illustrated by considering an example illustrated in figure 2.

There is a hierarchy of offices illustrated in the figure. An originating call will travel to the local central office where it is switched and sent to a toll center, From here it will be switched to a primary center. If a route is feasible, i.e. exists and is unutilized it will travel directly to the local toll office of the called phone. From there it can proceed directly to the called phone through the local central office. As the diagram illustrates there are a large number of alternative possibilities. The alternatives are numbered by the priority with which each would be selected. Routes with more switches are less desirable since more switching equipment is utilized and the quality of transmission falls as the number of switches increases.

The marginal cost differential of transmitting over any route other than route one is likely quite small today. Provided the quality of transmission does not fall too severely the available alternatives lowers the capacity required to handle peak periods. One of the re-

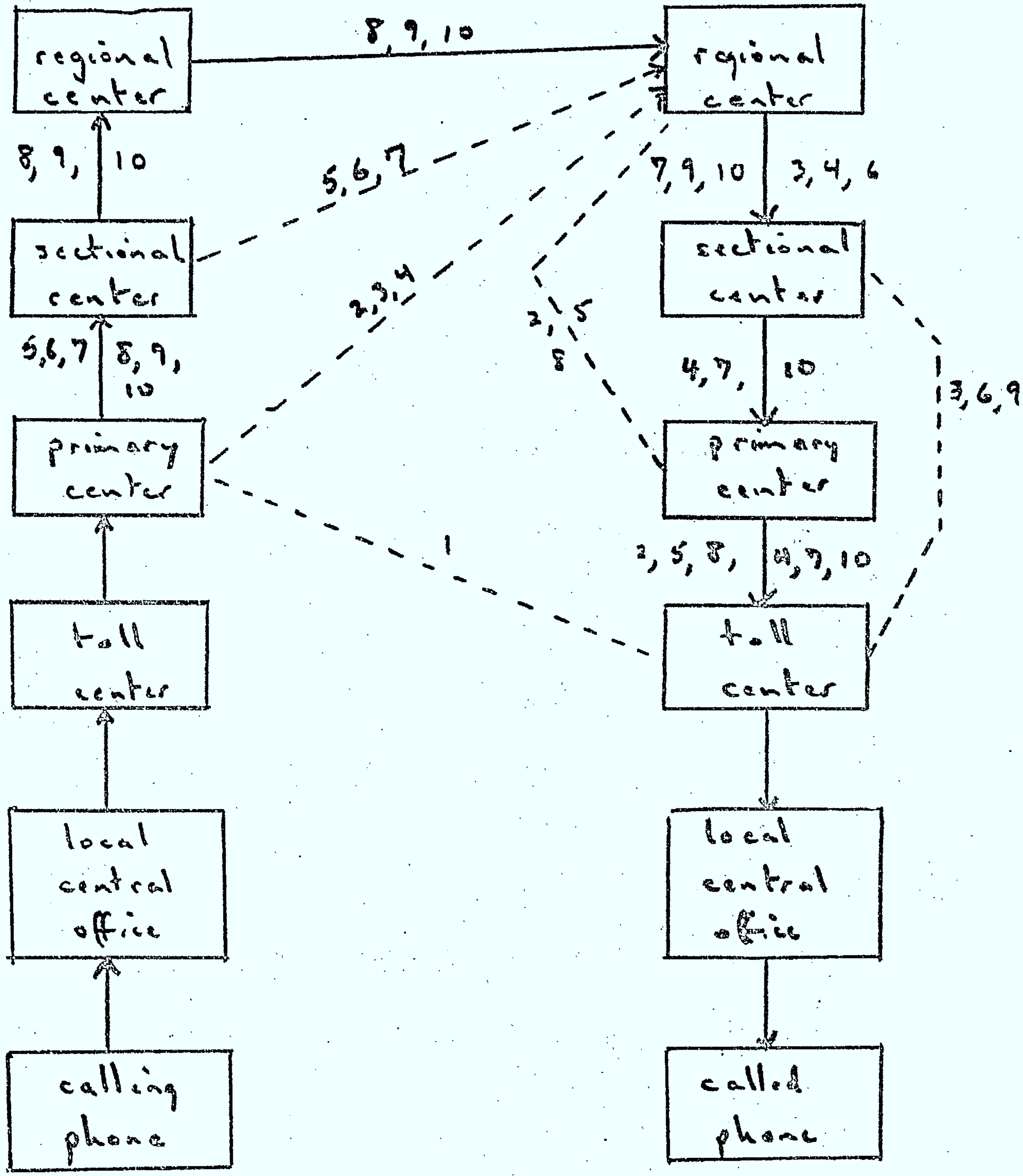


Fig. 2 : Alternative Routes

quirements of this type of flexible system is that searching for an open route must not take too long. The newer switching equipment with its greater flexibility and higher speed has enhanced the cost saving features of alternative route selection.

How are the characterististics of the network going to enter into an economic analysis? We have seen already that the switching equipment is going to be vitally important for holding costs down in the local network. Unless the local networks are redesigned to eliminate the permanent "private line" between the subscriber and the local central office, there will be large underutilized capacity in the local transmission network. This does not imply that the switching capacity is currently available to handle the capacity of local lines. It certainly is not in place. However the marginal cost of increased transmission is mainly related to the switching capacity in the local network. There may also be local trunks that are operating at capacity for at least part of the day. The changes in local loops may be dramatic in the next decade. The possibility of delivering many new non-voice services to the individual subscriber are evident. The channel capacity required will be much larger than the current local telephone line although the existing cable television lines have much larger channel capacity. The large time of day and day of week variations in demand provide the main network features that are not sufficiently analyzed. If the existing demand for service can be shifted into off peak period or if new services can be offered off the current peaks, output can be expanded without large increases in inputs.

Attempting to summarize the technology underlying a particular channel or link in a telecommunications network is difficult. Our intent is to cautiously portray an abstract version of the technological aspects that would appear to be most important for economic analysis. It is not conceivable that a few pages will substitute for centuries of engineering and scientific knowledge , but taken in context the important issues can be outlined.

It is perhaps simplest to consider the problems abstractly from the point of view of communications theory and concretely in the form of telegraphy, the oldest form of electrical communication. To maintain simplicity, there will be no network only a simple channel or link from point A to B. A message is to be sent from A to B via telegraphy. The message must be translated from its original form at A to the appropriate form for sending down the channel to B. There, it must be retranslated into the final form. The process of translation is usually referred to as encoding and much of the work in communications theory relates to efficient codes. The actual transmission process is more closely related to the physical properties of the media. The primary question is the speed at which we can send information down the channel. Affecting both of these problems is the noise associated with any electrical circuit. The presence of noise implies that there is always some probability of an error in the message received at B.

Initially, a system with DC signalling will be investigated. In the simplest single current system, electrical current is turned on and off at the sending end of the circuit. The receiving end perceives a similar although not identical pattern of on-off states. The message to be sent, for example an English text, must be encoded into a sequence of on-off pulses which are sent and received before translation back into an English language message.

By 1838, Morse had developed his code in which letters of the alphabet are represented by dots, dashes and spaces. The dot is an electric current of short duration, the dash an electric current of longer duration and a space is an absence of current. Dots, dashes and spaces were assigned to letters in a manner which minimized the length of time to send a message. That is, commonly used letters were assigned short combinations of the dots, dashes and spaces. It is known that a code using dots, dashes and spaces could be constructed that would be roughly fifteen percent more efficient than Morse code. However, the efficiency of Morse code is quite remarkable.

There is a limited speed at which Morse code can be sent over any simple telegraph line. A short pulse of current sent at one end of a link is received at the other end as a much more elongated smoother rise and fall in current. If one attempts to increase the frequency of short pulses sent, the receiver will find that the elongated pulses can no longer be distinguished.

Single current telegraphy uses only one level of sending current. It is possible to use more than one intensity of current. This was understood and utilized in the nineteenth century when double current telegraphy and much higher intensity differentiations were introduced. For each level of current, the speed of signalling is still limited by the elongation of the pulse signal at the receiving end. Moreover, as we increase the number of different intensities sent, there will be problems distinguishing between the different current intensities. Added on to these two difficulties is the presence of noise which is always present and which will tend to make identification of one or many current intensities difficult. Noise may be overcome by using more power, i.e., higher current intensities, but there are limits to the power that can be used on a link before a short-circuit will appear.

The early telegraph developpers realized in an informal manner most of the difficulties that would delineate the feasible technologies of the future. The rate at which signals can be sent over any line is limited. It is difficult to distinguish between many alternative current values particularily in the presence of noise. Finally while noise can be overcome by increasing the power of the signal, there are limits to the feasible power that can be applied to a particular line.

Why would an economist care about the technical characteristics of a communications channel. Fundamentally, as economists, we are interested in the characteristics of the constraints on the supply of telecommunications outputs. The output of the communications channel can be considered as the number of bits per second. We are interested in the possibility of devoting real resources to altering the signal to noise level or providing more sensitive symbol detectors or choosing new transmission machines in order to increase the output level. Are there scale economies in line transmission, can rapid substitution be made between different message types and sources etc.? Knowledge of the technical aspects of the supply system improve our ability to explore methods of altering efficiency and evaluating public policy.

The capacitance, inductance, resistance and leakage in a circuit will provide limits on the speed at which data may be transmitted. These characteristics of the circuit will affect the shape and amplitude of the arrival signal at the receiving end of a circuit. For a given circuit the speed of transmission will certainly fall with distance and in general the speed will fall more than proportionally as the distance is increased. To overcome this problem the use of regenerative repeaters is required. The fact that repeaters (regenerative or not) can be economically placed along the line does suggest that the major costs are not in transmission but in the line itself.

There are theoretical limits on the capacity and speed at which information can be sent down a channel. The most important is perhaps the rate at which information can be sent is proportional to the band-width. That is for a given quantity of information and a given bandwidth there is a minimum time that will be required to transmit the information.

In a noiseless channel the channel capacity, C, measured in terms of hits per record would be

$$C = 2W\log_2 L$$

where W is the bandwidth and L is the number of distinct signals being simultaneously. In more complicated channels, with noise for example, the expression will change. However it will still be true that the capacity of actual telecommunications systems is roughly proportional to the band-width. It is also true that the capacity can be increased by increasing the number of different distinct signal levels (L). However notice that the relationship is such that the capacity will not increase as quickly as the number of signal levels.

Most of the existing telecommunications networks use continu-ous analog, not digital, signalling. In those circumstances the capacity of a noisy channel will be

$$C = W \log_2 (1+S/N)$$

where S/N is the signal to noise ratio.. The same proportional relationship exists between the capacity and the bandwidth. However the ability to distinguish signals is now related to the power of the signal relative to the noise. If the signal-noise ratio can be increased at reasonable cost, we can obtain more capacity from a given channel.

It has not been possible to develop this section to the extent that we would have liked. Hopefully the examples, incomplete as they are will clarify our intent. The specific technological knowledge underlying a particular industry should be used to enrich any economic analysis of efficiency. The engineering or scientific information should clarify the concrete form in which increases in efficiency have been embedded into the capital of the network. To do this fully would be an enormous effort but our more modest efforts can be enhanced by a further attempt to incorporate the explicit characteristics of the technology. It is hoped that this can be done in conjunction with the telephone company engineers particularly in the measurement of capital.

III. Total Factor Productivity: The Theory and Practice of Output and Input Measurement

III.1 Introduction

In section II, we have discussed the conceptual foundations of efficiency measurement. This section investigates the problem of defining the real output and real input measures required to measure productivity. The relationship between the desired concepts and the feasibility of attaining these measures is studied. Several of the telecommunications firms are currently measuring productivity and their methods of measuring outputs and inputs are reviewed.

Why do the exact methods of measuring inputs and outputs receive so much attention? A trite answer is that measured productivity will vary with alternative measures. More seriously there are at least two strands of concern. Economists have developed a theory of efficient resource allocation and we would like the chosen measures to be consistent with that theory. The measures should also conform as closely as possible to the economic theory of welfare. In contrast to this emphasis, it may be useful to alter the measures to permit variants which are of more interest to the firm or the regulator.

Although not stressed in the text, the feasibility of achieving the desired measures will be severely constrained by cost. It is quite possible to begin with crude measurement and as the value of productivity measurement is perceived then further improvements can be made.

III.2 Outputs: Consumption and Production

To clarify the problems associated with measuring output it is useful to consider an abstract notion of an elemental production process. Such a process is to be understood as a complete description of a set of inputs and detailed processing that produce known specified outputs. The elemental processes are given at any moment of time and the existing ones are independent of current economic decisions. The latter will have important implications for future elemental processes. Why are these elemental processes different from a production function? The basic distinction has to do with the level of complexity of the technical processes and socio-economic organization. That is it is assumed that firms or sections of firms are involved in combining elemental processes to produce outputs. While the term production function could be applied to the elemental processes, for the purposes of this discussion, it could not then be used to describe the technology of the firm. Since the term production function has been used widely at the firm level, the major purpose of introducing a new term is to clarify the wider choice that is presumed to be available to the firm while maintaining a technical constraint on its behavior.

The notion of an elemental output is useful as an anchor. There is a huge range of detailed possibilities of "packaging" characteristics as an output through combination of elemental processes. An output from an elemental process that is literally packaged and shipped to a destination could be considered as the output from a combination of at least three elemental processes. The elemental processes, however numerous, produce outputs with characteristics that can not be unpackaged. Combinations of elemental processes prodcue outputs although the possible combinations may be limited by the technology. Technologies describe not only how inputs can be combined to produce elemental outputs but also how outputs that are combinations of elemental processes can be produced.

A problem arises at this point concerning the characteristics. The concept of a characteristic of an output or commodity is not well-defined. What we will accept is the idea that the list of characteristics exhausts the description of the commodity. We are assuming that information about the characteristics of a commodity are available at no cost to both producers and consumers. However the quantity or presence of any particular characteristic may be of no importance to either or both groups. Alternatively the relative importance of any particular characteristic for either group may vary substantially. This distinction will create some interesting problems for practical applications as we will see later.

The brief conceptual diversion can now be used to clarify certain problems in output measurement. These problems are associated with the methods of accounting for changes in quality. Changes in quality imply changes in the characteristics of the outputs through time or between firms. An excellent example of the controversy is contained in Griliches (1964). As he discusses, the controversy is related to the possibility of separating output measurement from welfare. The answer is surely no but the implications of this judgement must be explored.

From a consumer perspective, an individual has a utility function defined over commodities. The list of commodities for which the utility function is defined must be carefully specified. To maintain the link with the production sector, the commodities in the utility function will be identical to those specified as outputs of the production activities. The consumer is presumed to maximize his welfare which in this context implies maximizing the value of his utility. The important point to stress is that the consumer knows the relationship between the commodities and their characteristics and his welfare. If one wishes one can introduce a Becker-type consumer technology that translates the goods obtained from the producer into utility. We have subsumed this in the utility function. However what is crucial is the following link. If the characteristics of the commodities in the utility function are not directly measures of welfare they are certainly indirectly related to welfare. The demand for these commodities is a derived demand from utility maximization. The characteristics of a commodity that interest a consumer are those that ultimately affect his well-being. This does not mean that output is a measure of welfare but only that the characteristics of a commodity that must be included in its description are those that the consumer desires because they indirectly affect his welfare. The producer must also be concerned with the characteristics that matter to the consumer under any profit-maximizing economic system. This reinforces what is fundamental but often obscured. It is not possible or useful to try and completely separate welfare and production. If outputs are to be measured with a detailed specification then the charactersitics that are important to consumers as welfare maximizers must be held constant. If they are not then consumer behaviour will be altered and presumably producer behaviour will also change. The output from production is not a measure of welfare but if the characteristics of the commodity that effect welfare are altered output must change. If problems of quality change are to be avoided, the description of the output must include the welfare derived characteristics in which the consumer is interested.

On the production side, gross output is the desirable output magnitude. The gross output is simply the flow of output per time period that leaves the production unit. As noted above the producer is concerned with characteristics of the output not only due to his concern for consumer response but also due to his own cost considerations. That is, some characteristics of the output can be altered without affecting the consumers evaluation of the product although producers' cost may change.

For the firm, a continuous series of choices must be made about the package of characteristics that will be contained in the outputs it produces. These choices usually imply that through time the bundles of characteristics offered for sale will be altered and that the quality of products change both at the market boundary and at non-market boundaries.

There can be a long sequence of events between the initiation of production and the consumer's satisfaction from the consumption of the product. This rather vague sequence has to be split at some point to define output of the production sector and the arguments of the utility function. The existence of market transactions in commodities has provided a useful although arbitrary dividing line. It is not arbitrary in the sense that the location of the boundary between producer and consumer cannot be explained. It is arbitrary in the sense that it can shift and its location is not based on the notion of what is an output. Alternatively we can state that neither the concept of a production function not that of a utility function is sufficient to determine the commodities in which market transactions occur.

The Market Boundary

The sequence of events from inputs to satisfaction is divided by the market boundary. It is there that transactions take place that exchange money (or goods) for goods. On the one side of the boundary there is the firm with a production technology which limits its behaviour and describes the possible transformation of the inputs into outputs. A very large abstraction has to be made concerning the relationship of science, technology and the internal social ordering of the work force to define the production technology. We have subsumed all this in the conventional production function for both the elemental production processes and the complex production processes derived from the elemental ones. The output of the production function becomes the market output. This framework focuses the analysis onto the market value and the determination of market prices. On the other side of the boundary is the consumer whose satisfaction from the purchased output may require time and other resources not purchased in this transaction.

There is a tendency in most applications of economic theory to freeze the location of the market boundary by the definition of the outputs from production that enter into consumption. In general there is no theoretical reason not to determine the boundary endogenously within the model. Historically, the market boundary has shifted due to the use of the consumer's own resources to produce more or less of certain products outside of the sphere of market activity. Certain production processes have shifted across the market boundary. If we are interested in markets or economic activity on markets then the market defined commodities are what we should attempt to measure. To the extent that we wish to study a wider range of socio-economic activity then the market determined commodities may not be the most useful or the only output measure that should be collected. Even for the study of the market economy, the collection of non-market output will be necessary to understand changes that arise in non-market areas that shift particular commodities from one side of the production process to the other.

While it is certainly easiest to collect information on market transactions there will always be problems for which this in inadequate. Although thoroughly underdeveloped at the moment, it is feasible to develop models that explain the location of the market boundary. Although not formally developed they would analyze the problem of the existence of firms that use many elemental and complex production processes. Firms always face the 'make or buy' question on the input side and the question of produce or not produce on the output side. For most firms there is a complex range of internal transactions between users that are not market transactions. Exactly which transactions will be market and which non-market has not received very much attention.

In telecommunications, there are interesting and pertinent examples of all these problems. The outputs of the industry are often classified under three headings: access, usage and terminals. For example AGT used these categories to form groups from a detailed list of services at the second Teleglobe symposium on productivity. This is a natural procedure since there exist tariffs for all these types of services. The existing market boundaries define the conventional output categories.

Are these output categories satisfactory? We may not be able to reach a definitive answer but some limitations of these categories may be suggested. It is reasonable to argue that usage is what consumers of telecommunications services are buying. Consequently some quantitative measure of usage is required. Usage implies a time dimension but this is certainly not sufficient. For any service the quantity of time must be supplemented by the other characteristics of the service.

III.3 The Measurement of Outputs in Telecommunications

The method of measuring outputs is important. There is no natural notion of output independent of the purposes for which the data is intended. Firms undertake activities and provide physical goods and less tangible services, some of which are distinctly priced and others which are provided as part of a package whose components are not separately priced. We have attempted to present an abstract notion of an output in the first part of our discussion. The intent is to stress the ease with which alternative output measures can be defined, particularly if one selects the outputs as defined by the existing market boundary. A more detailed consideration of the outputs produced by telecommunications firms follows.

It is not sensible to select a single measure of output. Rather a multitude of physical and constant dollar measures, together with quality characteristics, should be collected. This will permit (1) an improved understanding of the attributes of 'output' that are changing, and (2) a better control over the difference in the output mixes produced in different firms. The latter is necessary for comparing efficiency.

There are at least three distinct purposes that may require some differences in the output measures. The companies are interested in profits and consequently the resources accruing from the rates of telecommunications services. In this context, outputs might be defined as the products that are priced in the markets. If there was no change in the quality or characteristics of outputs through time or between companies these measures would suffice. However there are many changes in telecommunications outputs. For this reason, output measurement will have to be investigated from the perspective of economic theory and efficiency and from the regulator's viewpoint. It is not necessary that large differences in output measures for different purposes exist but the possibility cannot be eliminated without careful study.

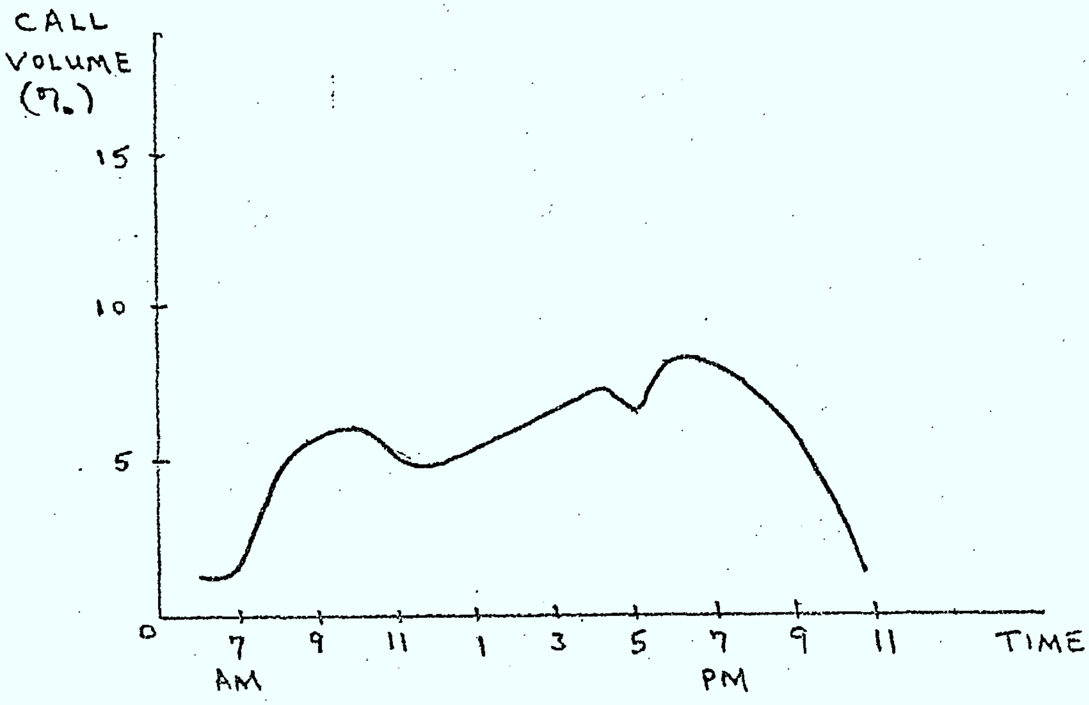
Local service revenue is generated by several major types of outputs as well as a very large number of minor ones. The largest share of local service revenue is derived from the flat fee charged for local service. These fees are differentiated by (1) business vs. residential service and by (2) rate group, i.e. the number of phones in the local calling area. The current definition of the market boundary is quite special and a variety of alternatives exist but were not selected. Local calls are not metered for time, distance or amount of switching required. Historically the metering costs were sufficiently large relative to the costs of providing local service that metering could not be justified. Recent widespread metering of local calls in Europe and the introduction of more metering in the U.S. implies that this is no longer true. The flat fee gives any user unlimited access to the local network. If local usage increases through more phone calls or higher average duration or higher average switching, there will be no increase in output if prices at the existing market boundary are used to deflate local service revenue. On the other hand for the same volume of usage, output will change if the number of main stations i.e. telephone lines, increases or if existing users switch between rate groups or between business and residential.

For the firm, concerned with revenue, changes that generate increases in revenue without price changes might be treated as output changes. However for other purposes, this definition is misleading. It may be possible to increase output with no changes in inputs e.g. increases in off-peak calling. This is not feasible with other definitions of output that monitor the number of calls. Flat fees that ignore the number of calls, combined with price discrimination across the class of user, do not recognize any of the usage characteristics of demand or supply. Only the inputs required to affect access are reflected in this type of output measure.

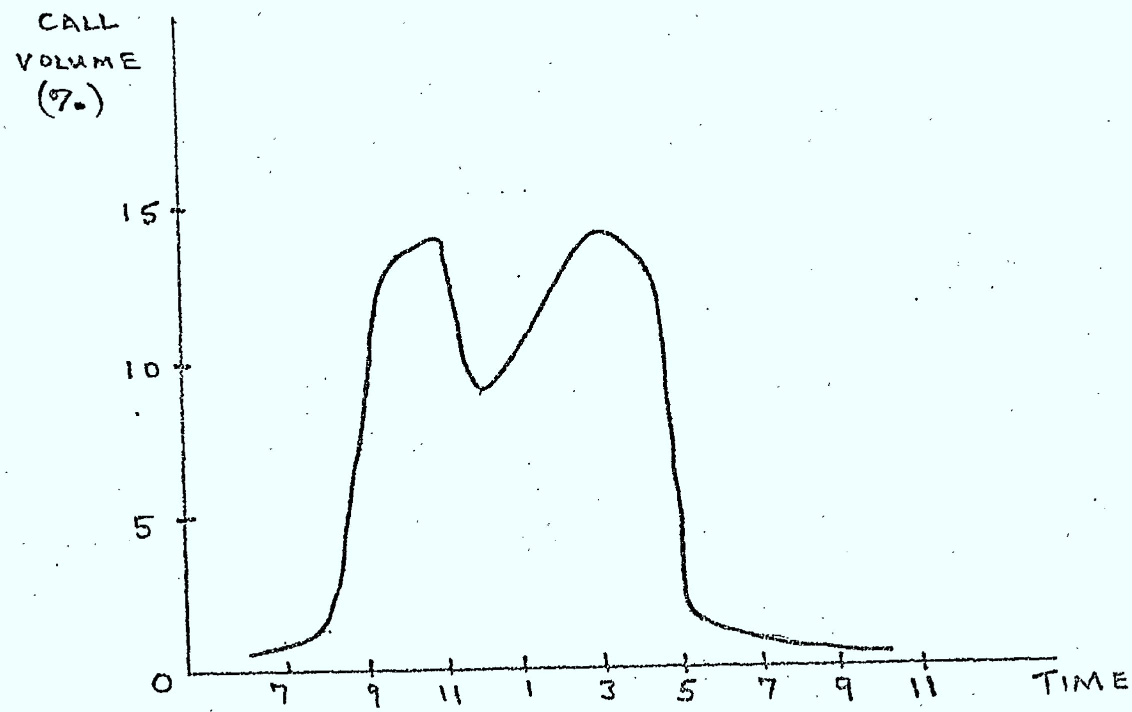
The role of time in the output measure is very important for local service. There are two aspects. First, is calendar time which inlcudes the month, date, day and time of day. The demand for local calls has a well established general pattern for most days of the week. There are weekend patterns and seasonal variations.

There is a relatively common temporal call pattern for residential and business exchanges on weekdays. These are illustrated in figures 1 and 2 . The distribution does not cover the night-time period since call volume is very low for both groups. The residential demand pattern is less peaked than the business pattern. Since telecommunications systems are engineered for peak demand, it is the business exchanges that have the largest excess capacity off the peak period. In both residential and business exchanges there are huge quantities of empty hours which might be filled by other services in the future.

There is also a very skewed distribution of usage classified by customer. That is a very large percentage of all calls are made by a relatively small number of customer lines. With a flat rate service, the average cost per call varies widely among customers. The marginal cost to the user is of course zero. With the rising interest in metering local service, it will be useful to obtain data on number of calls. In a system with no time of day pricing the capacity of the system is not utilized during most of the year. Although time of day pricing would lower the peaks it is not known to what extent calls can be shifted to off-peak periods.



RESIDENTIAL LOAD



BUSINESS LOAD

The experience of the New York Telephone Company (NYT) will provide an example. Hopley (1978) describes the experiment. NYT introduced a 27% discount on local calls placed from 9:00 P.M. to 9:00 A.M. relative to the rate of 8.2¢ per call charged in the period 9:00 A.M. to 9:00 P.M. The scheme was introduced in the NYC and Buffalo metropolitan areas. The regular call rate as well as the discount only apply after a limited number of "free" calls. It is not surprising that the shifts in calls from peak (9:00 A.M. - 9:00 P.M.) to the off-peak period was minimal. As Hopley notes this is not the only criteria. The major rationale for peak load pricing is to allocate the costs of the system to usage at different periods of time. This is a useful objective even if the call shifting is minimal. The basic problem with the NYT experiment was the minimal price differentials between only two large periods. As we have seen above substantial variation in demand exists within the peak period. However, the shifting of costs out of the off-peak night-time period certainly makes sense. The local network is engineered to have a very low probability of not being able to connect a subscriber who is not using their phone. This implies that the switching equipment or inter-office trunk lines are not utilized fully. For comparing companies the main implication is that if the call pattern over calendar time differs, then for the same network of subscribers, different real resources will be required.

The duration of the call is the other major time factor. If the length of calls or pattern of length of calls over calendar time varies across firms then different input resources will be required.

Since several phone companies distinquish local rates by the number of subscribers in the same local exchange, the consequences of this practice for output measurement is worth studying. The general pattern across

Canada has been for the population to become concentrated in urban areas. It is in these areas that the top level of the rate groups exist. Consequently the local service output level has been rising due to shifts in location of existing users and the location choices of new users. In making inter-firm comparisons, we want to know the extent to which output growth has been based on differences in rate groupings and shifts across these categories. For example, it is known that the introduction of Extended Area Service raised rates in Ontario and Quebec by shifting subscribers into higher rate groups. What quantity of extra output should be associated with this shift?

Local service revenue includes a large quantity of equipment rentals. For residences, some examples are touch tone phones, extensions etc. and for business there is a wider variety of specialized equipment such as PBX, data communications and Centrex. This is an interesting example of a complex market boundary. Certain equipment is priced separately from either access or usage of the communications links. Moreover it is perfectly reasonable to view the equipment that is rented separately as an intermediate input into the provisions of telecommunications' services. In that context, increases in equipment, even though rented separately, might be viewed as an increase in inputs rather than outputs. For most types of rented telecommunications equipment this would seem to be partially appropriate.
Equipment rentals provide access to the network with a different set of characteristics than the basic black box telephone. This suggests that one might experiment with hedonic measures of output that permit calls to have characteristics. However in the early stages of development this will not be feasible.

Two examples can provide some illustrations of the judgements that are required. Extension phones have grown more rapidly than main stations during the last twenty years. The charge for the extension is a price for improved access to the network from a larger number of locations in the user's building. Users are substituting more telephone equipment for their own walking time and providing flexibility in the use of rooms for a variety of purposes including telephoning. Users are willing to pay a price for this convenience and the output level certainly has changed. One may treat this as a change in the quality or quantity of output. The growing importance of this source of revenue should be carefully investigated so that if it is treated as output quantity not quality, the implications are understood. There may be no affect on total messages carried on the network due to extension phones. Increases in output via this source are qualitatively and quantitatively distinct from increases in usage.

A more significant change is currently occurring in Bell Canada's territory. Households are being converted to a 'new' type of access system. Phone jacks are being provided up to a maximum number at no extra charge. The basic monthly charge includes the use of the phone jacks and one 'black box' telephone. More phones can be bought or rented from Bell Canada or competing suppliers. The new system will alter the extension market for Bell Canada quite substantially. From Bell Canada's perspective, the most important part of this change is the servicing of equipment. Customers will not have phones serviced in their residences but will be asked to take the phone to a repair depot and exchange it for a different one. This is a decline in the quality of output provided although it may not show up directly as a decline in the quantity. In fact it should permit Bell to reduce the capital and labour used for maintenance. Efficiency could rise due to a decline in quality not a decrease in resource use for a given quantity and quality of service.

These examples highlight the care that is required if overall changes in productivity are to be understood for a single firm or compared between firms. Without precluding special items not discussed above the following data should be acquired.

Local Service revenue should be broken down into at least four groups: (1) residential basic charges, (2) business basic charges, (3) residential other local revenue, and (4) business other local revenue. Price indexes for these four items would have to be developed. This should be viewed as a minimum requirement. For at least the firms with better data bases, a more detailed investigation into (a) the components underlying these categories should be completed, (b) the usage and (c) the time varying demand pattern, should be undertaken. Data on the number of local calls and their duration should be collected under (b) while (c) would involve estimates of the demand patterns over time. This information is important for assessing the extent to which capacity exists within the local networks of different firms. Comparisons across firms will be improved by the inclusion of this information.

A more detailed study of the components of our four basic local output measures should cover the question of the changes in output due to shifts amongst rate groups and the introduction of EAS. There are a number of quite different services included in the other local output, items

(3) and (4) above. These include, equipment rental, non-recurring charges, public phones, private lines, message charges and some smaller items. It will be useful to consider variations across firms in the relative importance of these items in case the firms vary in the ways in which they price their services.

For Bell Canada, we know that Bell Canada's measure of constant dollar local service revenue grew by fifty percent more than the number of local calls from 1952-76. The number of local calls grew only slightly faster than the number of main stations. What is required is an explanation of this result. Growth in output through an increase in minutes of phone calls is very different than output growth through equipment rental, for example.

<u>Toll Revenue</u> : This is the seond large group of outputs and the rate of growth of these outputs has tended to exceed that of local outputs. Although technologically most of the service being provided is not different than for local output, i.e. voice message service, the method of pricing is substantially changed. This is an excellent example of redefining the market boundary. In toll service, metering of usage is done by distance, time of day, day of week and duration of the call. Output measurement can be much more easily calculated as a flow of services rather than the unlimited access to the local network. Historically the costs of long distance lines were sufficiently large that the metering costs were justified as a means of pricing. This was not true on local lines where metering costs would have been too large a proportion of total costs for local calls.

The measure of toll output should reflect all of the detail now used in the pricing of toll. That is distance, time of day, day of week, duration and type of call. The latter refers to the use of special handling methods e.g. person to person, charge account, operator assisted, etc. The current Bell Canada toll output measure is well constructed in this manner. The major difficulty with this procedure involves the weights to be used in aggregating. It is not obvious that the existing price structure provides the appropriate weights. However they will have to be used unless cost information can be collected.

The apparently improved situation for measuring toll outputs relative to local outputs is made more obscure by the problems of calls which originate in ones firm's jurisdication and end in another firm's. The revenue settlement procedures within Canada and between Canada and the U.S. and Overseas may create difficulties for output measurement. A station-to-station call of ten minutes duration on Monday from 10:00 to 10:10 A.M. from Newfoundland to Vancouver represents a certain quantity of output. However, this output is going to be split up between the operating companies. It is important that the price and output allocated to each company for this call bear an approximate relationship to the costs incurred and output generated by each firm. Transit traffic is particularly important in the Prairie provinces but the problem will arise elsewhere. Judgements will have to be made about the exact methods to be used in these intra-firm toll calls.

While message toll dominates the toll area, it is not the only output. The major other outputs are WATS and Private Line Service. Price and quantity indexes for these outputs will have to be developed.

Miscellaneous Outputs: Revenue will be received for a variety of services not associated with the direct production of telecommunications services. Probably the most important is the production of telephone directories. However there will be some rental and perhaps licencing income as well as scattered minor sources. Provided that these revenues are not a substantial portion of total revenue, relatively crude price indexes may be used to deflate revenue to obtain constant dollar revenue.

III.4 The Measurement of Inputs in Telecommunications

We are concerned with three broad classes of inputs, labour, capital and all other inputs. called materials. Many of the issues that were discussed in the output section are pertinent for the measurement of inputs. It may be useful to summarize the major conceptual problems.

First, the inputs ought to be measured in units of constant quality both through time and across firms. Since qualities are constantly changing, attempts must be made to evaluate the consequence of imperfect control over quality variations. Second, the inputs should be measured in flow units not stock units wherever possible. Man-hours and machine-hours are preferable to employees and number of machines. More precision can be given to these notions through a direction discussion of the various inputs.

Labour: Simple measures, for example, total employees are apt to be biased as input measures due to changes in both the quality and the man-hours per year. Man-hours worked (or at least on the job) should be used as the labour unit. Quality adjustments to man-hours must be made. Labour should be disaggregated into detailed occupational groups. The wages of these groups in each year will provide variable weights to aggregate over the occupational groups. Disaggregation by occupational group is particularly important when the occupational mix has been changing. In telecommunications companies, telephone operator man-hours have fallen and the hours of highly skilled well-paid technical, administrative and managerial occupations have grown during the last two decades. These are only rough examples of broad trends. The degree of detail that is feasible will depend on the costs but detail is desirable particularly where large relative changes in the importance of different skills groups have been occurring.

Total labour must be separated into man-hours on own-account construction and operating account. The disaggregated occupational man-hours will also have to be subdivided in this way although perhaps crudely.

For reasons that will be clearer in the discussion of capital, some attempt should be made to estimate labour used in maintenance. The companies provide Statistics Canada with maintenance labour expense data. If some approximation to man-hours can be calculated for this expense, it may be quite useful.

The major limitations to providing good labour usage measures will reside in the accounting records of the firms. How costly will it be to extract a historical record on disaggregated labour man-hours and costs. Labour costs including all benefits by occupational group will be required. Wage rate data will not suffice except as a rather poor approximation since the product of the wage rate times the man-hours worked summed over all occupational groups will not equal total labour costs. Since the accounting classifications for labour hours and costs may not be identical across firms, the effects of non-uniformity will have to be checked qualitatively or quantitatively if possible.

The methodology employed by Bell Canada in their productivity study is very useful. They have disaggregated by occupation, skill level and experience. The breakdown beyond occupation provides further controls on changing quality of the man-hours. If other companies can produce data of this quality, there should not be serious problems with the measurement of the labour input.

Materials: This term is used to cover all of the inputs, other than capital and labour, that are part of the operating expense of the company. Non-operating expenses are discussed separately in the section on management information systems and productivity uses. In telecommunication firms these expenses include supplies, rents and a variety of miscellaneous items. The diversity of items will create problems for the construction of accurate price indexes. In Bell Canada's productivity study, materials were deflated with the price index for GNE. This should be improved by an investigation of the components underlying the aggregate series. It may be necessary to use price indexes from Statistics Canada that reflect the bundle of goods included in materials. The possibility of improving on the Bell Canada procedure certainly exists and should be attempted. The treatment of material by B.C. Tel. is discussed below.

Indirect taxes should be allocated to either outputs or inputs dependinf on the type of tax. These taxes should not be treated as a separate input or amalgamated with the materials inputs.

Capital: Telecommunications is a very capital intensive industry. The methods chosen to measure capital will have an enormous affect on the measured efficiency of firms and on inter-firm comparisons of efficiency. The flow quantity of capital services utilized in any time period is very difficult to measure. Moreover the quality of major components of the stock have changed over the lifetimes of much of the equipment. For example, new transmission systems, new multiplexing procedures and new switching equipment exist with equipment of earlier vintages. Aggregation in this context provides special problems. A simple example will illustrate the problems.

Consider a truck with a given capacity for carrying a load and an expected lifetime of T years assuming a particular maintenance schedule. The capital service flow from this truck might be measured by flows such as the ton-miles carried or the operating hours during any time period. These flow measures are generally unavailable and stock measures are often substituted for them. The number of trucks might be used as the input measure. To the extent that the service flow per truck varies over time, there will be either an over-estimate or under-estimate of the capital services flowing from the stock. More serious problems arise if there are trucks of different durability; i.e. longer lifetime, but similar load carrying capabilities. A more durable truck need not provide any more services per operating hour than a less durable one. However, many standard practices would aggregate the two types of trucks giving the more durable machine, with the higher price, a higher weight and consequently implying that more services were provided per time period. If trucks that are more efficient are introduced during the lifetime of earlier trucks one must weigh the new and old trucks in such a way as to implicitly compare their relative efficiency. None of this is easy even when discussing trucks. For the diverse and complex telecommunications networks, the problems are severe.

Many of the major telecommunications companies have invested large quantities of resources in detailed revaluation of their existing capital stock at current reproduction value. This data will have to be used as the basis for the construction of the capital input. The methods used to revalue the stock need to be investigated in detail, in order to assess the impact of the particular methods selected on measures of comparative efficiency. It should be feasible at some modest level of disaggregation to consider alternative procedures for obtaining an aggregate capital stock that was purged of some of the problems that will arise.

Telecommunications capital can be divided up into a number of particular groups related to the network. It will be useful to have estimates of the reproduction value of outside plant, central office equipment, station equipment, other equipment and other capital not included elsewhere. This should provide weak information on the variability of types of capital across firms.

As we stress below in the section on the technology, it will be useful to have some physical dimensions of the network layout. Although the exact data required is not clear at this stage of our work certain rough notions can be provided. Data on the exchange lines and toll lines measured in miles, systems, channels and one-way channel miles will be useful. Included should be direct miles of conduit, microwave systems mileage and channel miles as well as the same type of data for high frequency radio systems. Data on the network characteristics will be necessary to attempt to perceive why there are efficiency differences. Without some information on the network characteristics it is possible that comparisons will be misleading. Much of this data is currently available since it is supplied to Statistics Canada by all firms.

The measurement of capital inputs will require the most careful attention in any comparison. Only a combination of constant dollar quantities of capital and physical measures will permit a high quality comparison. Fortunately much of the data is available in many of the firms through their own efforts.

The collection of data is an expensive process. While accurate data is important for the measurement of productivity the tradeoffs between desired data and costs will have to be part of the process of defining a precise list of data that will be collected. It is impossible to be definite without some participation by the companies since their practical knowledge is required in order to judge the costs. The sensitivity of our results to alternative data specifications is also a practical question and we should continuously be evaluating the cost effectiveness of any procedure as we proceed.

It is very important that the data collection for productivity not be unrelated to the other management information systems currently in place. The data required for productivity is either part of the data required for many managerial purposes or may be directly used for these purposes. If it is treated as a distinct entity then opportunities to realize benefits from the initial expenditure will be lost.

III.5 Measurement in Practice: an Overview

The purpose of this section is to place a practical perspective on the conceptual issues of examining inter-firm comparisons within the context of "Productivity, Employment and Technical Change in Canadian Telecommunications." While elsewhere in this study the questions of measuring efficiency, output, input and technology, index number choice, etc., which are fundamental and must all be thoroughly understood before undertaking inter-firm comparisons, are attached at a theoretical level, we approach them through a survey of existing and ongoing empirical work. In this vein we have chosen to present the practical aspect from a variety of viewpoints. It must be kept in mind that these pertain more to the fundamental questions rather than directly to the issue of inter-firm comparisons. They include: (1) a cataloging from published sources of the methodologies and measures developed at Bell Canada (B.C.), British Columbia Telephone Company (BCT), Alberta Government Telephones (AGT) and Teleglobe Canada (TC), these being representative of telecommunications. The first three, due to limited public information receive ralatively scant coverage (with B.C. receiving the most, followed by BCT and AGT), while TC is reviewed in somewhat more detail. The extensively developed and ongoing studies of Electricité de France (EDF) provide us with an excellent example of practical work outside of telecommunications; (2) a sample from the existing literature in the area of the ecomonic analysis of productivity in a regulated environment. These include studies of airlines, trucking and railways; (3) the operational aspect of productivity analysis which looks at its use as a management and regulatory tool in control and planning. This draws upon: applied work at EDF, the French government and AT&T; the automatic rate adjustment debate; and imminent applications at TG; (4) specifications of alternative operational productivity and economic efficiency measures.

An examination of (1) and (2) would be not only incomplete, but unfair as well, without at least a cursory look at certain theoretical aspects of the conceptual issues involved in the measurement of Total Factor Productivity. Since the theoretical view is given detailed treatment elsewhere in this study, we will confine ourselves to providing the reader with a brief look at its more salient features in order that the operational measures described be kept in appropriate perspective. We propose to look at existing measures in terms of: (a) the definition of output; (b) the measurement of input factors; (c) the level of disaggregation and (d) the method of indexation. While the form of items (a) to (d) crucially depends on the initial prioritization of desired results expected from the productivity index, we are here principally interested in the consequences of the choices rather than the initial decision to actually develop a measure. It should be emphasized, however, that whatever the original motivations for developing productivity indicators (whether for regulatory, management or both purposes) the choice of approach to items (a) to (d) can be cited as an important source of variation between the results of such measures, among different firms. Clearly, another source of difference, and one which may be far more important, lies in the reality behind the various approaches even after they have been standardized. While two firms, in the same industry, by conventional definition, (say, two "telephone companies") may have adopted identical measurement techniques, the composition differences in their output and consequently, in their choice of technologies, may be so vastly different as to produce entirely dissimilar productivity results. The convention of definition may in fact be quite inappropriate when a "telephone company" in one region is in fact not a "telephone company" elsewhere. While the importance of this issue cannot be exagerated, it is given more detailed coverage elsewhere and we can leave it in order to concentrate on the pure measurement and definition aspects.

The most important source of variation between measures of different firms is due to item (a), the operational definition of output. Take, for example, company A, which provides all domestic telecommunications services to a particular locality, where management has somehow determined, during the pricing and costing exercises that its main product was access to the telephone network. However, some indenpendant researcher (through his own observation) determined that the firm's clients were in fact purchasing usage of the network. Naturally, the time paths of output growth measured according to the different definitions would coincide only by chance. As a consequence, even with identical measures of input, the productivity growth results will diverge.

This is a fairly superficial example of output related difficulties. In general, the measurement of output is a very serious conceptual issue. In Einsteinian terms, it is all relative to the observer. The consumer might not be purchasing what the producer thinks he is selling. These difficulties arise because, essentially, any good is really the sum of its characteristics.

If you are an interior decorator, you may be purchasing both the design and communications characteristics of the extra cost coloured telephone, while if you are an elderly individual, communication is all that counts.

At Teleglobe Canada, for example, output could have been measured in terms of access to international telecommunications simply by counting and aggregating all circuits, appropriately weighted for grade. However, due to strong evidence, such as increasing time duration per message over time (which is also why messages, per se, were rejected as a measure of output), usage, in terms of message minutes was chosen as the output definition. Given the differential growth rates in minutes and circuits (leading in some cases to severe congestion) these would certainly have resulted in quite different time paths of output growth and, as a consequence, drastically different productivity results.

As far as input factors are concerned, while differences in their measurement do arise, they are related more to questions of classification, data availability and indexing. Economic theory provides some very clear indications as to the measurement of not only the capital stock but, as well its cost in terms of a stream of services flowing from the asset over its life. Opinion does differ, but only to a minor extent as to which items should be included in the cost of capital. However, any excluded item, say different types of taxes, will nevertheless still appear as part of total input, even if under another heading. Labour in terms of manhours worked is generally accepted as the correct version of manpower input. Intermediate inputs, being the mixed bag of unidentified as well as difficult to interpret (in terms of productivity) items such as various taxes, miscellaneous type of expenses, etc. has the choice of appropriate price vector as its main source of difference.

The third source of difference arises due to varying levels of data disaggregation. Formally, if

$$X = f\,(\Sigma_j X_{ij}, \ldots\ldots\ldots \Sigma_r X_{mr})$$

where the aggregator function (denoting some specific index), f, provides that each

$\sum_{k=1}^{Z_i} X_{ik}$, for $i = 1 \ldots\ldots m$ and $Z_s \gtreqless Z_t$ changes the value of the aggregate index, X, in proportion to some measure of its importance, S_i, then it will not, in general, be equal to

$$X^* = f\,(X_{11}, \ldots\ldots X_{15}, X_{21}, \ldots\ldots X_{25}, \ldots\ldots, \ldots\ldots, X_{m1}, \ldots\ldots X_{mn})$$

where each X_{ij} enters the aggregation through the same aggregator function, f, as above, but with a unique weight, S_{ij}. We can easily construct an example based on the distribution of the labour force within some firm. The table below is constructed with a view to developing a Laspeyres quantity index. We assume three major classes of labour, each with its own sub-categories.

PERIOD / CLASS	0 VOLUME	0 VALUE	1 VOLUME	(1 + %Δ) VOLUME	BASE WEIGHTS
1	60	750	67	1.12	.51
11	20	400	22	1.10	.27
12	30	300	33	1.10	.20
13	10	50	12	1.20	.03
2	19	280	23	1.21	.19
21	5	150	4	0.80	.11
22	6	90	9	1.50	.06
23	8	40	10	1.25	.03
3	35	440	34	0.97	.30
31	15	300	12	0.80	.20
32	20	140	22	1.10	.10
TOTAL:		1470			1.00
					1.00

The aggregate index, X, is

$$\Sigma_i \frac{Y_i}{Y_{io}}\left(\frac{V_i}{\Sigma V_i}\right) \quad \text{where } Y_i = \sum_{k=1}^{Z_i} X_{ik} \quad \text{for } i = 1.....m \text{ and } Z_s \gtreqless Z_t$$

$$\text{and} \quad V_i = \sum_{k=1}^{Z_i} V_{ik} \quad \text{for } i = 1.....m \text{ and } Z_s \gtreqless Z_t$$

$$= \text{the appropriate value of the ith class.}$$

Its numerical value is:

$$(1.12)(.51) + (1.21)(.19) + (.97)(.30) = \underline{\underline{1.09}}$$

The disaggregate index, X^*, is:

$$\sum_{i=1}^{m} \sum_{j=1}^{Zi} \frac{X_{ij1}}{X_{ij0}} \frac{V_{ij}}{\Sigma_i \Sigma_j V_{ij}} ; \quad Z_s \gtreqless Z_t$$

Its numerical value is:

$$\left[(1.10)(.27) + (1.10)(.20) + (1.20)(.03)\right] + \left[(.80)(.11) + (1.50)(.06) + (1.25)(.03)\right] + \left[(.80)(.20) + (1.10)(.10\right] = \underline{\underline{1.04}}$$

Although the difference may not always be as dramatic, it will usually be present. While disaggregation should not be carried to absurd extremes whereby within one grouping of, say, labour, distinctions are made for the "weekday-of-birth" differences, it should reflect degrees of importance due to skill, experience, education, etc. In the case of non-human factors, important distinctions can be based on quality or scarcity.

The final source of difference, although solidly grounded in economic theory, is purely mechanical. It is the choice of indexing method. The difference between the results of using a fixed base index from those derived through a chained version of the index (unless some incredibly stringent restrictions are met) will usually be far larger than the difference between two different indexes when they are both chained. The choice of index depends on a theory of what underlying forces are driving the system. Since index numbers have exact functional counterparts, they can be chosen to reflect (or test) the production technology which is believed to be endemic to the system under study. Happily index numbers range from the more restrictive Laspeyre's or Paasche to the more general (which allow for substitutability) such as Divisia and Diewert's quadratic forms (Diewert 1976). Some situations are more exigent than others. A "true cost of living index", for example, is well known to be bracketed by the most common fixed weight indices, Laspeyres and Paasche.

While all four items are clearly very important to an understanding of not only the actual measures developed by each of the firms, but also the observable differences in their results, only output and input measurement will be covered in detail with commentary on indexing and aggregation added only where they sharpen our analysis of the main points. Each section, in keeping with our goal of outlining the various existing operational productivity measures developed for telecommunications, will look at the general concept and then place it in the context of specific examples. Given certain confidentiality restrictions only that information which is either publicly available or has been put at our disposal through prior agreement, will be used in this study.

The uses aspect of productivity receives detailed coverage in a separate section.

III.6 Outputs

Differences in the measurement of output, between individual telecommunications firms stem from the definition of output and the choice of observable variables. With respect to the latter source of difference, in order to measure real output, one must be able to isolate the price and quantity components of the value term which, in our case, is the revenue expression. Furthermore, since value is naturally the product of price and quantity, or, by extension, value index = price index x quantity index, (a property known as "factor reversibility"), we need theoretically observe only two sets of variables, revenues and prices or quantities. We then construct, say, a price index, and derive the quantity index implicitly. Not all indices, however, have this desirable property. For example, the Laspeyres index does not,

$$\frac{P_1X_1}{P_0X_0} \neq \left(\frac{P_1X_0}{P_0X_0}\right)\left(\frac{P_0X_1}{P_0X_0}\right)$$

Fisher's ideal index, on the other hand does,

$$\frac{P_1X_1}{P_0X_0} = \left[\left(\frac{P_1X_0}{P_0X_0}\right)\left(\frac{P_0X_1}{P_0X_0}\right)\right]^{\frac{1}{2}}\left[\left(\frac{P_1X_1}{P_0X_1}\right)\left(\frac{P_1X_1}{P_1X_0}\right)\right]^{\frac{1}{2}}$$

$$= \frac{P_1X_1}{P_0X_0}$$

Thus, by choosing an index that has this property (or is at least a close approximation) either a quantity or price index can be constructed while implicity deriving the other. This is the essense of the "direct vs indirect" issue whereby quantity indexes required for TFP measurement, are calculated directly or derived indirectly by deflation of value indexes with directly calculated price indexes. Given, however, that very few indexes have this property, the choice of starting point will likely influence the result. This problem, as well as the various other aspects of index number construction is covered in greater detail elsewhere. In this section, we will simply point out, for each case, which of the two approaches have been used.

Differences due to output measures are based on compositional and definitional considerations. While any two firms may have an identical array of products, their distributions may be dramatically disparate. Beyond this purely market phenomenon lies the difficulty of perspective. Much of the telecommunications output embodies at least two characteristics; providing the means of accessing the network, and, ultimately of using it. The fundamental issue, then, is how can these principle characteristics of each output be, first of all, isolated, and then defined, i.e. how can we include the double criteria of access and usage in a usable definition of output?

The access criteria argues that service constitutes the availability of the telecommunications medium (or network), i.e. consumers are willing to pay a positive cost for an assurance of access even without any attendant usage. This is certainly somewhat extreme, but it highlights the importance of quantifying that aspect of output. For the measurement of access, attention is focused on items such as connector terminals, circuits miles, number of telephones sets, etc. (A more detailed description can be found in Werner, M., Routledge J.). However, this cannot be the only basis of measurement. For example, if output is measured as the number of telephones, then the price index will be a measure of the average revenue received per telephone and with more intensive demand for, say, toll calling, the increasing price index will be misleading.

Telecommunications service, defined as the use of the network, reflects the capacity aspect of the firm. This aspect constitutes a net addition to the size of any telecommunications system. By way of illustration, imagine the two extreme cases whereby on the one hand use, above and beyond access, is guaranteed with the capacity constraint such that only one subscriber can use the system at any time and at the other end of the spectrum all those who have purchased access can simultaneously use the network. Clearly the existing networks have been developed to operate at some intermediate point, and it is for this reason that the output measures developed for studies such as productivity, attempt to include, as well as the access aspect, good indications of usage. While the volume of access equipment (such as number of telephones) served as a fairly reliable guide to the growth in telephone company output in the past, the rapid development (and consumer acceptance) of the usage aspect of existing equipment (requiring relatively small enhancement), through the message toll venue, has rendered these simplistic measures, if not misleading, certainly far less useful. For example, for B.C., local service revenues, in constant 1967 dollars, declined from 64.4% in 1952 to 51.7% in 1976, as a proportion of total revenues. Taking into account that these revenues have grown faster than the number of main stations, it is clear that number of telephones is not a very reliable measure.

With these views on the various possible individual and collective definitions of output, let us examine some specific examples. Since it is the least restrictive, in terms of confidentiality (although it is the most different in terms of output and technology), we may begin with the "international" telecommunications sector.

A. International Telecommunications:

Given that production, as in any firm, is always subject to resource constraints (which, if it was not would make the price of output equal to zero), we should consider all those resources which push our production possibilities frontier outward. In the case of the individual firm, the basic constraining resource is financial. We may assume that all physical resources demanded will be available at constant prices in any period (ignoring difficulties such as a marginal efficiency of capital schedule). The problem of the firm is then to choose that combination of resources which will optimize its objectives, whether we regard them as profit maximization, cost minimization subject to a technological constraint or output maximization subject to a cost constraint. In this light, it is only natural that all sources of financial return be included as a measure of real output. That is, the real part of total revenues, from all sources, must be considered as the output of the firm. This includes the returns from that part of capital which is still in liquid form (in cash or short term assets) either because of lack of investment opportunities or structural rigidities (such as strict investment guidelines for government owned firms). With this view in mind, Teleglobe Canada defines the real value of its total revenues, in every period, as its output. As will become apparent in the subsequent breakdown, this includes some items whose connection with the business of the firm (i.e. the provision of international telecommunications services) is somewhat opaque. However, their significant size as well as their ready interpretation in quantity terms, does not allow them to be ignored.

Finally, as will be seen, most of the output categories enter the calculation as quantities while prices are implicitly calculated.

Given the behaviour variations of indices at different levels of aggregation, Teleglobe attempts to capture the sources of change in its index by collecting data at a very disaggregate level. They break output down into a large number of definable and observable categories. Teleglobe, in their breakdown, does not have to be overly concerned about the access - usage question, since most access is provided by firms outside of the Teleglobe net. A very small percentage of its business actually involves the sale of both access and usage. We can, of course, open the issue of access by considering that Teleglobe does in fact provide both access and usage to all its clients whereby they can use the international network of cables and satellites as a mean of access into other domestic systems around the world. But, as can be imagined, this sort of reasoning could continue ad infinitum where access and usage touch each other accross the entire range of economic endeavour. The line has been drawn at the usage aspect of the Teleglobe network.

The fundamental justification of viewing output at such a disaggregate level is that each individual item constitutes a different service, at least in terms of costs. There is a different structure associated with the receipt of a person to person telephone call, during business hours, from the U.K., than with the transmission of the same person to person message during the same time period and so on. The aggregation procedures which are designed to account for these differences use revenue shares as weights. These are, of course, to a large extent, output price determined. Ideally weights reflecting the cost elasticity of output would have been preferred, but since these are unavailable at such a detailed level of disaggregation, it is assumed that, in some sense, the selling price of each individual service type reflects, if not its absolute cost at least its relative share of the total. That is, an assumption if made that the distribution of prices somehow mirrors that of costs.

The basic output classifications, which form the disaggregated categories include (and bracketed numbers immediately after the headings, refer to the number of categories):

a) International Telephone (182), which includes 9 streams (each of which refers to the telecommunications between Canada and another country, group of countries or regions) accounting for over 70% of volume subdivided into 10 different types of service with two directions (incoming and outgoing). The remainder of the telephone traffic is aggregated into an "other" category with two directions. The full range of the telephone breakdown is listed in Table I, where an X indicates that data was available and the footnoted items indicate that data was available only for certain selected years. Its basic physical quantity unit of account is the "message minute".

b) International Telex (24), which includes 11 streams, accounting for over 70% of telex traffic, and another category aggregating the remainder. These are further broken down by direction. Table II summarizes the breakdown. Its basic physical quantity unit of account is the "message minute".

[redacted]

TFP

TELEPHONE BREAKDOWN

Call category / Stream	Initial								Overtime								Comb.		ISD					No breakdown	
	Person-to-Person				Station-to-Station				Person-to-Person				Station-to-Station				Sta-Sta		Full		Red.		Comb.		
	Full		Reduced		Full		Reduced		Full		Reduced		Full		Reduced		Full	Red.							
	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	In	In	Out	In	Out	In	In	At	In
Australia	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x[10]	x[1]	x[10]	x[1]			
Belgium	x	x	x	x	x	x	x[4]	x	x	x	x	x	x	x	x	x				x[1]		x[1]			
France	x	x	x[2]	x[3]	x	x	x[2]	x[3]	x	x	x[2]	x[3]	x	x	x[2]	x[4]				x[2]		x[2]			
W. Germany	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				x[6]		x[6]			
Greece	x	x			x	x			x	x			x	x									x[4]		
Italy	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				x[5]		x[5]			
Japan	x	x	x	x	x[5]	x[5]	x[5]	x[5]	x	x	x	x	x[5]	x[5]	x[5]	x[5]			x[10]	x[1]	x[10]	x[1]			
Switzerland	x	x			x	x			x	x			x	x									x[6]		
U.K.	x	x	x	x[7]	x	x	x	x	x	x	x	x[7]	x	x	x	x	x	x[9]		x[1]		x[1]	x[8]		
Other																								x	x

1 76 - 77
2 74 - 77
3 74 - 76
4 71 - 77
5 75 - 77
6 72 - 77
7 70 - 73, Total pers-pers red. 76
8 73 - 75
9 except 75
10 77
11 74 - 76

TFP

TELEX, TELEGRAPH BREAKDOWN

TABLE II

	TELEX[1]		TELEGRAPH	
	Inward	Outward	Inward	Outward
Australia	x	x		
Belgium	x	x		
France	x	x		
W. Germany	x	x	TOTALS	TOTALS
Greece	x	x		
HongKong				
India			ONLY	ONLY
Italy	x	x		
Japan	x	x		
Switzerland	x	x		
.K.	x	x		
Other	x	x		

1. Complete breakdown available only for the years 1975, 1976 and 1977. Totals are used for the other years.

c) International Telegraph (2), which is not only the smallest of what are known as the public services, but is as well a shrinking proportional of overall output. It enters the output calculation as a total broken down only by direction. Its basic physical quantity unit of account is the "paid word".

d) Leased Circuits (2), constitute the only access usage combination offered by Teleglobe Canada. However, as with the local services of some domestic telecommunications firms, there is no very clear division between access and usage pricing of leased circuits. The consumer leases a circuit at a fixed cost and determines his own usage level. Although the price is calculated on the basis of average expected usage, the fact that it is a competitive service reduces inequities, that may result through variations in usage, as a consideration. In other words, unlike local services, the consumer can choose between several different options for his access needs. These allow, as well, considerations of usage. The competitive aspect, however, is fairly complex. It exists between Canada and the U.S. and within Canada. In the former case, a consumer in one country can lease an inter-Canada/U.S. circuit and an international circuit in the other country through which his international traffic will flow. The choice will depend on the prices governing any individual's distribution of service needs; i.e., the volume split between voice, telex and data, as well as the final countries of destination, since prices will vary with traffic termination points.

d) Leased Circuits - cont'd.

Within Canada, the system of competition is far more complicated and, as well, highlights the entire access/usage debate. When Teleglobe leases a circuit it provides access to unlimited usage of the system at a fixed monthly charge, much like the local service network in domestic telecommunications. However, the client, at this stage, unlike the local service customers, can, through judicious use of definitions, such as service bureau, subdivide this circuit and release it to smaller, lower volume, users. Thus, I.P. Sharp can lease a circuit from Teleglobe Canada and provide access on demand, through the use of its own multiplexing equipment, to smaller clients. The provision of service on demand, as pointed out elsewhere in this report, may not be feasible, given the low ratio of variable to fixed costs which characterizes the local network. Another form of competition comes indirectly through the use of sophisticated peripheral equipment, such as Codex, which essentially doubles the capacity of a single voice circuit. Thus, the manufacturer of Codex is basically lowering the marginal cost of the extra circuit. Finally, Teleglobe itself, in offering competition to the vendors of service on demand, through its ICAS (International Computer Access Service), is in effect competing with its other circuit leasing service. Ultimately, then, as mentioned at the beginning, the inequities, as one may observe in the local service network,

d) Leased Circuits - cont'd.

through variations in the usage of the leased circuit service, are, to some extent, mitigated.

This category is broken down by grade of circuit, either voice or telegraph. Its basic physical quantity unit of account is the "circuit" (either voice or telegraph grade).

e) Other Services (1), constitute a mixed bag of very specialized offerings which, given their very small size, simply enter the output calculation as an aggregate. Their basic physical unit of account is the constant dollar. They include, television, transmission, datel, packet switching, etc.

f) CTFA Cost Recovery (1), is part of a financial arrangement to which only members of the British Commonwealth subscribe. The Commonwealth Telecommunications Financial Arrangements tries to pool resources that will avoid unneccessary redundancies on that part of the international network used and operated by its members. It operates on a cost apportionment basis, and is composed of two major items:

(i) Terminal Traffic Cost Recovery:

The subsidization of high unit cost administrations by the low unit cost ones constitutes the value of this component. As a simple illustration let us imagine that the entire commonwealth telecommunications network consists of only two partners, A and B. Total traffic (inward and outward of both administrations) is 100 units. Total revenue for the system is $500.00.

	A ⟷	B
Total Cost	250.00	150.00
Incurred Cost	2.5/unit	1.50/unit
Allocated Cost	2.00/unit	

B must provide the commonwealth clearing house with $50.00, since the cost to the system is that much larger than its own unit costs. A, on the other hand, will receive that $50.00 as compensation for its higher unit cost in providing its end of the facilities which make service possible. The outcome is that both A and B face equivalent unit costs.

If this transaction is now related to the larger picture, and its effects on profits are examined, we are able to draw some interesting conclusions. Given the accounting rate basis of revenue sharing (and assuming that ownership of facilities is equally shared), both A and B each received 250.00 of the 500.00 generated revenue. In the absence of CTFA, A would have had zero profits, with B reaping a net gain of 100.00. However, under CTFA each is left with a 50.00 profit. Thus, that portion of CTFA cost recovery due to differential unit cost subsidization may be legitimately viewed as "revenues for operations that were not collected due to originally incorrect pricing policies". That is, it can be regarded as a "unit profit correction scheme". And it is within this light that Teleglobe is afforded a means of treating CTFA cost recovery dollars as an output. If they were following the procedure of deriving price indices in order to deflate revenues, then CTFA would be added to all operational revenues prior to deflation. Under the present system, CTFA is deflated separately by the implicit price index derived from service operations output.

(ii) Transit Traffic Cost Recovery:

The direct connection of this component to operational revenues needs no further explanation. It is simply remuneration for the use of facilities to transit traffic originating and terminating outside of Teleglobe's jurisdiction. As a consequence, it is simply aggregated with the terminal traffic recovery portion and deflated by the implicit traffic price index.

g) Other Income, comprises the Interest Income from Teleglobe's investment in short term securities (mainly bank deposits), Intelsat Revenues, which are really a guaranteed return of 14% to Teleglobe as part owner of the International Satellite Network and Miscellaneous Revenues.

(i) Interest Income constitutes the revenues generated through the Corporation's maintenance of excess funds in cash and short term deposits. Since Teleglobe in effect sells the services of these funds, they must be considered as part of its overall output. Within this context, these funds can easily be converted to volume equivalents by deflation with the GNE price index and then weighted, in the TFP expression by the interest income generated. It should be noted that prior to deflation the value of cash and short term investments is averaged over two year periods. That is it becomes the sum of a year beginning, year end value, divided by two.

(ii) As for Intelsat revenues, they become the weight attributable to the volume of Teleglobe's investment in that organization. The satellite component of the Telecommunications Plant Price Index is used to convert Teleglobe's share of Intelsat from original to constant value. That is, the investment itself is considered as the output. This method is used, because unlike most of the other revenue categories, the Intelsat revenues cannot be directly associated to any of the existing output classifications such as telephone or telex minutes.

(iii) Finally, in keeping with the philosophy of including the real part of total revenues, all miscellaneous income is simply deflated by the PGNE and enters the output calculation as a constant value item. It is of almost insignificant size. Although it is essentially random in nature, making it both unforecastable and outside of management's scope of control, it does have a mean that hovers about a positive trend line, making part of the accounting structure with which the TFP measure attempts to maintain a close liaison.

The aggregator function for output is the Tornqvist discrete approximation to the continuous Divisia index. While it is true that the Divisia index (because it is a line integral and therefore path dependant) can, under certain conditions assume two different values at the same point, depending upon the path of approach, it should be kept in mind that their exists no useful atomistic index which does not have important shortcomings. The Laspeyre's index under certain conditions (of, say, great resistance to one particular brand of goods), will overestimate the size of either price or quantity changes (depending upon which one is explicitly measured.)

The main reason for choosing the Divisia index as Teleglobe's aggregator function is due to the fact that it is an approximation to an important flexible functional form, the translog fuction, which is itself a second order approximation to any linear homogeneous function. Thus the productivity measure will not be dependant upon the validity of any one particular production function specification.

B. Domestic Telecommunications

There are three large domestic telecommunications carriers that measure and use (mainly as a regulatory tool) Total Factor Productivity. These include Bell Canada (BC), British Columbia Telephone Company (BCT) and Alberta Government Telephones (AGT). The main distinction between BC and the other two (BCT and AGT) lies with the fact that the former measures output indirectly while the latter measure theirs directly. The categories, except for some services (to a fairly limited extent) which are unique to individual carrier, do not vary between firms. However, due to the different indexing methods as well as the direct vs indirect approach (explained above) the results display uncomfortable differences. BCT for example found that the average annual growth in quantity over the period 1966 to 1974, using the direct method and a chained Divisia index was 13% lower than that calculated indirectly through deflation by a non-chained Laspeyres price index. The price index average annual growth was greater (as would have been expected) by 166% over that derived directly.

The BCT categories, for telephone service, combining both the access and usage characteristics of a telecommunications network are as follows:[1]

1) Monthly Contract - Business Main
2) Monthly Contract - Business Extensions
3) Monthly Contract - Residence Main
4) Monthly Contract - Residence Extensions
5) Monthly Contract - PBX and Centrex
6) Service Connections
7) Local PL
8) PL Radio
9) Rent of Equipment
10) Other
11) WATS
12) Net Toll PL
13) Message Charges
14) Semi-Public Coin
15) Public Coin
16) Message Tolls - TC OPR
17) Message Tolls - TC DDD
18) Message Tolls - US OPR
19) Message Tolls - US DDD
20) Message Tolls - Alta OPR
21) Message Tolls - Alta DDD
22) Message Tolls - Intra OPR
23) Message Tolls - Intra DDD
24) Message Tolls - OVS via Montreal
25) Message Tolls - OVS via Vancouver

While 22 of these were observed directly as quantities, three required indirect estimation.

AGT has segmented its outputs into very specific categories which include access and usage as particular items. Because they recognize the advantage of directly observing quantity the following itemization of the AGT categories is meant to designate quantities:[2]

I. <u>Usage</u>
1) Local
2) Toll
3) Toll Operator
4) Other Special Services Messages

II. <u>Access</u>
1) Individual Access Lines
2) Multiple Access Lines
3) Public Access Lines
4) EFRC (AGT Edmonton Customers Only)
5) Private Lines
6) Mobile Radio Access

III.	Terminals	1) Total Telephones 2) Equipment Providing Multiple Access 3) Special Services 4) Special Assemblies 5) Mobile Radios 6) Miscellaneous Terminals
IV.	Other Special Services	
V.	Station Connection, Change and Moves	
VI.	Directory	1) White Pages 2) Yellow Pages 3) Directory Assistance
VII.	Other Services	1) Equipment Rentals 2) Other Rental Revenues 3) Custom Work 4) Other

Bell Canada uses the indirect method. Its output categories, as with BCT also reflect a mixture of usage and access. Price indexes, on a disaggregated basis, are developed and aggregated into seven major categories: one for local services;one for each of the message toll services, which include Intra Bell, Trans-Canada and Adjacent Members and United States and Overseas;one each for Other Toll, Directory Advertisement and Miscellaneous Revenues. These current dollar revenue amounts are then deflated and serve as constant dollar proxies for the output quantities required in the TFP measure.

As far as indexing,and consequently aggregation,is concerned, since AGT is still at the experimental stage in this area, we will only examine the BC and BCT cases. BCT uses the Tornqvist discrete approximation to a continuous Divisia index as its aggregator function. It assumes that output is priced at marginal cost (or at least that output prices and the respective marginal costs are identically distributed) and uses revenue shares as weights. The index is continuously chained and thus has weights that change annually. The implicit price indexes therefore measure the change in average revenue per unit of output, which takes account of both implicit (i.e., those due to a changing basked of goods) and explicit changes in the price paid by the consumer per unit of output. BC on the other hand, uses a fixed base Laspeyres index as its aggregator function for the component prices. The only changes captured by this index are those due to explicit tariff adjustments. However, from the firm's point of view, continuous chaining would seem to give a better indication of changing volumes. In addition, the final BC quantity index is difficult to interpret if the final result is in some sense supposed to represent the current quantity evaluated in base year prices, i.e. constant value = $\Sigma P_o X_t$ which does not result from deflation of current value b a Laspeyres price index where $\Sigma P_t X_t / (\Sigma P_t X_o / \Sigma P_o X_o) = (\Sigma P_t X_t \ \Sigma P_o X_o) / \Sigma P_t X_o \neq \Sigma P_o X_t$. In order to get the desired result, deflation must be effected by a Paasche price index or, alternatively the same result could have been obtained through inflation by a Laspeyres quantity index, i.e. by appropriate deflation we have $\Sigma P_t X_t / (\Sigma P_t X_t / \Sigma P_o X_t) = \Sigma P_o X_t$ or by appropriate inflation we have $(\Sigma P_o X_o)(\Sigma P_o X_t / \Sigma P_o X_o) = \Sigma P_o X_t$.

Finally, the difficulty with trying to obtain a good volume index under the possibly untenable assumption of an equivalence between marginal costs and output prices of course still remains. (Caves and Christensen, in a study of Canadian railroads, discussed below, have looked at possible solutions to this problem).

Before leaving this section on output measurement, it may be instructive to point out some of the difficulties with measuring output as the growth in its capacity rather than in terms of quantities actually demanded or sold. As Dhruarajan and Harris in their study of the Canadian airline industry point out, short term fluctuations in productivity that result from an understandable inability to match lumpy investment with short run factors causing deviations from some long term trend in demand, can, to some extent, be mitigated by using the growth in capacity as the relevant output measure. In the long run, however, management should be judged on their ability to match investments to demand and output sold therefore becomes the appropriate variable. Capacity measurement, while it may be difficult, is nevertheless feasible for the airline industry, which can find a more or less homogeneous definition of its output. For telecommunications, however, where there exist a far greater number of ways to combine a very hetergeneous set of products, capacity measurement would demand a far larger proportion of subjective input. Given that the potential of the system changes as the proportions of various outputs, such as toll, local, data, telegraph, etc., are varied, some subjective definition of composition would seem necessary before a consistent measure of capacity could be investigated.

III.7 Inputs

Total Factor Productivity gains, between any two periods, is generally understood to be measuring the differential rates of growth between total outputs and inputs, the latter being a weighted aggregate of the two primary input factors, Capital and Labour services and one intermediate input factor, Materials. The essential definition and measurement of these three basic components (except for such variation as is introduced as a result of examining gross value added, net value added, net output, etc.) are fairly similar from carrier to carrier. With Product Exhaustion as a basic constraint, once any two are defined, the third input is determined. That is, by beginning with the accounting identity that the total value of input is identically equal to the total value of output, once either one is established, and Labour and Materials, say, defined, the value of Capital Services becomes known. (It should be noted that this is only one possible view of product exhaustion. While Electricité de France, as it will be covered further on, also involves product exhaustion, at some stage in its calculations, it does not, as in the present case, assume short run optimal behaviour. It allows the existence of pure profits and losses, divergence between the cost and return to capital and under depreciation, thus making untenable the assumptions of marginal cost pricing and consequently invalidating the use of revenue shares as a proxy for cost elasticities). Beyond these basic definitions, it is the choice and use of an index number methodology that serves to introduce most of the difference into the input aspect of the productivity equation.

Beginning with product exhaustion we may write the basic accounting identify as:

$$PO = rK + wL + mM$$

Totally differentiating both sides and manipulating the terms we get;

$$P\Delta O - (r\Delta K + w\Delta L + m\Delta M) = (K\Delta r + L\Delta w + M\Delta m) - O\Delta P$$

which states that productivity gains calculated from either the physical changes (the left hand side) or from the price changes (the right hand side) must be equal as a necessary outcome of the product exhaustion assumption. Given their equivalence we will concentrate only on the physical side. Furthermore, having already discussed output measurement, we will limit ourselves to looking only at the input variables, of which there are six;

r = the cost of capital	$r\Delta K$ = Capital input (or service flow)	
ΔK = the changes in the capital stock		
w = the wage rate	$w\Delta L$ = Labour input	
ΔL = labour services changes		
m = the price of materials	$m\Delta M$ = Material input	
ΔM = material flows change		

Measurement of the changes in the physical quantities (ΔK, ΔL and ΔM) is fairly standardized across firms (with some difference between domestic and international telecommunications due to different classification practices). While the prices (r, w and m) are also similarly defined across firms, major differences arise as a result of index number choice. Let us first examine the physical changes.

Company accounts record <u>capital</u> data in original value terms. Thus, a transatlantic cable, put in service in 1960 at a total value of 20 million dollars, having undergone no measurable deterioration, will appear in the books at its original value of 20 million dollars in 1980. If, in 1980, a new cable with identical capacity (for simplicity sake), is put in service at a cost of 40 million dollars, then it would be highly misleading to say that the company has 60 million dollars worth of cable capacity in service. To see this, imagine that a physical measure of cables was being sought and that the value is divided by the price in order to get the quantity, then we would have 60/40 which would indicate a total of one and a half cables in services. Evidently accurate measurement of the capital stock would require, besides information on the original value of plant, vintage distributions and price indexes for physical plant as well. It is, in fact, through judicious use of this type of information that Canadian telecommunications carriers have been able to develop constant dollar proxies (through deflation by a Telecommunications Plant Price Index) for the physical volume of their plant in service.

Labour is measured in terms of man-hours worked in either weighted or un-weighted form. Classification, differing in minor ways between firms, is on the basis of management, operators, plant and craft, supervisory personnel and clerical. These, where the data is available (at least at Bell Canada) are further disaggregated by service age. Man-hours worked is essentially total paid man-hours adjusted for losses due to legal holidays, paid vacations, sickness, extraordinary leaves, lunch breaks and so on. In addition, the man-hours worked figure that appears as labour input in the productivity expression is further adjusted to pull out all those hours that have been capitalized and consequently, already included with the capital input.

Before continuing on to a discussion of the Materials input, it may be worth dwelling, for a moment, on the weighted-unweighted man-hour distinction. There appears to be a general misconception that the Bell Canada method of assigning weights to adjust for varying quality in different labour groupings is in some sense very different from that used elsewhere in the course of normal indexing procedures. It is fairly straightforward to demonstrate that in fact, the Bell Canada method uses nothing more complicated than a Laspeyres quantity index. They begin, in the base year with an overall, arithmetic average wage,

$$\bar{W}_o = \frac{\Sigma_i W_{io} L_{io}}{\Sigma_i L_{io}}$$

where W_{io} and L_{io} are the wage and number of man-hours worked, respectively for the ith labour classification. Next, a series of weights are derived

$$l_{io} = \frac{W_{io}}{\bar{W}_o}$$

from which the weighted man-hours worked (WMH_t) in each period are calculated as

$$\begin{aligned} WMH_t &= \Sigma_i \, l_{io} L_{it} \\ &= \frac{\Sigma W_{io} L_{it}}{\bar{W}_o} \\ &= \left(\frac{\Sigma W_{io} L_{it}}{\Sigma W_{io} L_{io}} \right) \Sigma L_{io} \end{aligned}$$

which, as can be seen is simply the growth of base year man-hours worked (which is identically equal to base year WMH_o) by a Laspeyres quantity index. Thus, the only difference between, say, the Bell Canada and Teleglobe Canada methodologies lies in the choice of index number.

Materials input includes all costs which are neither capital nor labour related. For the most part the PGNE (Price of Gross National Expenditures) is used to deflate the current value of Materials in order to derive a volume proxy. Only B.C. Tel. disaggregates this (materials) input into 10 distinct categories, deflating each with the appropriate price index. These categories include:

1) Material
2) Contract Labour
3) Vehicles and Tools
4) Rentals
5) House Services
6) Printing and Stationery
7) Travel and Transfer
8) Postage
9) Fuel and Utilities
10) Other

As far as prices, which are in any event essentially weights, are concerned, they are measured identically by all the carriers, but applied in two distinctly different ways: continuously updated and geometrically entered into the productivity equation as by Teleglobe Canada and B.C. Tel. (and perhaps AGT, taking their initial commitment to this method, in November 1977 as indication); unchanged base year value, applied through a Paasche quantity index to all past and future periods as by Bell Canada.

The cost of capital includes the sum of depreciation (when dealing with a gross TFP measure), property taxes and the cost of money which includes fixed charges, preferred dividend appropriation, net income available for common and income taxes (payable). Bell Canada will measure this amount in the base year (which is 1967) of their study and then calculate its proportion vis-à-vis the 1967 value of their capital stock, i.e.:

$$r_{67} = \frac{\text{Cost of Capital}_{67}}{\text{Value of Capital Stock}_{67}} = \frac{r_{67}K_{67}}{K_{67}}$$

where $K_{67} = \Sigma_i \ TPPI_{i,\ 67} \ K_{i,\ 67}$

TPPI = (the Telecommunications Plant Price Index)

and, as per product exhaustion, the absolute cost of capital is determined as the residual

$$r_{67}K_{67} = P_{67}O_{67} - W_{67}L_{67} - m_{67}M_{67}$$

It is this rate that is then applied to derive a physical proxy for the flow of capital services in any particular year. Although Teleglobe Canada (as well as B.C. Tel. and AGT) also involve product exhaustion, using the method of the residual, there are some differences. First of all, it is the value of materials that constitutes the residual, so that, in every year

$$m_tM_t = P_tO_t - w_tW_t - r_tK_t$$

Secondly, because $r_t K_t$ is directly observed in every year, the return to capital varies continually as opposed to its fixed value of r_{67}, as in the Bell Canada case. This implies that while the growth in the physical flow of capital services is exactly proportional to the growth of the capital stock, in the Teleglobe method it enters the productivity calculations with a variable weight. This can be seen when comparing the two different formulations, first for Bell Canada:

$$\text{TFP gain} = \frac{\dfrac{\text{Output}_t}{\text{Output}_{t-1}}}{\dfrac{(r_{67}\ \text{Capital Stock}_t + w_{67}\ \text{Labour}_t + m_{67}\ \text{Materials}_t)}{(r_{67}\ \text{Capital Stock}_{t-1} + w_{67}\text{Labour}_{t-1} + m_{67}\ \text{Materials}_{t-1})}}$$

$$= \frac{\dfrac{\sum_i P_{i,67} O_{i,t}}{\sum_i P_{i,67} O_{i,t-1}}}{\dfrac{\sum_i R_{i,67} I_{i,t}}{\sum_i R_{i,67} I_{i,t-1}}}$$

and then the Teleglobe/B.C. Tel. method:

$$\text{TF}\dot{\text{P}}\text{ gain} = \frac{\dfrac{\text{Output}_t}{\text{Output}_{t-1}}}{\left[\dfrac{\text{Capital Stock}_t}{\text{Capital Stock}_{t-1}}\right]^{r^*_{jt}} \left[\dfrac{\text{Labour}_t}{\text{Labour}_{t-1}}\right]^{w^*_{jt}} \cdot \left[\dfrac{\text{Materials}_t}{\text{Materials}_{t-1}}\right]^{m^*_{jt}}}$$

$$= \frac{\prod_i \left[\dfrac{O_{it}}{O_{i,t-1}}\right]^{P^*_{it}}}{\prod_j \left[\dfrac{I_{jt}}{I_{j,t-1}}\right]^{R^*_{it}}}$$

where the * denotes average share in revenues or costs, as the case may be, over the two periods (t) and (t-1).

III.8 Productivity Measurement in Regulated Non-Telecommunications Industries

Since the following studies are presented more in the way of a note to indicate that work in the area of TFP measurement is taking place in regulated sectors outside of telecommunications, all the aspects of TFP measurement, including output, input indexing and results for each of them will be completely covered within the confines of this section, highlighting only those points which can be legitimately related to similar work in telecommunications. We will examine four service industry studies. These include:

a) Caves, D.W. and L.R. Christensen, Productivity in Canadian Railroads, 1956-1976, Canadian Transport Commission, Report No. 10-78-16, Aug. 1978.

b) Dhruvarajan, P.S. and R.F. Harris, A Productivity Study of the Canadian Airline Industry, Canadian Transport Commission, Report No. 10-78-03, March, 1978.

c) Cairns, M. and B. Kirk ., Canadian for Hire Trucking and the Effects of Regulation; A Cost Structure Analysis, Canadian Transport Commission, Preliminary, November, 1979.

d) The Total Factor Productivity Measurement System at Electricité de France.

The first three, each deal, in some sense, with comparative productivity studies. The problems that face each of them span the entire gamut of output definition of a hetergeneous product, sustenance of certain key assumptions (such as marginal cost pricing), interfirm and interegional comparison problems and index number choice. Caves and Christensen introduce an interesting way of using proxy cost shares for the weighting of output. Dhruvarajan and Harris demonstrate an insightful method of accounting for the short run phenomenon of lumpy investment and its consequent contribution to a volatile TFP growth. The intricate empirical work of Cairns and Kirk,. in their attempt to appropriately define an enormously heterogenous industry and its equally vast set of differentiated outputs has very important potential for the intricacies of dealing with the study of comparative productivity in telecommunications. The final study, productivity at Electricité de France, provides an extremely important view of a methodology, which departs in one important respect from those currently used in Canadian Telecommunications, through its independant measurement of inputs and outputs without recourse to tradional notions of product exhaustion. Although the final accounting identity whereby the total values of input and output are identically equal, is ultimately satisfied by adding various profit and loss balancing items, their methodology admits the possibility of short run digressions from optimizing behaviour. Residually derived product exhaustion, on the other hand, does not.

Finally, it should be noted that the study of Total Factor Productivity has become fairly widespread, encompassing a far larger effort than only the four mentioned above. The American Productivity Centre[3] studies TFP, among many different variations of productivity being examined and has well over 100 members using its output. These cover the entire range of productive activity including consumer goods and service firms, capital goods manufacturers, insurance companies and so on. As well, the electric utilities in both North America and Europe have, for quite some time, been measuring and applying TFP.

Productivity in Canadian Railroads, 1956-1976
[D.W. Caves and L.R. Christensen]

The study attempts to measure the total factor productivity for (1) CN Rail, (2) CP Rail, (3) combined CN and CP and (4) CN and CP compared. Beginning with a multi output implicit production function, deriving the dual cost function and then invoking Sheppard's Lemma (whereby the cost minimizing input levels are equal to the first partial derivitives of the cost function with respect to its input price arguments), the authors derive the continuous Divisia index of productiviey whereby shifts in the cost function are represented as the difference between continuous Divisia indexes of outputs and inputs. The discrete approximation to this index is the same one used by Teleglobe Canada, B.C. Tel. and others. It can be represented as:

$$\text{TFP gain} = \left[\frac{O_t}{O_{t-1}}\right]^{1/2\,(P_{it} + P_{i,t-1})} - \left[\frac{I_t}{I_{t-1}}\right]^{1/2\,(R_{it} - R_{it-1})}$$

(the notation has been changed to conform with our text), where the P it are revenue shares at certain levels of output aggregation and cost elasticities with respect to output (as derived from the cost function) at others. The R it always refer to cost shares. While the standard Divisia approach of using revenue shares is applied to derive the two major output categories, passenger miles and freight (ton) miles (aggregated from more than 20 sub-categories) the authors felt that the implicit assumption (which allows revenue shares as weights) of prices reflecting marginal costs was untenable for the final aggregation of these two broad categories into total output. Instead, they used indirectly estimated cost elasticities (with respect to output), normalized to sum to one, as weights. The coefficients were obtained from cross-section studies of American railroads and then combined with the Canadian data, within a translog specification of the cost structure to calculate "second-hand" elasticities. Given the somewhat restrictive technological assumptions that are implicit in such a procedure, it is not at all clear that this is superior to simply using revenue weights, a point only partially admitted by the authors. However, barring these potential difficulties, this method does offer some interesting possibilities for similar studies in Canadian telecommunications where not all firms have the volume of data required for the estimation of flexible functional forms. The coefficients of a data rich firm may possibly be combined with the data of firms with shorter histories to add richness to any analysis of productivity.

Inputs are categorized along fairly standard lines, between capital, labour and materials as a residual. They are combined, at all levels, using the standard Tornqvist approximation to the continuous Divisia index, with cost shares as weights.

Comparisons are viewed on two levels, in terms of productivity growth rates and productivity levels. The growth rates are compared directly while for level comparisons, additional, and rather interesting further adjustments are applied. The firms are adjusted for size by reinterpreting the time subscripts in the indexing equations as referring to different companies. Clearly, when more than two firms are involved, comparisons, using this method, will only be possible in terms of each individual firm vis-a-vis some overall, 'all-firm' average.

Finally, the authors are quick to point out that due to the inability of their model to capture various intangibles, such as favourable environments with respect to existing networks at any point in time, their results "should be interpreted as productivity comparisons and not as comparisons of economic efficiency."

A Productivity Study of the Canadian Airline Industry
[P.S. Dhruvarajan and R. F. Harris]

The authors faced problems fairly similar to those present in any study of an industry with multiproduct firms. Outputs and inputs had to be both defined and aggregated with a view to not only examining individual firm productivity but to reaching meaningful interfirm comparisons as well.

As opposed to the Christensen study, Dhruvarajan and Harris used a Divisia index with revenue weights at all levels of aggregation for both inputs and outputs. They assumed prices and marginal costs to be similarly distributed. They did, however, make an important distinction between available (or capacity) output and revenue (or actually sold) output. This allowed them to distinguish between short run and medium to long term productivity movements. In the short run, due to indivisibilities in capital investment, increases in available capacity, as the measure of output, would avoid unjust penalties. In the medium to long term, however, management should have been able to marry investments to marketing and operating decisions with a view to maximizing productivity growth, making the appropriate measure 'output actually sold' (or revenue output).

While labour, materials and fuel are treated in the usual way, the authors have, interestingly enough, directly measured capital in physical terms contrary to the indirect deflation method employed in most other studies dealing with heterogeneous equipment. They essentially use two (and occasionally a third, consisting of number of aircraft) physical flow measures to represent the stock of capital - which is actually the inverse of the standard procedure, whereby the flow is assumed proportional to the growth in the stock - (1) the unweighted version, number of hours flown and (2) the weighted version, number of hours flown adjusted for size of aircarft. This ignoring of all other capital equipment requires two assumptions: (a) that the ratio of all other equipment in relation to these flows remains constant over time and/or (b) that if it is not constant then the large proportion of total capital accounted for by aircraft alone will in any event greatly reduce the significance of fluctuations in other equipment.

This distinction leads Dhruvarajan and Harris to some interesting analytical results. The difference in TFP, between using the unweighted and weighted versions (where the latter is lower) can now be directly attributed to "changes in aircraft size and quality". Productivity, apart from the airlines own contribution to input conservation derives partly from the manufacturer's contribution of a higher quality input. Nevertheless, the ultimate responsibility does rest with management.

Another interesting aspect is the complete ignoring of depreciation and the use of gross capital. The crucial assumption, which the authors claim has been born out by various studies, is that service capacity does not significantly deteriorate with age. This certainly avoids two major problems in the measurement of capital input: (a) its application as a volume concept particularly to the authors' specific definition of capital and (b) the derivation of vintage distributions become irrelevant. This is of particular relevance to a service industry such as telecommunications, which, it can be argued, also uses equipment which does not lose its service capability with age. The survivor curve for most equipment resembling that of buildings, almost perfectly horizontal to its replacement age.

As far as intercarrier comparisons are concerned, one wonders why, with all the disclaimers questioning their value, the authors even bothered.

> "The value of these comparisons is very much limited by the differences in the markets served by the various carriers.
> ...Again it must be stressed that many of these differences [in productivity] are caused by variations in operating environments rather than by basic differences in efficiency."

What comparisons are made are effected on the basis of partial output to partial input measures of absolute values such as Available Seat Miles per Employee. Besides having the usual drawback of any partial productivity measure comparison, as the authors duly note, they are further hamstrung by the fact that some of the carriers (CP in particular) began the study period with lower load factors than others (such as Air Canada) with very high load factors).

<u>Canadian For Hire Trucking and the Effects of Regulation: A Cost Structure Analysis (Preliminary Report)</u>
[M. Cairns and B. Kirk]

The basic goal of this study is to examine the effects of regulation on the cost structure of the trucking industry. It does not use the traditional accounting approach, employed by the other studies that have been examined, productivity is rather analyzed indirectly through the econometric estimation of single and multiple output translogarithmic cost functions. The most relevant aspects, however, concern the manner in which Cairns and Kirk treated the problems of industry segmentation (whereby the business activity of different firms were too diverse for legitimate comparison or aggregation), heterogeneous output and the difficulties associated with interfirm comparisons operating in different regulatory and demand environments.

The first step was to investigate segmentation and thus choose that group of carriers considered appropriately homogeneous for investigation. Of a potential 2756 carriers reporting information in 1975 (the study year), 2538 were eliminated as a result of inadequate or questionable information reported in the various surveys and also due to the fact of having a group of specialized contract carriers as a recognizable industry segment. Through a further examination of output content, the group was narrowed down to 178 carriers whose main business was Canadian intercity transport. Further natural segmentation of this last group was investigated, empirically, by carrying out a principal component analysis on selected financial and traffic related operating characteristics. The variables were selected in order to differentiate the activities of each truck carrier with respect to such operating features as: size and profitability; extent of pick-up, delivery and terminal operations; quantity of equipment leasing;type of carriage provided and so on. Except for some clustering in Alberta, due to large terminal operations, no further segmentation was detected, allowing these firms to be used as the individual unit of investigation within an empirical cost function.

The methods used by Cairns and Kirk to measure output should be very instructive for studies in telecommunications. The similarity to trucking lies in the enormous diversity of the output. Like telecom., interest is focussed on distance (kilometrage bands), handling (operators), type of commodity (data or voice), origin and destination, density and so on. After elaboration, of the above, Cairns and Kirk identify 17,280 discrete traffic characteristics, which, on the basis of similar marginal costs (looking only at Alberta which would not have regulation distorted cost curves), were partitioned into 1,246 different characteristics. These were then classified as belonging to either one of four possible output denominations (all estimated from sampled shipments), including number of shipments, tons, miles and ton miles. The final choice was narrowed to: (1) three output measures, for the multiple output cost function, denominated as shipment miles and including less than truckload, truckload short haul and truckload long haul and (2) a collective output measure, for the single ouput cost function, denominated in ton miles.

There are three interesting points, which we may note, concerning the rather elaborate selection procedures. Firstly, although most other studies have, in any event, chosen ton-miles as the appropriate output measure, Cairns and Kirk have empirically demonstrated its validity. The second point, of direct interest to our study, lies in the potential application of these selection procedures towards a manageable and acceptable set of telecommunications output definitions. And finally, the authors chose to define and measure output only from a supply point of view, demand considerations being ignored.

The rest of the study examines the cost structure of the industry through the estimation of single and multiple output cost functions under a translog specification. While this type of analysis is certainly of interest in and of itself, we will only note the conclusions which bear on some longstanding controversies in telecommunications. However, before quoting these, verbatim, it should be noted that the estimation procedures were able to account for the differences in utilization, as well as (viz. the selection procedures discussed above) those due to operating environment and output distribution. The conclusion then, are:

> "Overall therefore the differences in traffic mix between Alberta and Quebec and Ontario [the two regulated provinces] were closely related to the differential rates of capacity utilization and it was these differential rates of utilization that accounted for a significant proportion of the observed differences in marginal costs. These conclusions ... suggest that differences in unit costs between Alberta and Quebec and Ontario may be largely manifestions of different demand conditions for transport service.
>
> Lastly, it is open to question whether the residual differences in marginal costs between provinces, having accounted for differences in selected traffic characteristics and utilization, were a result of still unaccounted differences in traffic characteristics or whether they were the result of the effects of economic regulation. Given the demonstrated impact of traffic differences on unit cost, unaccounted traffic differences such as type of commodity must remain a strong candidate for the source of residual differences in unit costs".

Electricite de France

The productivity measurement technique known as UNIPEDE (Union Internationale des Producteurs et Distributeurs d'Energie Electrique), which is used by most of the European electric utilities as well as, with some minor variations, by all the large French government owned enterprises, was first developed by C.E.R.C. (Centre d'Etudes des Revenus et des Coûts) and EDF (Electricité de France). The following discussion, although relying directly on EDF sources and examples, is, therefore, representative of a fairly widely used methodology.

Apart from details such as calculating quantity indices directly or indirectly through deflation by prices indices, the EDF methodology differs from the Teleglobe, Bell, B.C. Telephone and others in one major respect, EDF does not invoke the product exhaustion condition for calculations of productivity growth. That is, EDF has no mechanism whereby the total value of output is constrained to equal the total value of all productive inputs. At Bell Canada, for example, the absolute return to capital, in the base year, is calculated as a residual after observing the values of labour and material inputs. Thus,

$$P_o O_o = r_o k_o + W_o L_o + m_o M_o$$

where $P_o O_o$, $W_o L_o$ and $m_o M_o$ are known, therefore constraining

$$r_o k_o = P_o O_o - W_o L_o - m_o M_o$$

Product, in all subsequent years is as well assumed to be exhausted. At Teleglobe Canada, it is the current value of materials that plays the role of "constraining residual" and,

$$m_t M_t = P_t O_t - r_t K_t - W_t L_t$$

To avoid confusion it should be noted that "t" has replaced "o" as the subscript for Teleglobe only because of differences in indexing techniques and that the idea of a constraining residual is identical.

While the use of a constraining residual has certain theoretical merit, practically, it may not be entirely justified. This issue will be considered below. First of all, let us examine, more closely, the EDF methodology. Inputs and outputs are calculated independantly and differences between their current values originate from three sources:

1) <u>Profits/Losses</u>:

The independant tabulation of observed revenues and costs, known as "effective" values or rates (as the case may be), will lead to complete equality, after accounting for returns to capital labour and materials, only by coincidence. The final "pure" profit or loss ("resultant" in the literature) is a first source of difference. This may be described as:

$$\Sigma_i P_{it}O_{it} = \Sigma_j R_{jt}I_{jt} + PL_t$$

where PL_t is the profit or loss.

2) Under Depreciation:

With a positive rate of inflation at r%/year, the replacement value of a price of equipment at age T with original cost in t-T of $k_{t-T}{}^T$ (and assuming no technical change over the period from (t-T) to t) is:

$$\text{Replacement Value } (RV_t) = k_{t-T} \left(1 + \frac{r}{100} \right)^T$$

However, in normal accounting procedure, when determining profits subject to tax, depreciation is calculated on an original cost basis. Clearly, if d_t is the depreciation rate, then

$$d_t k_{t-T} < d_t RV_t \qquad \text{when } r > 0$$

and at EDF, for their TFP study, $d_t RV_t$ is used, thus giving rise to

$$d_t(RV_t - k_{t-T}) = \text{under depreciation } (UD_t)$$

3) EPCF (Ecart Provenant des Conditions de Financement):

EDF draws on two sources for its long term financing needs, internally generated funds and debt. The latter is a combination of direct low interest government loans and bond market activities. Within an accounting framework charges are inputed to capital on the basis of observed (or effective) costs equal to zero for internal funds (except for interest charges to ongoing construction activity), the low interest paid on direct government loans and the normal bond market rates for the remainder. For TFP purposes, however, a competitive rate (on the basis of hypothetical leasing arrangements) is applied to the value of physical plant in service (as determined above in studies of under depreciation) which yields a cost of capital which will normally differ from the accounting results. In the latter case,

$$r_t L_t = \Sigma W_i r_i L_i \qquad 1 = 0 \ldots\ldots T$$

where W_i is the weight and r_i the applicable rate for the i^{th} type of liability, L_i. Thus, if internally generated funds constituted the j^{th} liability, then $r_j = 0$. The theoretical rate, on the other hand is calculated by applying a competitive rate, r_t, to revalued physical plant in service, RV_t and

$$EPCF = r_t^* RV_t - r_t L_t$$

The final accounting equation would then be

$$\Sigma_i \; P_{it} O_{it} = \Sigma_j \; R^*_{jt} \, I^*_{jt} - UD_t - EPCF_t + PL_t$$

where: $\Sigma_j R^*_{jt} I^*_{jt} = W_t L_t + (r_t L_t + EPCF_t) + (D_t + UD_t) + m_t M_t$

TFP is measured on the basis of $\Sigma_i P_{it} O_{it}$ for output and $\Sigma_j R^*_{jt} I^*_{jt}$ for input, and output and input values, as can be seen, are unlikely to be equal. The implication of this inequality can be examined in the following manner. The average cost function for any firm,

$$AC = \frac{\Sigma r_i X_i}{\Sigma W_i Y_i} = \frac{\Sigma r_i X_i}{O}$$

where O is some index of aggregate output volume and $\Sigma r_j X_j$ is the total value of inputs and can also be expressed as the product of aggregate input price and quantity indexes, RI. Then the AC index

$$AC = \frac{RI}{O} = \frac{R}{O/I}$$

when O/I is easily recognized as the standard TFP expression, we can therefore write:

$$TFP = O/I = R/AC$$

From this we can state the following rule: If maximum profits are equal to zero and in addition we have output prices equal to both marginal and average cost, TFP results become invariant with respect to measurement through prices or quantities, i.e.,

$$TFP = O/I = R/P \quad \text{where } P = MC = AC \text{ and } \Pi = 0$$

As can be seen, product exhaustion, is a necessary condition for the above equality to hold. More importantly, however, if we wish to avoid invoking the restrictive Euler theorem and homogeneity condition, we note that for the marginal productivity theory of distribution to hold the crucial assumption is zero maximum profits. Any firm that meets this condition and, as well, behaves accordingly , will have prices equal to marginal cost. Thus, from

$$\Pi = PO - \Sigma_j r_j X_j = 0$$

and the normal first order conditions for profit maximization, where

$$r_j = P \frac{\partial O}{\partial X_j} \text{, we get}$$

$$PO = \Sigma_j P \frac{\partial O}{\partial X_j} X_j$$

and with marginal cost pricing were $P_i = \frac{\partial C}{\partial Y_i}$ or $P = \frac{\partial C}{\partial O}$, we get,

$$\frac{\partial C}{\partial O} O = \Sigma_j \frac{\partial C}{\partial X_j} X_j$$

However, if prices, due to short term deviations in competitive conditions, are less than perfect competition marginal costs (as in the case of favourable financial conditions which may allow for lower prices), then

$$PO < \Sigma_j \frac{\partial C}{\partial O} \frac{\partial O}{\partial X_j} X_j$$

and calculations of productivity, whereby the weights for output are based on observed prices and the weights for inputs on theoretical costs, will be misleading. This can be seen from

$$O/I = R/AC < R/P \quad \text{if } P < AC = MC$$

where $O = \Sigma \frac{P_i Y_i}{\Sigma P_i Y_i} Y_i$ which is the weighted index of aggregate output with observed revenue proportions as weights. But, because

$$P_i < \frac{\partial C}{\partial Y_i} \text{ for some i and } PO < C = \Sigma_j r_j X_j = RI \text{ then}$$

$$\frac{P_i Y_i}{\Sigma P_i Y_i} \neq \frac{\partial C}{\partial Y_i} \frac{Y_i}{C}$$ for some i (where $\frac{\partial C}{\partial Y_i} \frac{Y_i}{C}$ is the cost elasticity of the ith output) which implies that the value of measured TFP is different from the theoretical value. While this problem is reconciled in the Teleglobe and Bell studies by forcing the product exhaustion condition through the use of a residual, it is ignored at EDF. There are two ways of resolving this issue. Either theoretical

cost elasticities can be found or that the variations in financial, depreciation and profit/loss results are assumed to have a distribution identical to that of the effective results. Although the first possibility is probably unmanagable and the second unlikely the practical advantage of the EDF method may be such as to allow its theoretical drawbacks to be ignored. These, as are discussed in the "Uses" section, pertain to the insights that management can gain by examining the impact of varying financial conditions on the firm's overall performance and behaviour.

COMPARATIVE EFFICIENCY IN CANADIAN TELECOMMUNICATIONS: AN ANALYSIS OF METHODS AND USES

PART II: Uses and Indexes

Phase II: Productivity, Employment and Technical Change in Canadian Telecommunications

Michael Denny
Institute for Policy Analysis
University of Toronto

Alain de Fontenay
Department of Communications
Ottawa

Manuel Werner, Consultant

Final Report for Department of Communications
(03SU. 36100-9-9527-DSS)

Table of Contents

* These sections should be read by the non-specialist

* These sections should be read by the non-specialist

IV.1 Introduction

One major purpose of this report is the development of methods for measuring and comparing productivity in telecommunications firms. The other primary goal is the evaluation of possible uses for productivity measures. Increases in productivity are the foundations of the growth in real wealth. Consequently the interpretation of a company's performance by both management and the regulator is enhanced by the measurement of productivity.

In this section, several possible and actual uses for productivity measures are discussed. Even before any productivity calculations are attempted, the existence of an improved set of data about the firm's activities is a benefit. Existing accounting conventions are slowly shifting to provide useful economic data for management decision-making. The current difficulties with "inflation" accounting is an excellent example. While accountants are slowly reorienting their conventions towards information systems for decision-making and away from record-keeping, progress has been quite slow. The data generated for a productivity calculation should (or could) be the data used in demand studies, cost, studies, manpower planning, investment decisions and almost any other area of management decision-making. This aspect of the project has not been emphasized in this report but will become clearer during the next phase.

We have divided our discussion into two major sections. In the first, the actual and potential uses of productivity measurement for managerial purposes are discussed. Two examples from the area of Planning and Control are developed below. Both of them pertain to the integration of productivity and profits. At th foundation of any progress that a firm is able to attain lies the increases in productivity. These gains can be distributed in a number of internal and external ways. In the developments below this distribution will be carefully shown.

For the regulator, productivity measures have a potentially larger role. Our work is incomplete in this area but the direction is clear. Efficiency in production is a goal that regulators must not ignore. Careful monitoring of productivity changes are required to evaluate the efficiency of the firm. More particularly, efficiency may be partially lost through the pursuit of other goals. Since these latter goals may be important, the regulator must carefully evaluate the costs and tradeoffs of the various goals that it may wish to pursue.

IV.2 Management Control and Planning

Although the methodologies developed by various firms for the analysis of productivity gains are essentially similar, given that Electricité de France (EDF) has had their system in operation and practically applied for a number of years now, most of the following will draw heavily on their sources. Other companies either using or developing these type of analyses include Teleglobe Canada, AT&T and many European electric utilities.

A. Distribution of Gains (Sources and Uses)

Within the context of a multi-input, multi-output firm, the prices at which output is sold is in some sense closely related to the cost of production. If we were to define the cost of producing one unit of aggregate output as

$$AC = \frac{RI}{O}$$

where R, I and O are indices of aggregate input price, input volume and output volume respectively, with R including the required return to capital under any given competitive situation (whether it be monopolistic, perfect, etc.), then we would expect AC to also equal P, the aggregate price of output. Furthermore, since

$$AC = \frac{R}{O/I} = \frac{R}{TFP}$$

then,

$$d\ln AC = d\ln P = d\ln R - d\ln TFP$$

or the proportional change in prices would have to equal that part of the proportional increase in input prices that was not offset by proportional increases in productivity gains. Within this context, the importance of not only monitoring, but as well, explicitly planning for gains in productivity, is clear. The ultimate goal will differ with the situation. In a highly competitive situation price increases are anathema and a heavier reliance must be placed on productivity gains. In a closely regulated industry, some proof of reasonable productivity performance may be required in order to justify price changes and in the case of EDF, productivity measurement and analysis are absolutely essential in order to ensure that government guidelines are being followed.

There are two methods, in current use, through which the absolute value of productivity, price and volume gains are analysed. One uses the method of the residual whereby the total value of output is always constrained to equal the total value of input, thus implying that the firm always behaves in an optimal fashion, and the other does not use the equality constraint thus permitting pure profits and losses and allowing for short term sub-optimal behaviour. Although, as we have seen, there are some theoretical drawbacks to the latter method, as employed by EDF, we will draw upon their analyses as illustrative of management uses.

Before continuing, we should note the advantages of using TFP as a management tool. First of all, it is not meant to replace the traditional financial measures of management success but only to complement them at certain weak points. TFP measures can sometimes identify abnormalities faster than classical accounting methods, above all, by demonstrating that it is often feasible to obtain a better return without relying on prices. In the long run, the profitability of an enterprise is, in large measure, a function of its relative productivity performance, although short term fluctuations may violate this relationship.

Distribution of Gains Analysis:

Distribution of gains comprises an analysis very much akin to the "Sources and Uses" exercise found in the annual reports and financial statements of most firms. The difference lies with the breakdowns, as well as the detail. While the financial statement, on the one hand, deals only with current dollars, our analysis (henceforth referred to as DG) examines the distribution in terms of separate price and quantity effects, which of course include TFP, and on the other hand, whereas the financial statement is only concerned with the overall sources and uses, DG carefully examines their composition. The potential benefits of such an analysis will become evident through the detailed example presented below.

The basic equations of DG can be derived by decomposing the periodic changes in revenues and cost into their essential elements.

$$\Delta R = P_1 O_1 - P_0 O_0$$

where $P_0 = (PGNE_1 / PGNE_0)\, P_0^*$ such that the unit price of output O in the previous period, P_0^*, is adjusted by the price of gross national expenditures, PGNE, in order to ensure that price comparisons reflect equivalent purchasing power. It should be noted that a purchasing power adjustment will have no effect on the results of TFP gain measurement, since we are dealing with a ratio $d\ln P_oO_o / d\ln R_oI_o$ where both the numerator and denominator undergo identical adjustments, thus cancelling each other out. R. is equal to total revenues. From,

$$P_1O_1 = (P_o + \Delta P)(O_o + \Delta O)$$

we can rewrite the change in revenues as

$$\Delta R = (P_o + \Delta P)(O_o + \Delta O) - P_oO_o$$

$$\Delta R = O_1 (P_1 - P_o) + P_o (O_1 - O_o)$$

and by adding and substracting $R_o(I_1 - I_o)$, the input volume change,

$$\Delta R = \underbrace{O_1(P_1 - P_o)}_{\text{Price Gains}} + \underbrace{\left[P_o(O_1 - O_o) - R_o(I_1 - I_o)\right]}_{\text{TFP Gains}} + \underbrace{R_o(I_1 - I_o)}_{\text{Volume Gains}}$$

It should be noted that had the above process been repeated beginning with a decomposition of $P_oO_o = (P_1 - \Delta P)(O_1 - \Delta O)$, in order to denominate quantity changes in terms of constant value with current prices, the final results would not have changed.

A similar procedure is followed for the decomposition of costs. Define the change in total costs as:

$$\Delta C = R_1 I_1 - R_o I_o$$

where R and I are, respectively, the prices and quantities of inputs. However, these inputs have both theoretical and effective components. These classifications apply to the cost of capital, including depreciation. Theoretical refers to the real cost of the input while effective denotes its actual cost to the firm. Thus, for the firm, the effective cost of capital, excluding depreciation is equal to some weighted average of its debt and equity expenses, including, in some cases, some fixed charge imputed to its own invested internally generated funds. The real price, however, when calculated, say, on a rental cost basis, with the rental rate applied to a capital stock value adjusted for replacement (or inflation), may be different. At EDF the difference is known as EPCF (écart provenant des conditions de financement). Similarly, theoretical depreciation calculated on an original value basis will differ from that calculated on a replacement base. This difference is known as under-depreciation. It should be noted that the theoretical rates are always applied to constant values. Thus, if we define $C = \sum_{i=1}^{m} R_i I_i$ and the cost of capital (excluding depreciation) and depreciation as the s^{th} and t^{th} components, respectively, where s and t occupy the $(m-1)^{th}$ and m^{th} positions in the vector of inputs, then

$$C = \sum_{i=1}^{m-2} R_i I_i + R_s I_s - R_s' I_s' + R_t I_t - R_t' I_t'$$

where $(R_s I_s - R'_s I'_s)$ = effective cost of capital

and $(R_t I_t - R'_t I'_t)$ = effective depreciation

and $\Delta R_s = \Delta R_t = 0$ (i.e. theoretical prices are donstant)

Then, if we let $\sum_{i=1}^{m-2} R_i I_i = R^* I^*$

$$\Delta C = R_o^* (I_1^* - I_o^*) + R_{os}(I_{1s} - I_{os}) + R'_{os} (I'_{os} - I'_{1s}) + R_{ot} (I_{1t} - I_{ot})$$

$$+ R_{ot}' (I'_{ot} - I'_{1t}) + I_1^* (R_1^* - R_o^*) + I_{1s}(R_{1s} - R_{os}) + I'_{1s}(R'_{os} - R'_{1s}) +$$

$$I_{1t}(R_{1t} - R_{ot}) + I_{1t}'(R_{ot}' - R_{1t}')$$

where $R'_{oj} = R_{ij}$; $i = 0,1$ and $j = s, t$ by definition of base year values. It should be noted that $R_{oi}(I_{1i} - I_{oi}) + R'_{oi} (I'_{oi} - I'_{1i})$; $i = s, t$ are the differences in "effective" volumes. In the long run, assuming that RI also contains the required return to capital, pure profits or losses may in fact be equal to zero, thus assuring all the desirable properties of the marginal productivity theory of distribution, such as marginal cost pricing, etc. Annual calculations, with which we will be dealing, will always have some $e \neq 0$ where $\Delta e = \Delta R - \Delta C$. This, however, should not pose any serious problems for measurement, where we assume marginal cost pricing and revenue shares as mirroring the distribution of cost elasticities, particularly if we suppose that the e_t are purely random with a zero mean. For management, however, the pure profit and loss item is of crucial importance and is a result explicitly considered in our analysis. Thus from,

$$e = \Delta R - \Delta C$$

we can substitute for ΔR and ΔC to get:

$$\text{(1)}\quad \underbrace{\overset{\substack{P/L\\+}}{\Delta e'} - \overset{\substack{\text{Output Price Reductions}\\-}}{\sum_{i=1}^{\ell} O_{1i}(P_{1i} - P_{oi})} - \overset{\substack{\text{Prod. Losses}\\-}}{T\dot{F}P'} + \overset{\substack{\text{Input Price Increases}\\+}}{\sum_{j=1}^{g} I_{1j}(R_{1j} - R_{oj})}}_{\text{uses}} =$$

$$\underbrace{\overset{\substack{P/L\\-}}{-\Delta e''} + \overset{\substack{\text{Output Price Increases}\\+}}{\sum_{i=l-1}^{n} O_{1i}(P_{1i} - P_{oi})} + \overset{\substack{\text{Productivity}\\+}}{T\dot{F}P''} - \overset{\substack{\text{Input Price Reductions}\\-}}{\sum_{j=g-1}^{m} I_{ij}(R_{ij} - R_{oj})}}_{\text{sources}}$$

where $\Delta e' = o; \Delta e'' = \Delta e$ when $\Delta e < 0$

$\Delta e'' = o; \Delta e' = \Delta e$ otherwise

$$O_{1i}(P_{1i} - P_{oi}) < 0 \qquad \forall i;\ i = 1 \ldots\ldots\ldots \ell$$

$$O_{1i}(P_{1i} - P_{oi}) \geq 0 \qquad \forall i;\ i = \ell+1 \ldots\ldots\ldots n$$

$$\dot{TFP}' = 0; \quad \dot{TFP}'' = \dot{TFP} \quad \text{when } \dot{TFP} > 0$$

$$\dot{TFP}'' = 0; \quad \dot{TFP}' = \dot{TFP} \quad \text{otherwise}$$

$$I_{1j} \, (R_{1j} - R_{oj}) > 0 \qquad \forall j;\ j = 1 \ldots\ldots\ldots g$$

$$I_{1j} \, (R_{1j} - R_{oj}) \leq 0 \qquad \forall j;\ j = g+1 \ldots\ldots\ldots m$$

Furthermore:

$$\sum_{j=1}^{m} I_{1j} \, (R_{1j} - R_{oj}) = \left[\sum_{j=1}^{m-2} (I^*_{1j} - R^*_{oj}) \right] - \left[I'_{1s} (R'_{os} - R'_{1s}) + I'_{1t} \, (R'_{ot} - R'_{1t}) \right]$$

where the 2nd term on the right hand side is simply the price effect of changes in "effective" depreciation and cost of capital. Also,

$$\dot{TFP} = \left\{ \left[\sum_{i=1}^{n} P_{oi} \, (O_{1i} - O_{oi}) - \left[\sum_{j=1}^{m-2} R^*_{oj} \, (I^*_{1j} - I^*_{oj}) \right] - R_{os} \, (I_{1s} - I_{os}) + R_{ot} \, (I_{1t} - I_{ot}) \right] - \left[R'_{os} \, (I'_{os} - I'_{1s}) + R'_{ot} \, (I'_{o1} - I'_{1t}) \right] \right\}$$

where we can isolate the separate effects of the EPCF and under-depreciation on TFP. This is due to the last term in square brackets on the right hand side.

The aggregate equality, equation (1) formally distinguishes between the various sources and uses of both financial and physical resources. The term $\sum_{i=1}^{l} O_{1i}(P_{1i} - P_{oi}) < 0$ since $O_{1i}(P_{1i} - P_{oi}) < 0$ $\forall i;\ i = 1.......\ell$ and denotes the value of resources used by the firm to lower the unit price of goods and services to the consumer. In other terms, the firm is doner and the consumer recipient.

$\sum_{i=1}^{\ell} O_{1i}(P_{1i} - P_{oi}) > 0$, on the other hand, constitutes a source. TFP, combining the effects of scale and technological progress is alternatively a use or source, depending on its sign. It, of course, refers to real resources. The reasoning behind $\sum_{j=1}^{g} I_{1j}\ (R_{1j} - R_{oj})$ and $\sum_{j=g+1}^{m} I_{1j}\ (R_{1j} - R_{oj})$, except that it refers to the interaction between the firm and the suppliers of productive inputs, is entirely similar to that outlining the effects of output price changes. Finally, if $\Delta e > 0$ then it is classified as a use item, applied to improving the position of the firm and vice versa for $\Delta e < 0$.

While the decomposition of any set of accounts, after TFP has been accurately measured, into its price and quantity (or financial and physical), components, is a fairly elementary exercise, the same cannot be said for its usefulness as a management tool. The above DG analysis is one of a useful methodology. Subsequent to a numerical example presented below, we will examine the same set of components presented, this time, to highlight the individual impacts on the change in net income, certainly one of the more important among the set of management's measures of success.

Numerical Example of DG:

Columns 1 and 2 of Table I give the current values of total output and both theoretical and effective input for the years 1975 and 1976. Theoretical capital costs of depreciation and the various financial charges are listed in rows 5a and 5b respectively, while the difference between these and the effective capital costs, known as "Under Depreciation" and "EPCF" are listed in rows 9 and 10, respectively. The pure profit/loss (P/L) item is listed in row 15. As can be seen, although it is negative in both years, the fact that it is larger in 1976 indicates an increase in profits.

The next step, after having adjusted for differences between theoretical and effective charges, is to adjust all 1975 values to reflect purchasing power in terms of 1976 values. This is accomplished by inflating all 1975 values by 1.05, the index of PGNE for the period 1975 to 1976. This adjustment is reflected in column 4 of Table I.

TABLE I

	1	2	3		4	5
	Current Value 1976	Current Value 1975	1975 Unit Prices 1976		1975 Value by PGNE 75-76 1975	Volume Index 1975-76
1. Output	76,468	68,032	83,578		71,434	1.17
2. Growth (76-75)				12,144		
EXPENSES						
3. Materials	20,895	20,536	19,837		21,562	0.92
4. Labour	17,510	13,349	16,118		14,016	1.15
5. Capital						
a) Theoretical Depreciation	18,000	15,000	18,000		15,750	1.16
b) Theoretical Financial	25,000	23,000	25,000		24,150	1.04
6. Total Theoretical Expenses	81,405	71,885	78,955		75,479	1.05
7. Growth (76-75)				3,476		
8. TFP (2-7) (Theoretical)				8,668		
9. Under Depreciation	1,500	800	1,400		840	
10. EPCF	3,000	2,053	2,700		2,150	
11. Total (7 + 8)	4,500	2,853	4,100		2,996	
12. Growth (76-75)				1,104		
13. Total Effective Expense (6-10)	76,905	69,032	74,855		72,483	1.03
14. Growth (76-75)				2,372		
15. Profit or Loss (1-12)	-437				-1,049	
16. TFP						
a) (1 ÷ 6)				11.4%		
b) (2-14) or (8 + 12)				9,772		
17. PGNE (75-76)				1.05%		

The final step in organizing the accounts into the required format, is to remove the effects of price changes from the 1976 current values, i.e., convert the

$$X_{i1}Y_{i1} \text{ to } X_{io}Y_{i1}; \quad X = P, R; \quad Y = O, I$$

As mentioned previously, this transformation can be effected either through deflation by an appropriate price index or, alternatively, through inflation by a quantity index. Using the latter method, the quantity indexes of column 5, Table I, are applied to inflate the 1975 (in the purchasing power of 1976) values (in column 4) into 1976 quantities, i.e.,

$$X_{i,75}Y_{i,75} \, (Y_{i,76} \,/\, Y_{i,75})$$

The results are listed in column 3 of Table I. It should be noted that while the theoretical values have undergone no change, the EPCF and under depreciation items have. This is due to the fact that EDF, whose methodology we are here drawing upon, initially calculates the theoretical values on a constant value basis and hence requires no further adjustment. The effective rates, on the other hand, are subject to exactly the same considerations as any other item and it is through these that the EPCF and under depreciation categories are affected. This can be seen by recalling that both, EPCF and under-depreciation are simply the differences between the respective theoretical and effective values.

TABLE II

	1 Current Value 1976	2 Current Value 1975 Unit Prices 1976	3 To Firm	4 To Others
1. Output	76,468	83,578		7,110 (to consu[illegible]er
2. Materials	20,895	19,837		1,058 (to supplier
3. Labour	17,510	16,118		1,392 (to labou[illegible])
4. Capital Expenses Effective				
4. Depreciation	16,500	16,600	100 (capital mining)	
5. Financial	22,000	22,300	300 (from capital suppliers)	
6. Total Effective Expenses	76,905	74,855		
7. Profit/Loss(1-6)	-437	8,723		
8. Profit/Loss (1975)	-1,049			
9. Total to Others				9,560
0. Improvement (7 - 8)				612
1. TFP gain			9,772	
2. TOTAL:			10,172	10,172

We are now in a position to calculate the respective Y_{1i} $(X_{1i} - X_{oi})$, TFP and Δe. The theoretical TFP is the difference between the changes in output volume and theoretical input volume. It is equal to 8,668. In addition, given that the combined volumes of EPCF and under-depreciation actually increased, signifying that the firm was able to get away with a lower than expected cost of capital, this added 1,104 to the value of TFP for a total gain of 9,772 (which, it should be noted, is equal to the effective gain in TFP). The fact that under-depreciation increased should not necessarily be interpreted as a positive event. It demonstrates a certain degree of capital mining whereby part of today's profits are earned at the expense of future plant replacement.

The P/L in terms of 1975 viewed from a 1976 purchasing power perspective is a loss of -1,049 making Δe = -437 - (-1,049) = 612.
For ease of calculation, columns 1 and 3 are transcribed as columns 1 and 2 of Table II. These allow us to derive the Y_{1i} $(X_{1i} - X_{oi})$. As can be seen, (in columns 3 or 4) the price of output has dropped thus benefiting the consumers to the extent of 7,110. Suppliers of intermediate goods, as well as labour were also not beneficiaries to the tune of 1,058 and 1,392 purely as a result of higher input prices. The supplier of funds, due to lower than expected rates, contributed 300 to the firm and depreciation, once again due to capital mining, was the source of another 100.

We can now summarize the total sources and uses statement:

		% Contribution
To consumers through output price reductions:	- 7,110	70%
To suppliers through input price increases:	- 1,058	10%
To labour through wage increases:	- 1,392	14%
To improvement of firms P/L:	- 612	6%
TOTAL USES:	-10,172	100%

To firm from under-depreciation:	100	1%
To firm from EPCF:	300	3%
To firm from TFP:	9,772	96%
TOTAL SOURCES:	10,172	100%

It should be noted that we have been using an example that is higly aggregated on two planes. Firstly it can be further broken down by input-output categories and secondly, for a multi-branch firm, by sector of operations. EDF, for example, has detailed disaggregation by input-output categories but by activity as well. Thus, in their analysis, they will have total company and also those activities which deal with the same product breakdown. Some sample tables, from EDF's 1975-1976 analysis are provided as Tables III, IV, and V. These include one variance report based on the sources and uses statement, indication that planning explicitly for productivity gains is an important consideration, and two sources and uses tables, one showing overall performance and one classified to include the contributions of the separate activities of "electricity production and Transportation" and "electricity distribution".

TABLE III

RESSOURCES ET EMPLOIS DE PRODUCTIVITES

(MF 1975)

	Valeurs 1975	Volumes 1975	Heritages	Emplois
PRODUITS				
Ventes on haute tension	4 336	3 876	460	
Ventes a l'etranger	130	127	3	
Ventes on moyenne tension	7 420	7 074	346	
Ventes en basse tension	12 988	13 743		755
TOTAL DES PRODUITS	24 874	24 820		
FACTEURS				
Combustibles fossiles	5 407	4 817		590
Combustibles nucleaires	246	247	1	
Achats aux tiers	1 807	1 849	42	
Personnel	5 606	5 429		177
Entretien (hors main d'oeuvre)	975	1 000	25	
Impots	1 159	1 164	5	
Loyers et redevances	705	747	42	
Autres depenses	1 341	1 362	21	
Provision C.N.R.	47	51	4	
Prais d'etablissement	361	379	18	
Amortissement industriel EFFECTIF	4 516	4 499		17
Charges financieres EFFECTIVES	3 166	3 236	72	
TOTAL DES FACTEURS	25 336	24 782		
RESULTAT 1975	- 462	+ 38		
RESULTAT 1974 (Rappel)	-1509	-1509		
Amelioration du resultas	+1047			1047

en valeur en volume

+1 546 +1 039 = 2 585

SURPLUS + HERITAGES=EMPLOIS

*Source: Les Progres de la Productivite Globale des Facteurs a Electricite de France Commission d'Exploitation du 24 juin 1976.

TABLE IV

RESSOURCES ET EMPLOIS DE SURPLUS DE PRODUCTIVITE
Reclasses par postes principaux

Ce reclassement entraine une contraction pour les postes ou apparaissement a la fois des heritages et des emplois: clients, combustibles et Etablissement.

(MF 1975)

	ENSEMBLE E.D.G.		PRODUCTION-TRANSPORT		DISTRIBUTION	
RESSOURCES						
SURPLUS EFFECTIF		1 546		549		997
HERITAGES						
. des clients						
Ventes H.T.	460		460			
Ventes a l'etranger	4		4			
Ventes M.T.	346				346	
Ventes B.T.	- 755				- 755	
Livraisons aux C.D.			633			
Total Clients		55		1 097		- 409
. des fournisseurs						
Entretien	25		15		10	
Loyers & redevances	42		42			
Autres depenses	25		12		13	
Total fournisseurs		92		69		23
. des preteurs		72		84		23
. de l'Etat & des Collectivite		4		- 8		- 12
Total Ressources		1 769		1 791		611
EMPLOIS						
. Combustibles & achats						
fossiles	589		587		2	
nucleaires	- 1		- 1			
achats aux tiers	- 42		- 44		2	
achats aux CIME					633	
Total Combustibles		546		542		637
. Personnel		177		57		120
. Etablissement						
amortissement	17		- 4		21	
frais d'etablissement	- 18		1		- 19	
amelioration de result	1 047		1 195		- 148	
Total etablissement		1 046		1 192		- 146
Total EMPLOIS		1 769		1 791		611

*Source: Les Progres de la Productivite Globale des Facteurs a Electricite de France.
Commission d'Exploitation du 24 juin 1976.

TABLE V

INDICES DE VOLUME ET TAUX DE PRODUCTIVITE

COMPARAISON DES RESULTATS ET DES PREVISIONS

	RESULTATS (Volumes)	PREVISIONS (Volumes)	ECARTS ponderes	
PRODUITS				
Ventes H.T.	95,76	103,54	- 1,3	
Ventes a l'etranger	101,56	93,80	-	
Ventes M.T.	102,40	105,31	- 0,8	
Ventes B.T.	112,37	109,11	1,8	
Livraisons aux C.D.	--	--	-	
	106,5	106,9	- 0,4	
FACTEURS				
Combustibles fossiles	94,80	107,18	- 1,8	
Combustibles nucleaires	120,53	129,31	- 0,1	-2
Achats aux tiers	97,36	114,17	- 0,9	
Achats aux CIME				
Personnel	100,48	101,89	- 0,2	-0
Entretien (hors main d'ouevre)	93,79	102,47	- 0,3	
Impots	109,83	104,24	0,2	
Loyers et redevances	97,85	100,00	- 0,1	-0
Autres depenses	103,42	100,31	0,1	
Provision C.N.R.	100,00	100,00		
Frais d'establissement	127,37	90,22	0,3	
Amortissement indust. theorique	98,86	98,39	0,1	1
Charges financieres normatives	103,04	100,81	0,7	
TOTAL	100,6	102,6	- 2,0	
TAUX de P.G.F.	5,85%	4,27%	+ 1,58	

*Source: Les Progres de la Productivite Globale des Facters a Electricite de France. Commission d'Exploitation du 24 juin 1976

B. Net Income Analysis:

Another way of presenting the same components, which is used for control and planning in the AT&T system companies, involves an analysis that focuses attention on the change in Net Income. If we begin with the basic equation,

$$\Delta e = \Delta PI - \Delta RI$$

and expand it into the components, then,

$$\Delta e + I_1^*(R_1^* - R_o^*) + \left[I_{1s}(R_{1s} - R_{os}) + I_{1s}'(R_{os}' - R_{1s}')\right] + \left[I_{1t}(R_{1t} - R_{ot}) + I_{1t}'(R_{ot}' - R_{1t}')\right]$$

$$= O_1\ (P_1 - P_o) + T\dot{F}P$$

and by adding $R_{os}(I_{1s} - I_{os}) + R_{os}'(I_{os}' - I_{1s}') + R_{ot}(I_{1t} - I_{ot}) + R_{ot}'(I_{ot}' - I_{1t}')$ to both sides, where,

$$I_{1s}(R_{1s} - R_{os}) + I_{1s}'(R_{os}' - R_{1s}') + R_{os}(I_{1s} - I_{os}) + R_{os}'(I_{os}' - I_{1s}') =$$

change in Net Income (ΔNI)

and

$$I_{1t}(R_{1t} - R_{ot}) + I_{1t}'(R_{ot}' - R_{1t}') + R_{ot}(I_{1t} - I_{ot}) + R_{ot}'(I_{ot}' - I_{1t}') =$$

change in current depreciation (ΔD)

and,

$$R_{os}(I_{1s} - I_{os}) + R'_{os}(I'_{os} - I'_{1s}) + R_{ot}(I_{1t} - I_{ot}) + R'_{ot}(I'_{ot} - I'_{1t}) =$$

growth in total constant value capital input (ΔK)

We have,

$$\Delta NI + \Delta D + I^{*}\ (R_1^{*} - R_o^{*}) + \Delta e = O_1\ (P_1 - P_o) + TFP + \Delta K$$

and if we modify ΔK to reflect the value of theoretical capital growth, which, if we are interested in real resources, is in fact the figure we should use then:

$$\Delta NI + \Delta D - I_1^{*}(R_1^{*} - R_o^{*}) + R'_{os}(I'_{os} - I'_{1s}) + R'_{ot}(I'_{ot} - I'_{1t}) + \Delta e = O_1(P_1 - P_o) + TFP + \Delta K^{I}$$

or, using more compact notation,

$$\Delta NI + \Delta D + I^{*}(R_1^{*} - R_o^{*}) + \Delta EPCF + \Delta UO + \Delta e = O_1\Delta P + TFP + \Delta K^{*}$$

where:

$O_1\Delta P$ = the impact of rate changes, both implicit and explicit, i.e., those amounts due to deliberate changes in output prices as well as those due to changes in average revenue per unit of output as a result of changes in the mix of outputs.

$I^{*}(R_1^{*} - R_o^{*})$ = the <u>inflationary impact</u> on (1) $L\Delta w$ = labour expense changes due to increases in the wage rate and (2) $M\Delta m$ = intermediate or material input expense changes due to increases in the relevant price variables (such as indices for electricity, fuel, stationery, etc.)

ΔK^* = the contribution of physical capital growth. Clearly, apart from that portion of output increase which is due to a higher capital productivity and which has already been captured by the TFP term, net income must grow by at least the growth in invested capital, times the required rate of return to invested capital, net of the portions going to effective depreciation, EPCF, UD and an improvement in the firm's P/L position.

ΔD = the impact of effective depreciation, in current value.

$\dot{TFP}$ = the contribution of productivity gains.

$\Delta EPCF$ = the amount by which ΔNI did not have to increase (decrease) due to favourable (unfavourable) financial conditions.

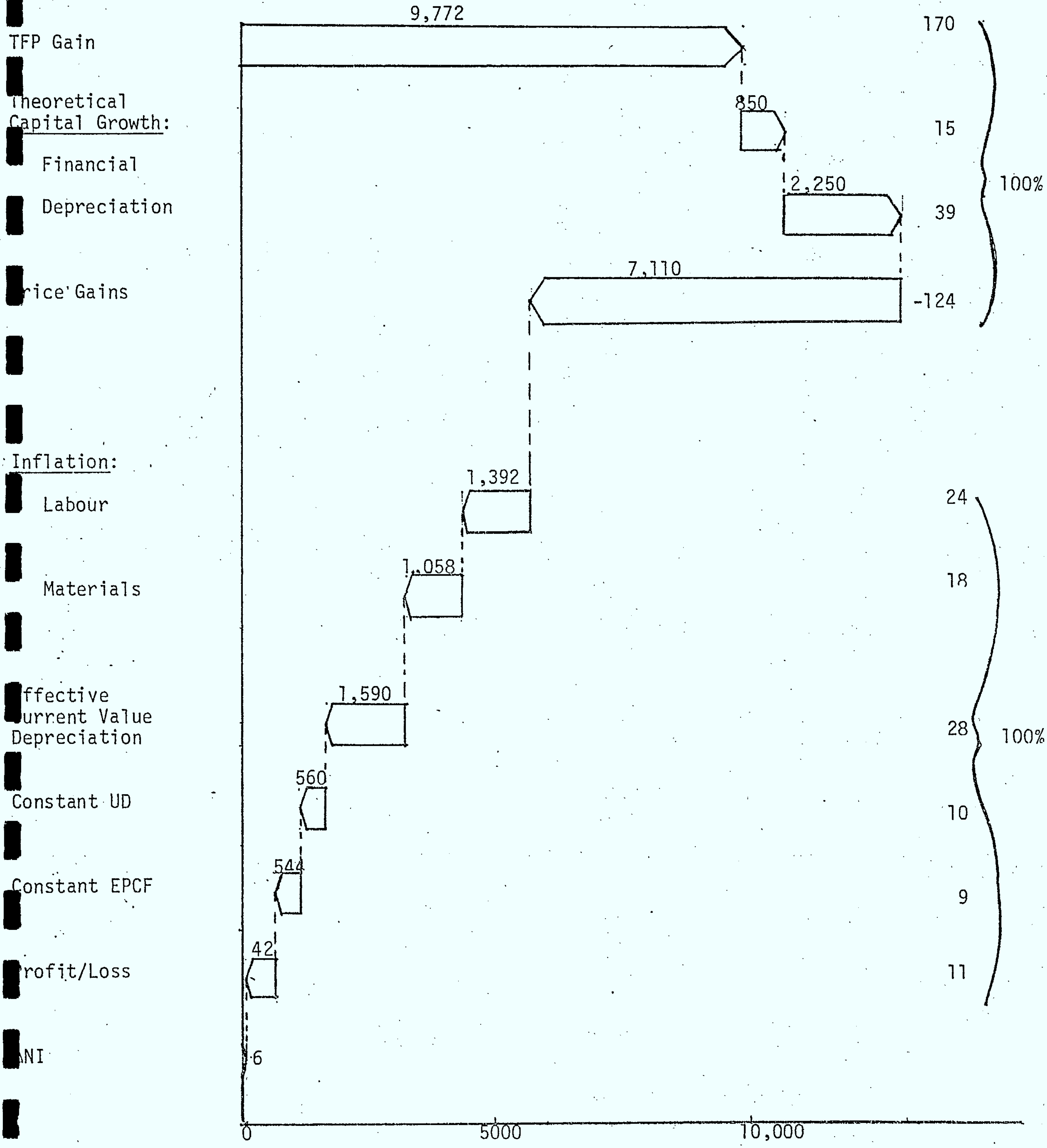
ΔUD = the amount by which ΔNI did not have to increase (decrease) due to capital mining (over depreciation).

Δe = the amount be which, beyond the required return to capital the firm was able to improve (deteriorate) its pure profit (loss) position.

ΔNI = the current value of the change in capital costs (which in the example provided is equal to the required change in debt service costs).

The analysis can be followed through Fig. 1. As can be seen the gains in productivity accounted for the lion's share of all contributes to the total positive impact on ΔNI. The growth in capital contributed 54% and the reduction in average price per unit of output, reduced the total impact quite substantially. In total, however, not only did NI increase by at least the amount of capital input growth, but an excess, of 612 was available in order to

Fig. 1


9,772
TFP Gain
170
Theoretical
Capital Growth:
Financial
850
15
Depreciation
2,250
39
100%
rice Gains
7,110
-124
Inflation:
Labour
1,392
24
Materials
1,058
18
ffective
urrent Value
Depreciation
1,590
28
100%
Constant UD
560
10
Constant EPCF
544
9
rofit/Loss
42
11
NI
6
0
5000
10,000

improve the firm's loss position over the previous year. It is interesting to note that the improvement, as well as the required increase in NI was possible because of effective capital costs that were lower than theoretically estimated values, by 544, on the one hand, and capital mining, or under-depreciation of 560, on the other hand.

C. Planning:

The mechanics of DGA and net income analysis, whereby the budgetary accounts of previous or future periods are scrutinized in terms of the relative contribution to growth from productivity, prices and volumes on the one hand and their distributional implications on the other hand, suggest the feasibility of developing an inverse procedure to derive a set of forecast accounts beginning with productivity, price and quantity information. Teleglobe Canada is in fact developing such a model (Werner, 1979). Essentially, it builds an entire set of accounts, including the income statement and balance sheet, based on certain key management targets which include the growth in demand, desired growth in TFP and the required rate of return to invested capital. It is not meant to replace the traditional planning process, but only to provide management with a set of budgetary guidelines that already embody their basic objectives. They are thus in a position to provide much more informed direction during the budgetary planning process.

IV.3 Regulation and Efficiency

The goals to be pursued in regulation are often stated in broad and fuzzy terms. For example in 1973 the Canadian Transport Commission interpreted its responsibility to be the protection of the public interest. This required the 'best' telecommunications system to be provided at the 'lowest possible cost' not only today but in the future. The telephone rates themselves were to be 'just and reasonable' and free from 'undue discrimination'. The generalities are pleasant and hardly rigorous guidelines for 'good works'. In actual practice, Federal regulation has concentrated on a relative few major issues with more minor ones appearing as time has progressed. The regulation by the Provinces has been similar in some cases although generally less intensive. Where the government owns the telephone system, explicit regulation has been quite minor until very recently.

No detailed treatment of existing regulation is intended in this report. Rather we wish to consider several major regulatory practices and investigate their relationship to efficiency. Before turning to this task, it may be useful to re-state an underlying premise. <u>Efficient production is always desirable and growth in efficiency should always be a goal of the regulator</u>. It is not necessary to elaborate on this condition and it will be presumed throughout this section.

Regulation in telecommunications has implied the creation of a legal monopoly. That is certain types of services are to be provided solely by the legal monopolist. If any ambiguity about the definition of the services exists, these can be settled in regulatory hearings, or law courts. It has been suggested, e.g. Shepherd, that the creation of an artificial boundary defining the monopolist is one source, (potential or actual), of inefficiencies. With a given technology, The boundary provides an exclusive tariff that shields the monopolist from any competitive pressures. This shield implies that the forces generating efficient production must come from within the monopoly company or from the regulator. If society does not wish to rely on the goodwill of the management to create an efficient operation then it must strive to use regulatory means to assure efficiency.

The problems may be even more severe in a dynamic context of technical change. A particular boundary may be sensible with a given technology and not with a very different one. The current controversies in policy making with regard to pay TV, satellites, cable television and optimal fibres are sufficient examples. However, the point may be re-phrased to emphasize a persistent problem. With a fixed boundary, innovative activity by the telephone company monopolist is constrained to activities within the boundary. This may hinder the efficient discovery and application of knowledge outside the boundary. The reverse is also true and the stress has often been placed on the freedom of the monopolist from competitive pressure in innovation activities. The incentive for other firms outside the boundary to innovate in activities that might be inside the boundaries has been diminished. In practice, there is the further fear that the boundary may be altered in order to incorporate new innovations on the boundary.

There is no possibility of avoiding the definition of a boundary but there are probably efficiency losses in defining any particular boundary and the boundary must change over time. We have not been able to clearly indicate how the boundary should be selected in order to minimize the inefficiency loss although some answers would seem possible. This problem will be directly investigated in the next phase of the project. For the purposes of this report, we should note that the type of efficiency measures proposed in this report will be sensitive to the boundary definition. However, it is unlikely that without further work one could separate out the precise magnitude of the effect.

Two practical policies utilized in Federal regulation have been selected for discussion. They are rate of return regulation and rate regulation, and will be studied in that order.

If a monopoly is created and the goal of the monopoly is profit maximization it is well known that the monopoly may charge a price that is above marginal cost and consequently produce an inefficiently small output level. In addition, the profits earned by a monopoly may be higher than the minimum level required to keep the capital invested in the production of the monopoly service. Rate of return regulation is directed towards restraining the profits earned by the regulated monopoly. It is not directed towards the first problem. In its simplest form the monopoly is permitted to earn a return on its capital comparable to the return achieved elsewhere in the economy for investments with similar risks. This is equivalent to average cost pricing in the context of a monopolist producing a single output with cost curves which include the appropriate rate of return. Achieving an appropriate rate of return does not ensure that the efficient output level is produced. With price discrimination and/or multiple outputs (with or without joint production) there may be many output vectors that will yield the revenue required for the appropriate rate of return. Rate of return regulation alone will not produce the socially optimal output level.

The last decade has been replete with studies of the inefficient input choice possible under rate of return regulation, Averch and Johnson (1962). A recent paper by Atkinson and Halvorsen (1980) indicates the presence of these distortions in electrical utilities. Denny, Fuss and Waverman (1979) derive an explicit expression for evaluating the effects of rate of return regulation on measured productivity. Fuss in some unpublished work has explicitly evaluated the effects of the detailed type of rate of return regulation used in Canada by the Federal government.

The possible inefficiency created by rate of return regulation is well documented. One of the uses of productivity measurement for regulatory purposes is to offset these inefficiencies through either incentives or other policies.

There is a widely held belief throughout the telecommunications industry that local services are subsidized by toll services and that some local services, e.g. residential and rural, are subsidized more than others. While evidence has not been produced which would convince the sceptics, let us assume that some types of cross-subsidization does occur. More generally, if prices do not equal marginal costs, appropriately defined, then there will be inefficiencies in production. In Denny, Fuss and Waverman (1979), the consequences for productivity measurement of non-marginal cost pricing were developed. The methodology is complete although the quantification of the inefficiencies, measured in terms of changes in efficiency, is not necessarily easy. The quality of the available data will determine the quality of the answer.

The basic function of productivity measurement for regulation should be to encourage efficient production. However, it may also be used to assess the trade-off in terms of lost efficiency, measured in productivity units, of pursuing alternative goals.

It may be useful to suggest, in rough form some possible practical methods of using productivity in regulation. The work on this aspect is only beginning and the form in which the proposals are presented here are not final.

(a) <u>TFP target</u>. The regulator can announce a target level of TFP growth that a firm must meet or else suffer some penalty. The level chosen could be based on average experience in the industry or perhaps some specific percentage above average performance. Since year to year variations can be large, a moving average would probably be sensible. The precise methods used to measure productivity would become contentious although many of the issues already arise with rate of return regulation.

The penalties and rewards for productivity performance could be derived in many alternative ways. For example, one might link the allowed rate of return to the performance on productivity.

(b) 'Slice of the Pie' Regulation: This type of regulation would utilize an analysis similar to the gains analysis discussed earlier. Productivity gains become distributed amongst shareholders, customers and perhaps workers and suppliers. The slice of the pie type of regualtion would guarantee the firm and consequently the shareholders a given proportion of the productivity gains. The rest would be largely distributed to consumers. The firm has an incentive to increase productivity since it will gain profts. Obviously this procedure will conflict with any rigid control over the rate of return. The rate of return might fall sharply or rise above traditionally permitted levels. Some control should be available through rate change limitations. That is the only provision for permitting rates of return to rise above certain levels would be if tariff rates were not rising. Alternatively, a ceiling or perhaps a floor on the rate of return could be imposed which would permit the firm to locate with the range based on its productivity performance. While no detailed suggestions are made here this type of incentive may be preferable since it recognizes both the changes in efficiency and its distribution. The two types (a) and (b) can be tied together by requiring some minimal target before "slice of the pie" contributions are considered.

A. Government Guidelines:

While much effort has been devoted, in various firms, to the development of comprehensive, well documented Total Factor Productivity studies, their potential as management tools has received remarkably little attention. This, at least, is the case in telecommunications. The opposite is true outside of the telecommunications industry. Some firms, such as IBM, ALCOA, Texas Instruments and so on, while they have calculated productivity number of questionable accuracy, they are, more and more, becoming the basis for important decisions. In France, not only has productivity measurement in the "state" owned industries reached a very advanced state, but in addition, it has been extensively applied as part of the central planning process to which these industries are subject. The state, based on past performance, sets TFP targets and includes them in the contracts between the state and each of the large state owned enterprises. For example, in the 1970 contract between the State and the French Electric Utility (Electricité de France) article 3 states:

> "A l'échéance de la période de cinq ans couverte par le Contrat, les gains de productivité réalisé par l'Etablissement sur <u>l'ensemble de ses facteurs de production</u> devront avoir atteint un total correspondant à un taux moyen de progrès annuel au moins égal à 4.85%,....".

Clearly, with this type of very specific target the firm has no choice but to include productivity growth directly in its planning mechanism. The methodology at Electricité de France is iterative - ex post, i.e. the productivity gains implicit in any current budget are measured and any deviations from the State central plan are noted and either explained or used to modify the initial budget, a process which eventually converges to a solution incorporating an explicit effort to meet both financial (rate of return) and productivity targets. The entire control apparatus is actually quite detailed, incorporating a complete analysis of distributional changes, over time, due to tariff modifications, input price variations and productivity .[4]

B) Automatic Rate Adjustment (ARA):

Although regulation of telecommunications has not yet become automatic, the merits of ARA have been quite hotly debated. While everyone seems more or less agreed that a method whereby the present lengthy and costly full scale rate hearings can be streamlined is desirable, and that ARA could be the vehicle through which they at least become less frequent, there is far less concurrence as to both the specification and composition of an acceptable ARA formula, with generally less agreement on the latter question. The first issue, specification, is basically technical. Does a formula exist which, at once, ensures that the firm can maintain an acceptable financial return and, as well, provides incentives for the maintenance and improvement of efficiency? The most popular specifications, designed to meet these goals, have included productivity variables along with allowed earnings ranges such that better productivity performance can be rewarded. The best known of this formula type is the one developed and presented by Illinois Bell. Briefly it was specified as

$$\Delta R^{*} = \min \left[(a\Delta C + b\Delta VTFP), \Delta R_{max} \right]$$

where ΔR^{*} = allowed change in revenues

ΔC = test year change in allowable costs

$\Delta VTFP$ = value of changes in productivity

ΔR_{max} = maximum allowable change in revenues with respect to ensuring that the rate of return does not exceed some allowable upper limit

a,b = coefficients

The coefficients, a and b would be estimated from historical data. The logic is quite straightforward and appealing. The regulator would set an upper limit on the rate of return and then using some forecast of financial and economic conditions would also set the maximum ΔR_{max} that would ensure an upper limit on returns to capital. In addition, the regulator would guarantee to allow rate increases to cover a fraction "a" of all allowable cost increases. The remainder of the cost increase, up to a maximum of $(\Delta R - a\Delta C)$ would be covered by gains in productivity, $b\Delta VTFP$, where b is the multiple of these gains allowed in the calculation of ΔR^*. Then natural strategy of the firm would be set to

$$\left[a\Delta C + b\Delta VTFP\right] - \Delta R_{max} \geq 0$$

by maximizing $\Delta VTFP$, which serves the purpose of promoting efficiency. With "b" set to reflect the multiple of historical VTFP that was required to ensure that $(a\Delta C + b\Delta VTFP)$ falls within the "zone" of a reasonable rate of return, after "a" has been set (arbitrarily), the firm will approach ΔR_{max} as its growth in productivity increases. This reasoning of course implies that all the elements of ΔC are uncontrollable.

The contentious aspects of an ARA procedures centres around two issues[5]: (1) The appropriate definition of productivity; (2) The division between controllable and uncontrollable cost and (3) The indexing of the appropriate set of costs. We elaborate on these as follows:

(1) The first issue, concerning the choice of an acceptable "productivity offset", questions:

a) The acceptability of existing measures in terms of volatility and uniqueness.

b) The feasibility of using general, economy-wide measures of productivity either due to the absence of company specific information or in order to ensure consistency of application.

c) The merits of using partial measures to offset individual component cost changes, such as labour or total measures to offset aggregate changes.

(2) The division of between controllable and uncontrollable costs was examined by the DOC Economic Policy and Statistics Branch (DOC, July 1975) and we summarize their findings, with respect to individual cost components, below:

a) Wages and Salaries: while wages and salaries are subject to collective bargaining for unionized employees, the rest, which may comprise a good proportion, are subject to the discretion of management. In addition, even those subject to collective bargaining are, to a certain extent, influenced by the efficiency of the management bargaining unit. Finally, labour costs are highly sensitive to productivity changes and must be appropriately adjusted.

b) Taxes: Excluding income taxes (which offset would interfere with government fiscal policy) most profit maximizing firms cannot, in the short run, pass the entire burden of increased taxes onto the consumer. Therefore, although they may be unambiguously uncontrollable, an offset may discriminate in favour of regulated firms.

c) Depreciation: All changes in depreciation rates, assuming that this is one component of depreciation expense changes, are not necessarily uncontrollable. To the extent that such rates are based on forecasts of technological obsolsence, they are not entirely outside of management control.

d) Cost of Capital: The question in this case is not so much whether certain components of this cost, such as interest payments due to changes in the embedded cost of capital, are uncontrollable, but rather, even though they are beyond management's control, what are the implications for the overall rate of return if one component is allowed to be indexed. It is argued, on the one hand, that indexation would cause alterations in the overall rate of return without recourse to a rate hearing, which is contrary to the current spirit and intent of the regulatory mechanism, and on the other hand, that the existence of indexing only for operating expenses may provide an incentive for buy rather than make decisions.

There is no doubt that productivity can be used to advantage by both management and regulators. Our examples do not exhaust the possibilities and they are certainly not precise enough to guide practical implementation, particularly in regulation. In the managerial area the companies currently using productivity can certainly assist with practical details. Further specifics will be developed with the companies during future phases of the project. One interesting development is the possibility of incorporating more direct engineering efficiency measures into this work. This is particularly true with respect to the evaluation of investment decisions by firms and changes in the boundary, e.g. control of pay T.V., by the regulators.

The regulatory uses of productivity requires an intensive effort during the next phase of this project. The details of our proposals are suggestive at this stage although the principles underlying them are sound.

V. Index Numbers

1. Introduction

This section will be a rapid and partial overview of the index number problem. More on these index numbers will be presented in subsequent phases of this project. Even though it is not exhaustive, it attempts to present some recent results in this domain. Starting from the elementary index, Laspeyres and Paasche indices are considered. The natural step then is to follow Fisher's approach in the making of index numbers and to derive a class of ideal index numbers. The two underlying properties one might desire are time and factor reversal, and an alternative approach which has been proposed by Divisia is to define over infinitely small changes in time an index which meets both properties. One of the main critics of Divisia indices, levied among others by Usher, relates to the path of integration. It is shown however that path independence is directly related to positive linear homogeneity (PLH). PLH , it is indicated, may not be as serious a problem as it may seem as long as a chain index is adopted and as long as the changes are sufficiently progressive. The alternative, or rather complementary approach to Divisia indices consist in using exact and superlative indices. Since index numbers, in productivity analysis are used for very specific purposes which relates to the study of the production process, the theory of exact and superlative index numbers has the added advantage taking index numbers from a "mechanical" approach and to integrating them into the context of economic theory.

The problem begins at the level of data availability. If over the years there was just information on quantities, (or prices), a crude index using solely that information could be constructed as a simple average of elementary indices. Given, in addition, for at least one period, which can be referred to as the base period, observations on prices, (or quantities), a Laspeyres type of index could be generated. As soon as observations are available for all periods for both prices and quantities, a desirable class of indices to be considered is the class of superlative indices. However at this stage, given the inability to find in that class one index which seems to be on the whole superior, it would seem wisest to develop at the individual company level, a data base on the raw data upon which alternative index aggregation could readily be applied depending upon the analysis. The flexibility of this latter approach appears particularly desirable when one considers the rapid development in the theory of index numbers over recent years and the number of questions which still remain unanswered. Where this is not a feasible approach, a second best approach would be to select either a Tornqvist index which has the advantage to being related to translogarithmic functional forms or a Fisher ideal index which is also superlative; i.e., related to a flexible functional form to describe production and which meets the factor reversal test in addition to the time reversal test.

Finally even though recent results by Diewert indicate that a chain index approach is advisable, results by Hall and Star suggests that it may not be sine qua non to have, for every period, data on both quantities and prices.

V.2 Elementary Indices

The aim of index analysis is to study fluctuations in time, space, ... of certain variables. In this note, the analysis will be restricted to the time dimension.

Before tackling the aggregation problem, in an index number, one must tackle the comparison problem, i.e. before talking of the index problem of an aggregate of variables, one must tackle the index problem of a single variable. To illustrate that point, we may consider local telephone services. Clearly, there is no point of talking of a price index for local services if one cannot handle the price index of some unambiguously defined entity such as contract primary/residential/single party service for some given tariff group.[6] The construction of index numbers for such well-defined micro-entities is what concerns us in this section.

Let some such elementary entity be denoted by x_t where x denotes either a quantity, say number of residential single party main lines, in a given tariff group, or a price, and t denotes the date of the measurement. The index number problem consists in finding a measure to compare x in period t' to x in period t.

Given x_t and $x_{t'}$, a possible measure would be

$$(x_{t'} - x_t) \quad .$$

While there is nothing wrong with such a measure, it is normally more convenient to deal with measures which are dimension-free. If x_t denotes a number of certain type of main lines, $(x_{t'} - x_t)$ will also have the dimension "main line". One way to eliminate that dimension is to make that measure "relative" to a number of that type of main lines, say x_t, i.e. to take the ratio

$$\left(\frac{x_{t'} - x_t}{x_t}\right)$$

Such a type of measure is said to be commensurable. The elementary index may now be defined.

<u>Definition</u>: Given a variable denoted by x_t observed at time t, the elementary index of change, $E_t^*(x)$, of x_t at time t is defined as the proportional rate of change of x_t at time t, i.e.

$$E_t^*(x) = d\ln x_t = \frac{\dot{x}_t}{x_t}$$

Even though this form will be useful for developing the Divisia index, one does not normally have continuous data. Even when one has continuous data, as in the case of regulated tariffs, the data are constant over periods of time and change at specific time period by a discrete quantity, i.e. the series x_t is not a differentiable function of time. In practice the change will be taken in a discrete fashion from period to period, and it can be seen, if need be, as some approximation of an elementary index of change:

$$E_t(x) = \frac{x_{t+1} - x_t}{x_t}$$

In face, this new index depends upon two periods, which happen to be here periods t and $(t+1)$. Clearly we could consider any other time interval to yield

$$E_{t,s}(x) = \frac{x_s - x_t}{x_t}$$

Furthermore, we observe that, by adding one, a new measure is obtained which is simply the ratio of x_s to x_t .

Definition: Given a variable x_t , the elementary index between periods t and s , $I_{t,s}(x)$, of x_t is defined as the ratio of x_s to x_t , i.e.

$$I_{t,s}(x) = (\frac{x_s}{x_t})$$

Elementary indices can be analyzed in terms of properties which are considered desirable. These will be considered in terms of $I_{t,s}(x)$.

Definition: An index $I_{t,s}(x)$ is said to be time reversible if and only if, for all t and s ,

$$I_{t,s}(x) \cdot I_{s,t}(x) = 1$$

This property is somewhat comparable to stationarity in time series analysis. It implies that the index is not affected by time in that time intervenes only to situate the event. It is easily generalized to more than two periods to yield the circularity property:

Definition: An index $I_{t,s}(x)$ is said to be transitive (circular) with respect to a set of periods $t_0, t_1 \ldots, t_n$, if and only if

$$I_{t_0,t_1} \cdot I_{t_1,t_2} \cdot \ldots \cdot I_{t_{n-1},t_n} = I_{t_0,t_n}$$

It follows immediately that a circular index has the chain property.

<u>Definition</u>: An index $I_{ts}(x)$ is said to be chained with respect to periods t and $s = t+n$ if and only if

$$I_{t,t+1} \cdot I_{t+1,t+2} \cdot \ldots \cdot I_{s-1,s} = I_{t,s}$$

At this stage very little has been said about the entity x_t . It has only been assumed that it is a scalar, given t , however nothing prevents x_t from being derived from a set of other variables say $x_{i,t}$, $i=1,2,\ldots,N$, say as a weighted sum:

$$x_t = \sum_{i=1}^{N} a_i x_{i,t}$$

Here the weights have been assumed to be independent of time. In economics, these coefficients will generally be assumed to be quantities or prices, depending on the nature of $x_{i,t}$.

Now the elementary index of x_t , will be the weighted sum of the $x_{i,t}$, $i=1,2,\ldots,N$'s elementary indices:

$$I_{t,s}(x) = \frac{\sum_{i=1}^{N} a_i x_{i,s}}{\sum_{i=1}^{N} a_i x_{i,t}}$$

$$= \sum_{i=1}^{N} \left(\frac{a_i x_{i,t}}{\sum_{i=1}^{N} a_i x_{i,t}} \right) \left(\frac{x_{i,s}}{x_{i,t}} \right)$$

$$= \sum_{i=1}^{N} w_{i,t} \cdot I_{t,s}(x_i)$$

where

$$w_{i,t} = a_i w_{i,t} / \sum_{i=1}^{N} a_i x_{i,t}$$

While there were no restrictions on $(\sum_{i=1}^{N} a_i)$ and while a_i was assumed independent of time, this new set of weights is different:[7]

$$\sum_{i=1}^{N} w_{i,t} = 1$$

and each weight depends upon the period t which will be called the base period.

Let's consider some simple example. Let $x_{i,t}$ be Bell Canada residential contract primary single party price (or tariff), i denoting the tariff group. Let a_i be the forecasted quantity of such main lines in period t, where t is, say, 1979. Then

$$x_t = \$31,716 \text{ K}$$

Let $x_{i,s}$ be Bell Canada's rates effective in 1978. Then

$$x_s = \$30,203 \text{ K}$$

$$I_{t,s}(x) = 1.0501$$

$$= \sum_{i=1}^{N} w_{i,t} \cdot I_{t,s}(x_i)$$

where $w_{i,t}$ and $I_{t,s}(x_i)$ are as Bell's forecasted share of residential-contract primary individual line revenues in rate group i, $i=1,2,\ldots,17$ and the ratio of prices in the same rate group. Given $i = 8$, then

$w_{i,79} = .095$ and $I_{79,77}(x) = 1.047$.

However, this is not the only relationship between $I_{t,s}(x)$ and the elementary indices of the components $x_{i,t}$ since

$$I_{t,s}(x) = \left\{ \sum_{i=1}^{N} \left(\frac{a_i x_{i,s}}{\sum_{i=1}^{N} a_i x_{i,s}} \right) \left(\frac{x_{i,s}}{x_{i,t}} \right)^{-1} \right\}^{-1}$$

$$= \left\{ \sum_{i=1}^{N} v_{i,s} \, I_{s,t}(x_i) \right\}^{-1}$$

where

$$v_{i,s} = a_i x_{i,s} \Big/ \sum_{i=1}^{N} a_i x_{i,s}$$

The result can easily be generalized in terms of the r-mean introduced in the appendix. If x_t is obtained as a weighted r-mean of the set of $x_{i,t}$ such that

$$x_t = \left\{ \sum_{i=1}^{N} a_i (x_{i,t})^r \right\}^{1/r}$$

then

$$I_{t,s}(x) = \left\{ \sum_{i=1}^{N} w_{i,t} (I_{t,s}(x_i))^r \right\}^{1/r}$$

where

$$w_{i,t} = \frac{a_i (x_{i,t})^r}{\sum_{i=1}^{N} a_i (x_{i,t})^r}$$

i.e., $I_{t,s}(x)$ is an r-mean of the elementary indices $I_{t,s}(x_i)$. Similarly

$$I_{t,s}(x) = \left\{ \left[\sum_{i=1}^{N} v_{i,s} (I_{s,t}(x_i))^r \right]^{1/r} \right\}$$

where

$$v_{i,s} = \frac{a_i (x_{i,s})^r}{\sum_{i=1}^{N} a_i (x_{i,s})^r}$$

i.e. it is also the inverse of the r-mean of the elementary indices $I_{s,t}(x_i)$.

If we let r tend to zero, then

$$x_t = \prod_{i=1}^{N} (x_{i,t})^{a_i}$$

$$I_{t,s}(x) = \prod_{i=1}^{N} I_{t,s}(x_i)^{a_i}$$

and the same weights are used to generate both the aggregate variable x_t and the elementary index $I_{t,s}(x)$. It is now much more desirable to have $(\sum_{i=1}^{N} a_i) = 1$ since

$$x_t = \left(\prod_{i=1}^{N} x_{i,t} \right)^{\left(\sum_{i=1}^{N} a_i \right)}$$

A possible value for a_i could be the expenditure share in period t :

$$a_i = \frac{p_{i,t} \cdot x_{i,t}}{\sum_{i=1}^{N} p_{i,t} \cdot x_{i,t}}$$

Still another possibility could be a simple arithmetic average of the expenditure shares in periods t and s :

$$a_i = (\tfrac{1}{2}) \left\{ \frac{p_{i,s} \cdot x_{i,s}}{\sum_{i=1}^{N} p_{i,s} \cdot x_{i,s}} + \frac{p_{i,t} \cdot x_{i,t}}{\sum_{i=1}^{N} p_{i,t} \cdot x_{i,t}} \right\}$$

The weighted r-mean can be approached in still another way whenever $\left(\sum_{i=1}^{N} a_i \right) = 1$:

$$x_t = \left\{ \sum_{i=1}^{N} a_i [(x_{i,t})^r - 1] + 1 \right\}^{1/r}$$

$$= \left\{ 1 + r \sum_{i=1}^{N} a_i x_{i,t}(r) \right\}^{1/r}$$

where

$$x_{i,t}(r) = \frac{x_{i,t}^r - 1}{r}$$

$x_{i,t}(r)$ is none other than the Box-Cox transformation of $x_{i,t}$ (1964). Then the Box-Cox transformation of the aggregate x_t is the weighted arithmetic mean of the Box-Cox transformation of the $x_{i,t}$'s :

$$x_t(r) = \sum_{i=1}^{N} a_i x_{i,t}(r)$$

where

$$x_t(r) = \left(\frac{x_t^r - 1}{r} \right)$$

It follows immediately that

$$I_{t,s}[x(r)] = \sum_{i=1}^{N} w_{i,t}(r) I_{t,s}[x_i(r)]$$

where

$$w_{i,t}(r) = \frac{a_i x_{i,t}(r)}{\sum_{i=1}^{N} a_i x_{i,t}(r)}$$

All the above results hold equally for the elementary change index $E_t^*(x)$, which, where x_t is a linear combination of the $x_{i,t}$, can alternatively be written as:

$$E_t^*(x) = \left\{ \sum_{i=1}^{N} w_{i,t}[E_t^*(x_i)]^r \right\}^{1/r}$$

or

$$E_t^*(x) = \left\{ \left(\sum_{i=1}^{N} v_{i,t}[E_t^*(x_i)]^r \right)^{1/r} \right\}^{-1}$$

Alternatively x_t may be obtained as a linear combination of the $x_{i,t}$'s where the coefficients are dependent upon time, say $a_{i,r}$:

$$x_t = \sum_{i=1}^{N} a_{i,r} \cdot x_{i,t}$$

$$I_{t,r}(x) = \frac{\sum_{i=1}^{N} a_{i,r} \cdot x_{i,s}}{\sum_{i=1}^{N} a_{i,r} \cdot x_{i,t}}$$

Two possibilities are considered most often, namely assuming that $r=s$ or $r=t$. In the first eventuality, the index derived,

$$I_{t,s}(x) = \frac{\sum_{i=1}^{N} a_{i,t} \cdot x_{i,s}}{\sum_{i=1}^{N} a_{i,t} \cdot x_{i,t}}$$

is a Laspeyres index, and it can alternatively be written as a weighted sum of elementary indices $I_{t,s}(x_i)$ with weights $w_{i,t}$ where

$$w_{i,t} = \frac{a_{i,t} \cdot x_{i,t}}{\sum_{i=1}^{N} a_{i,t} \cdot x_{i,t}}$$

or as an harmonic weighted average of those same indices with weights $v_{i,t,s}$ where

$$v_{i,t,s} = \frac{a_{i,t} \cdot x_{i,s}}{\sum_{i=1}^{N} a_{i,t} \cdot x_{i,s}}$$

The alternative is to have $s=t$, and then the index obtained is a Paasche index:

$$I_{t,s}(x) = \frac{\sum_{i=1}^{N} a_{i,s}\, x_{i,s}}{\sum_{i=1}^{N} a_{i,s}\, x_{i,t}}$$

The Paasche index can also be expressed in terms of its components elementary indices since

$$I_{t,s}(x) = \left[\frac{\sum_{i=1}^{N} a_{i,t}\, x_{i,s}}{\sum_{i=1}^{N} a_{i,t}\, x_{i,t}} \right]^{-1}$$

and since the term is bracketed it is identical to a Laspeyres index, t and s being inverted,

$$I_{t,s}(x) = \left\{ \sum_{i=1}^{N} \left(\frac{a_{i,s}\, x_{i,s}}{\sum_{i=1}^{N} a_{i,s}\, x_{i,s}} \right) I^{-1}_{t,s}(x_i) \right\}^{-1}$$

i.e., the Paasche index is a harmonic weighted average of the elementary indices, the weights being those of the current period s.

Using the results obtained earlier, the weights of the arithmetic mean form of the Paasche index, $w_{i,t,s}$ depend upon both time periods since

$$w_{i,t,s} = \frac{a_{i,s}\, x_{i,t}}{\sum_{i=1}^{N} a_{i,s}\, x_{i,t}}$$

in

$$I_{t,s}(x) = \sum_{i=1}^{N} w_{i,t,s} \, I_{t,s}(x_i)$$

To conclude this section, the last property to be considered is that of multiplication: the elementary index of a product (ratio) is the product (ratio) of the elementary indices:

$$I_{t,s}(x \cdot y) = I_{t,s}(x) \cdot I_{t,s}(y)$$

$$I_{t,s}(x/y) = I_{t,s}(x)/I_{t,s}(y)$$

V.3 Laspeyres and Paasche Indices

In the preceding section, elementary indices were introduced. As a result of the analysis, while considering the possibility that the variable for which an elementary index was to be constructed, x_t, was itself the linear combination of a set of variables $x_{i,t}$, Laspeyres and Paasche indices were introduced.

It was shown that, in terms of the elementary indices $I_{t,s}(x_1)$, while the Laspeyres index was an arithmetic mean, the Paasche index was an harmonic mean. It follows that the Paasche index of a set of variables would be expected to be smaller than the Laspeyres one. Even though the harmonic mean is smaller than the arithmetic mean, it does happen that the Paasche index is greater than the Laspeyres index. This would follow from the fact that the two weighting schemes are distinct, the Paasche's being based on current shares, while the Laspeyres weights are based on base period shares. If this were to happen, it would have to be the case that, on average, the weights of those elementary indices which are the highest would have to be those which, from period t to period s, increase the fastest. This is in fact what happens in the earlier example of a 1979 price index using 1977 as base period for individual residential main line/contract primary: while the Laspeyres price index is 1.0492, the Paasche one is 1.501.[8] In this case it seems that this can be attributed to the dominant rate group, group 12 for Montreal and Toronto which from 1977 to 1979 experienced both a slightly smaller than average rate increase and a relative and absolute decrease in terms of number of individual main lines (by 8%).

In economics, the Paasche index is expected to be smaller than the Laspeyres index. This follows from the fact that while $x_{i,t}$ would alternatively be $p_{i,t}$ or $q_{i,t}$, $a_{i,t}$ will then be $q_{i,t}$ or $p_{i,t}$ accordingly, and that price changes and quantity changes are expected, usually, to be negatively correlated. In fact the divergence between Paasche and Laspeyres indexes has been studied by von Borthiewicz (1924) who show that, where the divergence between both indices, D, defined as

$$D = \frac{P_{t,s}(x) - L_{t,s}(x)}{L_{t,s}(x)}$$

where $P_{t,s}(x)$ denotes the Paasche index of x_s given the base period t

$L_{t,s}(x)$ denotes the corresponding Laspeyres index,

$$D = \frac{r\sigma_a\sigma_x}{L_{t,s}(a)}$$

where

$$r = \frac{1}{n^2} \sum_{j=1}^{n} \left[\frac{a_{j,s}}{a_{j,t}} - \frac{1}{n} \sum_{\ell=1}^{n} \left(\frac{a_{\ell,s}}{a_{\ell,t}} \right) \right] \left[\frac{p_{j,s}}{p_{j,t}} - \frac{1}{n} \sum_{k=1}^{n} \left(\frac{p_{k,s}}{p_{k,t}} \right) \right]$$

i.e. where r is the correlation coefficient between the elementary indices $I_{t,s}(a_j)$ and $I_{t,s}(p_j)$

σ_a is the variance of a_j's elementary indices, $I_{t,s}(a_j)$

and σ_x is the variance of x_j's elementary indices, $I_{t,s}(x_j)$

Economic theory tells us that, where the commodities considered are normal, r is expected to be negative, a relative price increase being expected to generate a relative decrease in the quantity demanded. Since σ_a and σ_x are squares and since $L_{t,s}(a)$ is a quantity index, they are all positive and D's sign is determined by r.

The expression is in fact very useful. As both Laspeyres and Paasche indices can be considered, as a first approximation, bounds to most price indices (Fisher, 192 ; Diewert, 1979), it gives us a general idea of how sensitive the computed index will be to the formula selection. Given a telecommunications output price index, σ_p should be very small, in general, for toll services, but not necessarily small for local services where say, in Bell Canada, recent increases for business rates have been considerably higher than corresponding residential rate increases. On the other hand σ_g should be relatively smaller in local services, especially in terms of residential services, than in message toll. Finally, r may be relatively high in terms of message toll, given the relatively high estimated demand elasticity (of the order of .3, as estimated by Bell in Bell Exhibit B.78-545 for intra-Bell services, of the order of 1 as estimated for intra-B.C. Tel. (Dreessen 1978), and of the order of 1.35 for all of Bell Canada's message toll services as estimated by Breslau and Smith (1979). However r will be relatively low for the greatest share of local services, especially for resident. It should be noted that the r obtained this way does not solely reflect demand effects, thus, in the earlier example the decrease in the number of main lines in rate group 12 is not associated with a shift in demand: residents do not move from one rate group to another as a result of a rate change.

Even though neither Laspeyres and Paasche indices are time reversible, there exists interesting relationships between them since

$$P_{t,s}(x) = L^{-1}_{s,t}(x) \quad ,$$

$$L_{t,s}(x) = P^{-1}_{s,t}(x) \quad .$$

Furthermore, both indices share another property which is particulary attractive, namely the fact that a Laspeyres (Paasche) index of Laspeyres (Paasche) indices of elementary indices is a Laspeyres (Paasche) index of those elementary indices.

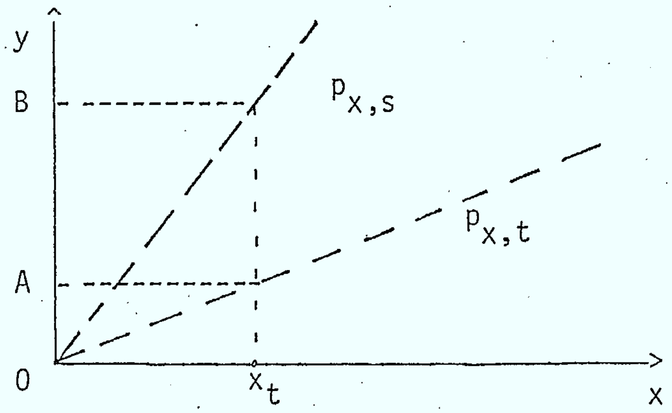
It is interesting to note that, where $a_{i,j,t}\, x_{i,j,t}$ is the expenditure on the (i,j) commodity, the form of the Laspeyres index of Laspeyres indices are, if say $x_{i,j,t}$ is the price of the (i,j) commodity, based on the product of the expenditure times the price. This does not raise any problem because this is measured at the base period t, hence the price corresponding to the quantity would be one.

V.4 The Geometric Analysis of Index Numbers

We have seen that, given a commodity, the change in price can be described by an elementary index. Such an elementary index can easily be illustrated with the following diagram, where the quantity of the good or service is represented by x on the abscissa while the money budget is represented by y on the ordinate. The price line is given by

$$y = p_x \cdot x$$

i.e. by a line passing through the origin, with slope p_x.



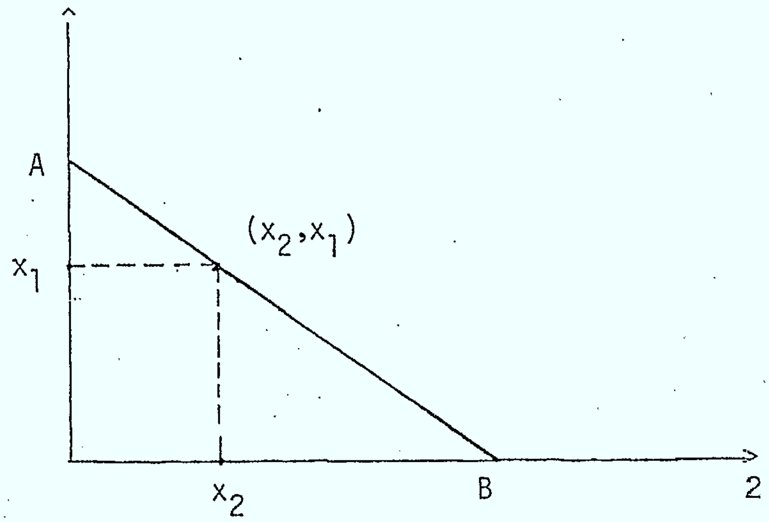
Given a quantity x, say 4, and given the price of x in the same period t, $p_{x,t}$ = \$.5, then the expenditure on x in t is \$2. If in period s the price rises to $p_{x,s}$ = \$1.5, to obtain the original 4 units of x, one needs a budget of \$6, i.e. the expenditure on the same quantity of x in s will be \$6. The elementary index is given by

$$\left(\frac{p_{x,s}}{p_{x,t}}\right) = \left(\frac{p_{x,s} \cdot x}{p_{x,t} \cdot x}\right) = \left(\frac{0B}{0A}\right)$$

We may now consider the situation in which there are two commodities, x_1 and x_2. Let x_2 be represented on the abscissa and x_1 on the ordinate. In period t, given the prices $p_{1,t}$ and $p_{2,t}$, x_2 can be traded for x_1. Then the endowment measured in terms of commodity 1, to $[x_1 + (p_{2,t}/p_{1,t})x_2]$ which will be denoted by A. In terms of commodity 2, it would be $[(p_{1,t}/p_{2,t})x_1 + x_2]$, which will be denoted by B. The line which goes through A, B and (x_2,x_1) is the budget line, it is given by

$$y_t(1) = x_1 + (p_{2,t}/p_{1,t})x_2$$

the 1 in $y_t(1)$ denoting the fact that the revenue y_t is measured in terms of commodity 1, commodity 1 being the numeraire.

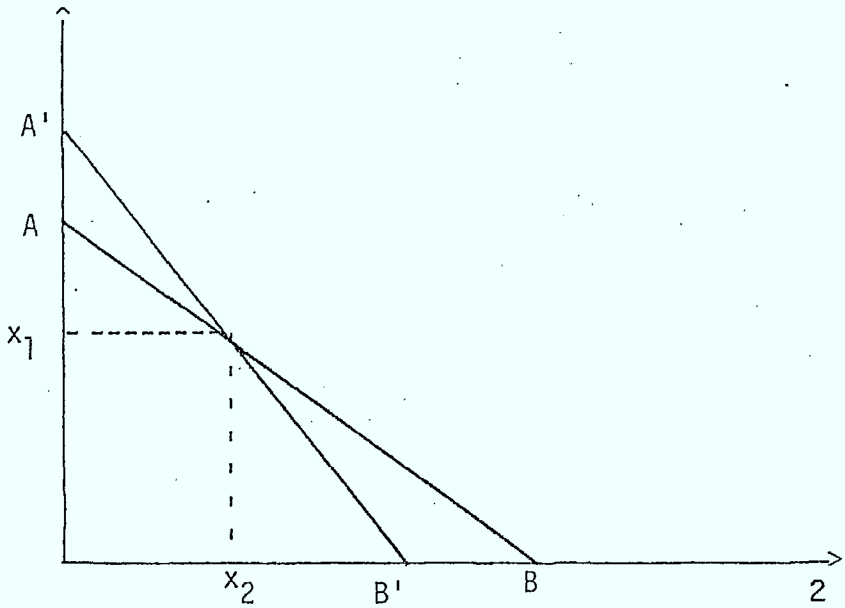


Let's assume that from period t to period s, the prices go from $p_{1,t}$ and $p_{2,t}$ to $p_{1,s}$ and $p_{2,s}$ respectively. In terms of x_1 and x_2 two budgets can be considered, y_t and y_s where

$$y_t = p_{1,t} \cdot x_1 + p_{2,t} \cdot x_2$$

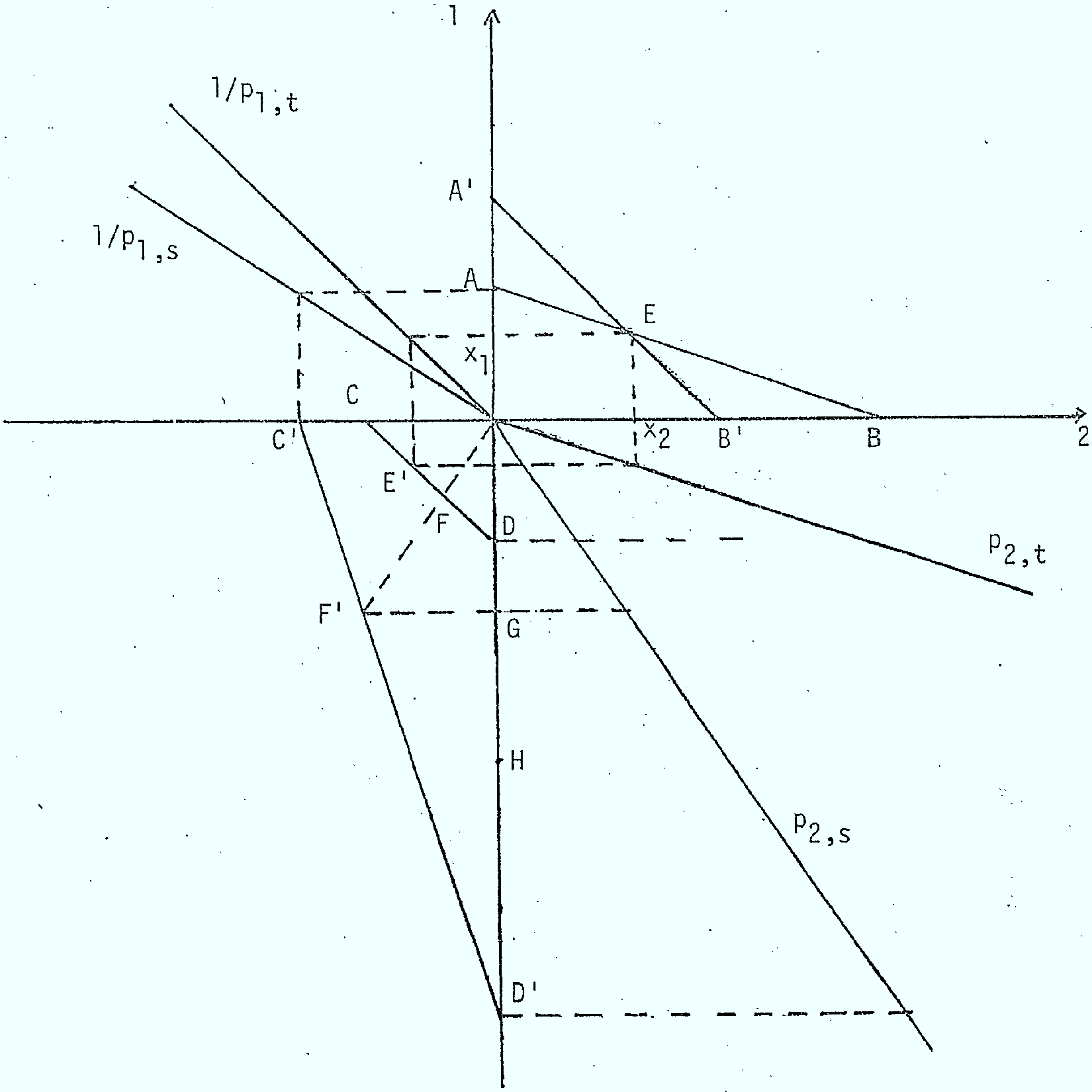
$$y_s = p_{1,s} \cdot x_1 + p_{2,s} \cdot x_2$$

Similarly, two budget lines can be considered, $\overline{AB}$ corresponding to y_t , with intercepts $[x_1 + (p_{2,t}/p_{1,t})x_2]$, and $[p_{1,t}/p_{2,t})x_1 + x_2]$ respectively and $A'B'$ corresponding to y_s , with intercepts $[x_1 + (p_{2,s}/p_{1,s})x_2]$, and $[(p_{1,s}/p_{2,s})x_1 + x_2]$



To obtain a diagramatic representation of aggregate price indices, it suffices to merge both this diagram showing the relative prices with the diagrams of the elementary indices. Let's measure in the positive quadrant the quantitites of commodities 1 and 2. Let's represent in the N.W. quadrant commodity 1's price lines and in the S.E. quadrant commodity 2's price lines.

Let's assume $x_1 = 2$, $x_2 = 3$, $p_{1,t} = \$1$, $p_{2,t} = \$1/3$, then $y_t = \$3$. A, B and (x_2,x_1) are equivalent, and 3 units of commodity 1 yields $\overline{OC}$, i.e. \$3, as does 9 units of commodity 2, yielding $\overline{OD}$. $\overline{CD}$ is the transformation of $\overline{AB}$ in taking into account the nominal value of both commodities. E' is the mapping of E , E' corresponds to the \$2


1
1/p1,t
1/p1,s
A'
A
E
x1
C
C'
x2
B'
B
2
E'
F
D
p2,t
F'
G
H
p2,s
D'

necessary to purchase two units of commodity 1 and to the \$1 needed to buy the three units of commodity 2.

Now if one considers how much $\overline{OA}$ costs in period s, one obtains $\overline{OC}'$, the rate of inflation in terms of commodity 1 being $\left(\frac{OC'}{OC}\right) = \left(\frac{4.5}{3}\right)$ $= 1.5$. $\left(\frac{OC'}{OC}\right)$ is in fact $I_{t,s}(p_1)$, the elementary index of p_1, and, since it does not take p_2 into consideration, given that there are only two commodities, it can be seen as a bound - here, the lower bound - to any price indices. Similarly, starting from commodity 2, while $\overline{OB}$ had cost $\overline{OD}$ in period t, it now costs, in period s, $\overline{OD}' = \$13.5$. $\left(\frac{OD'}{OD}\right)$ is in fact $I_{t,s}(p_2)$, and since it is the greatest elementary index, it can be seen as the upper bound of any price index.

In fact, both commodities were included in the budget hence both prices and both elementary indices are relevant to describe the price increase. x_2 at the old price $p_{2,t}$ implies an expenditure of $(p_{2,t} \cdot x_2)$, Weighting the elementary price index $I_{t,s}(p_2)$ by this factor - and, for simplicity taking $\overline{OD} = \$3 = (p_{1,t} \cdot x_1 + p_{2,t} \cdot x_2)$ as unit of measurement - the point OG corresponding to $(p_{2,t} \cdot x_2) I_{t,s}(p_2)$ is determined. Adding to it commodity 1's elementary price index $I_{t,s}(p_1)$ weighted by the (relative) expense on that commodity, $(p_{1,t} \cdot x_1)$, the point F' is obtained. F' corresponds to a weighted combination of (OC'/OC) and (OD'/OD) with as weights the share of the expenditures, in terms of period t prices, allocated to each commodity. It can be projected on either axis through a budget line $\overline{HH}$. A new price index, (OH/OD) is obtained which is bounded by x_1's elementary price index (OC'/OC) and by x_2's elementary price index (OD'/OD).

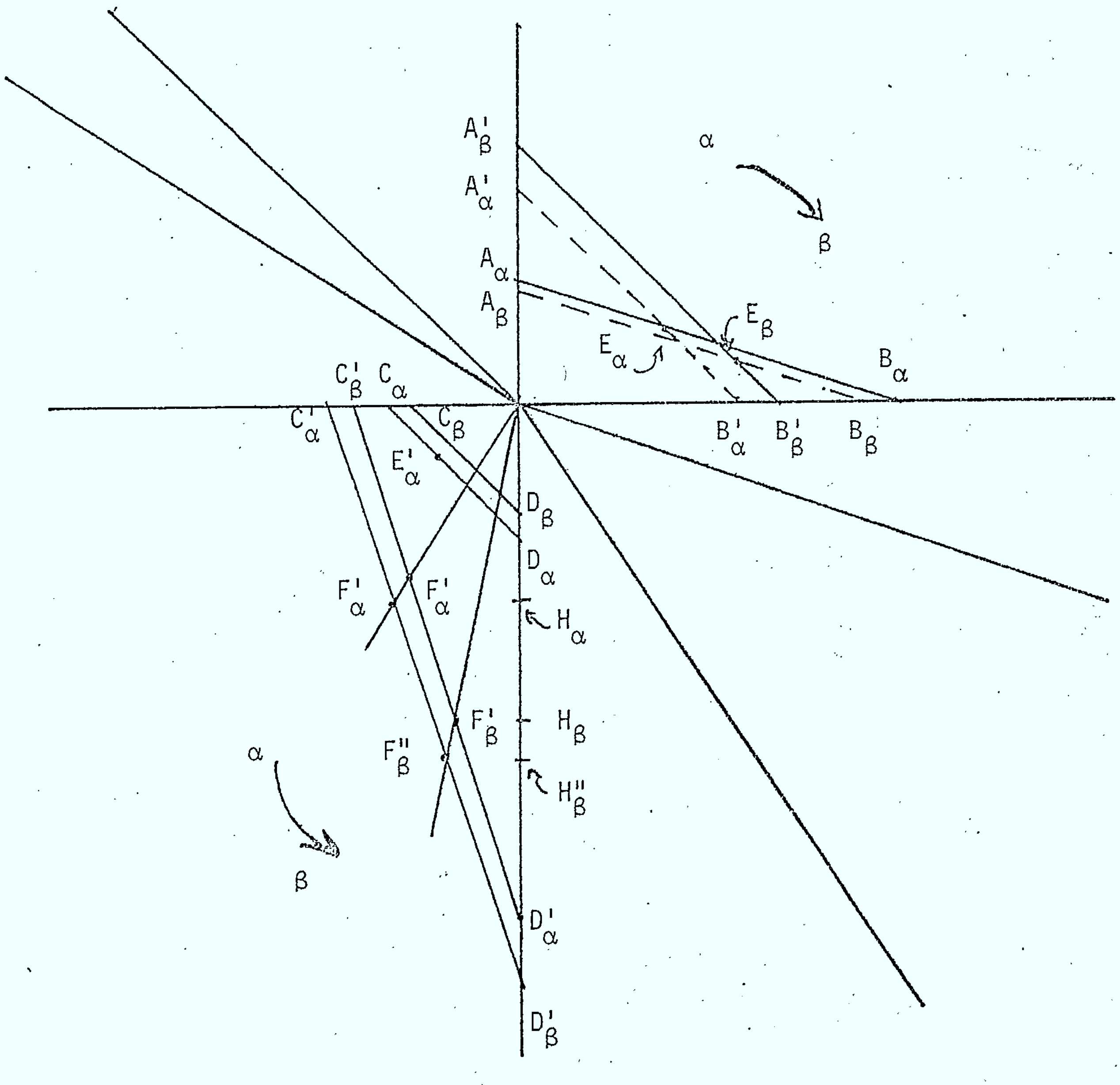
Until now the quantities x_1 and x_2 had not been situated in time.

One possibility would be to date them in period t, then the index obtained is the Laspeyres index since (OH/OD) is the weighted average of the two elementary indices with the shares of the expenditures in period t spent on each commodity as weights.

Alternatively, where $x_1 = x_{1,s}$ and $x_2 = x_{2,s}$, then (OH/OD) would be the arithmetic weighted mean of both elementary price indices, the weights being the share of the expenditures determined in terms of period t's prices, i.e. (OH/OD) would be the Paasche index.

In fact given quantities $(x_{i,t}, x_{2,t})$ in period t and $(x_{1,s}, x_{2,s})$ in period s, there is no reason that the quantities should be only in terms of t or s. A possibility is to average the quantities in period s and t to yield quantities of $x_1 = \frac{1}{2}(x_{1,t} + x_{1,s})$ and $x_2 = \frac{1}{2}(x_{2,t} + x_{2,s})$. In this case (OH/OD) will be an Edgeworth index.

The last question to arise is to determine the relative positions of such indices with respect to one another. We may begin by observing that any other quantities of each commodity, x_1', and x_2', such that x_1' and x_2' determine a point in the positive quadrant on the line OE $(x_1'/x_2' = x_1/x_2)$, will determine an index. Given $p_{1,\alpha}$, $p_{2,\alpha}$, $p_{1,\beta}$ and $p_{2,\beta}$ the index determined will also be (OH/OD), i.e. the index is independent of scale. Then it suffices to determine what happened to the index when $(x_1'/x_2') \lessgtr (x_1/x_2)$. Let E_α correspond to (x_1, x_2) and (x_1', x_2') to E_β, such that $(x_1'/x_2') < (x_1/x_2)$. Then indices (OH_α/OD_α) and (OH_β/OD_β) are derived. To compare them it suffices to consider, say H_β'' which corresponds to H_β rescaled to be consistant with OD_α. It follows immediately that $(OH_\beta''/OD_\alpha) > (OH_\alpha/OD_\alpha)$, i.e. that the index is inversely related to the ratio (x_1/x_2).


A'β
A'α
Aα
Aβ
α
β
Eβ
Eα
Bα
C'β
Cα
C'α
Cβ
B'α
B'β
Bβ
E'α
Dβ
Dα
F'α
F'α
Hα
F'β
Hβ
F"β
H"β
α
β
D'α
D'β

Now if x_1 and x_2 represent quantity demanded by consumers and if p_2 increases more than proportionately than p_1, then consumers which maximize utility will increase their relative demand for commodity 1 and $(x_1'/x_2') > (x_1/x_2)$ implies that the index based on (x_1',x_2') will be smaller than that based on the original quantities (x_1,x_2). Since the former is a Paasche index and the latter is a Laspeyres index, we obtain the common result in the cost of living index literature, namely that normally a Paasche index is smaller than the Laspeyres index. However, if E_α represents, the output combination of a producer, a more than proportional increase in commodity 2 will lead a profit maximizing producer to increase more than proportionately its output of the commodity which is becoming dearer. Then we obtain Fisher and Schell's result that the Paasche index, based in the diagram on E_β, will be greater than the Laspeyres index, based on E_α. In all these cases, the Edgeworth index will be bounded by both the Laspeyres and the Paasche indices. The Fisher ideal index which is a geometric mean of the Laspeyres and the Paasche indices will itself be smaller than the Edgeworth index, an arithmetic mean of those two indices, and it will be bounded by the Edgeworth index and the smaller of the Laspeyres and Paasche indices.

V.5 The Making of Index Numbers

In the preceding sections elementary indices were introduced and two particular forms of indices which could alternatively be seen as elementary index of an aggregate x_t or as synthetic index of elementary indices of the components. This suggests a first approach to the construction of indices, in which indices are constructed alternatively as average of elementary indices and elementary indices of average weighted values $x_{i,t}$. First we may consider a general class of indices which includes as a particular case both the Laspeyres and the Paasche index. For this, we observe that both Laspeyres and Paasche indices could be considered in terms of the theory of averages (introduced in the appendix) as related to harmonic and arithmetic means, and we generalize by using the r-mean as applied to either formulation.

Given a set of variables, $x_{i,t}$ and a corresponding set of weights, $a_{i,t,s}$, the r-mean of the $x_{i,t}$'s, x_t^r can be calculated as

$$x_t^r = \left\{ \sum_{i=1}^{N} a_{i,t,s} \; x_{i,t}^r \right\}^{1/r} .$$

An elementary index can be derived from these r-means:

$$I_{t,s}\,(x^r) = \left(\frac{x_s^r}{x_t^r}\right)$$

Earlier results yield

$$I_{t,s}(x^r) = \left\{ \left(\frac{\sum_{i=1}^{N} a_{i,t,s} \; x_{i,s}^r}{\sum_{i=1}^{N} a_{i,t,s} \; x_{i,t}^r} \right) \right\}^{1/r}$$

$$= \left\{ \sum_{i=1}^{N} w_{i,t,s} \quad I^{r}_{t,s} \; (x_i) \right\}^{1/r}$$

$$= \left\{ \sum_{i=1}^{N} v_{i,t,s} \quad I^{-r}_{t,s} \; (x_i) \right\}^{-1/r}$$

where

$$w_{i,t,s} = \frac{a_{i,t,s} \; x_{i,t}^{r}}{\sum_{i=1}^{N} a_{i,t,s} x_{i,t}^{r}}$$

$$v_{i,t,s} = \frac{a_{i,t,s} \; x_{i,s}^{r}}{\sum_{i=1}^{N} a_{i,t,s} \; x_{i,s}^{r}}$$

The generalized r-mean Laspeyres index will be given by restricting $a_{i,t,s}$ to $a_{i,t}$ and the corresponding generalized r-mean Paasche index will be given by restricting $a_{i,t,s}$ to $a_{i,t}$.

It could be indicated that as $r \Rightarrow 0$, at the limit, the o-mean is the geometric mean and the index will be given as a ratio of Cobb-Douglas forms.

The impact of various r is to give more or less emphasis to the greater elementary indices relatively to the smaller ones. Given the mean selected, the next problem to be handled is the weight schemes. Again it is possible to work in terms of Laspeyres and Paasche methods.

i) <u>Laspeyres weights</u> $a_{i,t}$

$$w^{r}_{i,t} = \frac{a^{r}_{i,t} \; x^{r}_{i,t}}{\sum_{i=1}^{N} a^{r}_{i,t} \; x^{r}_{i,t}}$$

$$v^r_{i,t,s} = \frac{a^r_{i,t}\ x^r_{i,s}}{\sum_{i=1}^{N} a^r_{i,t}\ x^r_{i,s}}$$

ii) Paasche weights $a_{i,s}$

$$w^r_{i,t,s} = \frac{a^r_{i,s}\ x^r_{i,t}}{\sum_{i=1}^{N} a^r_{i,s}\ x^r_{i,t}}$$

$$v_{i,s} = \frac{a^r_{i,s}\ x^r_{i,s}}{\sum_{i=1}^{N} a_{i,s}\ x^r_{i,s}}$$

iii) Edgeworth weights $1/2\ (a_{i,t} + a_{i,s})$

$$w_{i,t,s} = \frac{(a_{i,t} + a_{i,s})^r x^r_{i,t}}{\sum_{i=1}^{N} (a_{i,t} + a_{i,s})^r x^r_{i,t}}$$

$$v_{i,t,s} = \frac{(a_{i,t} + a_{i,s})^r x^r_{i,s}}{\sum_{i=1}^{N} (a_{i,t} + a_{i,s})^r x^r_{i,s}}$$

The aim, in developing an index number, is to somehow define some "typical-average" elementary index number. The added condition that such a "typical-average" elementary index number be itself representable under the form of an elementary index of average components was a condition fulfilled by both Paasche and Laspeyres indices. Even though it has some advantages, it is fruitful to consider a wider class of indices, including indices which do not fulfill this condition, by centering one's attention on the index as an average of elementary indices. In this case we may define the generalized (r,u)-mean index as the r-mean index with weights in terms of the exponent u . Denoting it by $J^{r,u}_{t,s}$, then

$$J^{r,u}_{t,s} = \left\{ \sum_{i=1}^{N} w^{u}_{i,t,s} \; I^{r}_{t,s}(x_i) \right\}^{1/r}$$

This generalization appears to be particularly attractive for two values of u . The first is u = r , in which case we obtain the generalized r-mean Laspeyres, Paasche and Edgeworth indices together with that class of indices which can be represented as an (1/r)-elementary index of r-means. The other interesting value of u is 1 because given economics where either "a" or "x" will denote price, the other denoting quantities, w_i will denote some sort of budget share going to commodity i . Since $I^{r}_{t,s}(x_i) = I^{-r}_{s,t}(x_i)$ by reversability, it is still true that, denoting by $J^{r,u}_{t,s}(u)$ the index the weights of which depends upon period u ,

$$J^{r,u}_{t,s}(t) = \left[J^{-r,u}_{s,t}(t) \right]^{-1}$$

i.e., that the (r,u)-mean index base s at t is the inverse of the (-r,u)-mean index base t at s .

Given the proposed generalization, the u-weights previously presented can also be used as w-weights, since the share of expenditures going to commodity i can be determined in terms of prices in period t or in period s , or ...

(iv) cross-Laspeyres weights

$$w^r_{i,t,s} = \frac{a^r_{i,t}\, x^r_{i,s}}{\sum_{i=1}^{N} a^r_{i,t}\, x^r_{i,s}}$$

(v) cross-Paasche weights

$$w^r_{i,t,s} = \frac{a^r_{i,s}\, x^r_{i,t}}{\sum_{i=1}^{N} a^r_{i,s}\, x^r_{i,t}}$$

(vi) cross-Edgeworth weights

$$w^r_{i,t,s} = \frac{(a_{i,t} + a_{i,s})^r x^r_{i,s}}{\sum_{i=1}^{N} (a_{i,t} + a_{i,s})^r x^r_{i,s}}$$

To this we can add one more class of weights by combining Laspeyres and Paasche weights:

(vii) Tornqvist weights

$$w^r_{i,t,s} = \tfrac{1}{2} \left\{ \frac{a^r_{i,t}\, x^r_{i,t}}{\sum_{i=1}^{N} a^r_{i,t}\, x^r_{i,t}} + \frac{a^r_{i,s}\, x^r_{i,s}}{\sum_{i=1}^{N} a^r_{i,s}\, x^r_{i,s}} \right\}$$

(viii) cross-Tornqvist weights

$$w^r_{i,t,s} = \tfrac{1}{2} \left\{ \frac{a^r_{i,t}\, x^r_{i,s}}{\sum_{i=1}^{N} a^r_{i,t}\, x^r_{i,s}} + \frac{a^r_{i,s}\, x^r_{i,t}}{\sum_{i=1}^{N} a^r_{i,s}\, x^r_{i,t}} \right\}$$

The major problem of all single indices but the (r,r)-mean indices is their failure to meet the time reversability criterion. The latter criterion implies

$$J^{r,u}_{t,s} \cdot J^{r,u}_{s,t} = 1$$

However as

$$J^{r,u}_{s,t} = (J^{-r,u}_{t,s})^{-1}$$

one would have

$$J^{r,u}_{t,s} \cdot (J^{-r,u}_{t,s})^{-1} = 1$$

which would imply, raising terms to the r-th power, that

$$\sum_{i=1}^{N} w^u_{i,t,s} \; I^r_{t,s} (x_i) = \{ \sum_{i=1}^{N} w^u_{i,t,s} \; I^{-r}_{t,s} (x_i) \}^{-1}$$

Even if $w_{i,t,s} = w_{i,s,t}$ the left-hand side is the arithmetic mean of $I^r_{t,s} (x_i)$ while the right-hand side is the harmonic mean of the same term. Now unless, $I^r_{t,s} (x_i)$ is a constant independent of x_i, it is shown in the appendix that the harmonic mean is smaller than the arithmetic mean provided all the elementary indices are sensitive to which of the two periods, t or s, is chosen as a base.

The one other exception is the particular case of the index which is obtained as r_o tends toward zero with weights symmetric about time, such that $w_{i,t,s} = w_{i,s,t}$ since then one has

$$\lim_{r \to 0} J^{r,u}_{t,s} = \prod_{i=1}^{N} I_{t,s} (x_i)^{w^u_{i,t,s}}$$

and denoting that limit by $J^{o,u}_{t,s}$, since then $w^u_{i,t,s} = w^u_{i,s,t}$,

$$J^{o,u}_{t,s} \circ J^{o,u}_{s,t} = 1$$

Given the class of $J^{o,u}_{s,t}$, i.e. given the sets of possible weights it would seem simplest to select the Tornqvist weights, i.e.

$$w^u_{i,t,s} = \tfrac{1}{2} \left\{ \frac{a^u_{i,t} \, x^u_{i,t}}{\sum_{i=1}^{N} a^u_{i,t} \, x^u_{i,t}} + \frac{a^u_{i,s} \, x^u_{i,s}}{\sum_{i=1}^{N} a^u_{i,s} \, x^u_{i,s}} \right\}$$

Whenever $u = 1$, this is Fisher's index 124 which, in view of modern usage, may be called Fisher's Tornqvist index.

V.6 Ideal Indices: Reversability

As long as one doesn't restrict the analysis to that of the limit when r tends to zero, one would like to develop a time and factor reversible index, i.e. an index $K^r_{t,s}$ such that

$$K^r_{t,s} \cdot K^r_{s,t} = 1$$

$$K^r_{t,s}(p) \cdot K^r_{t,s}(q) = \frac{\sum_{i=1}^{N} p_{i,s}\, q_{i,s}}{\sum_{i=1}^{N} p_{i,t}\, q_{i,t}}$$

Now, given the equality between $J^r_{t,s}(t)$ and $J^{-r}_{s,t}(t)$, it must be the case that if $J^r_{t,s}(t)$ is an element of $K^r_{t,s}$, then $J^{-r}_{t,s}(t)$ must be an element of $K^r_{s,t}$, i.e., $J^r_{s,t}(t)$ must be an element of $K^r_{s,t}$. In this case, both $J^r_{t,s}(t)$ and $J^r_{t,s}(s)$ must be elements of $K^r_{s,t}$.

Either one of two possibilities may arise. First of all, it may be that in fact

$$J^r_{t,s}(t) = J^r_{t,s}(s)$$

this will happen, given the definition of $J^r_{t,s}(u)$:

$$J^r_{t,s}(u) = \left\{ \frac{\sum_{i=1}^{N} a_i(u)\, x^r_{i,s}}{\sum_{i=1}^{N} a_i(u)\, x^r_{i,t}} \right\}^{1/r}$$

whenever $a_i(t) = a_i(s)$. But for the uninteresting case where a_i is in fact independent of time, it may be observed that the condition is not met with the Lespeyres or the Paasche-type of weights, but that, on the other hand, it is met by Edgeworth and Tornqvist-type of weights.

Given

$$a_i(t) = a_i(s)$$

it is sufficient to set

$$K^r_{t,s} = J^r_{t,s}$$

Two examples would be

$$\text{(i)} \qquad \left\{ \sum_{i=1}^{N} \left[\frac{q_{i,t}\, p_{i,t}}{\sum_{i=1}^{N} p_{i,t}\, q_{i,t}} + \frac{p_{i,s}\, q_{i,s}}{\sum_{i=1}^{N} p_{i,s}\, q_{i,s}} \right] \left(\frac{p_{i,s}}{p_{i,t}} \right)^r \right\}^{1/r}$$

$$\text{(ii)} \qquad \left\{ \frac{\sum_{i=1}^{N} (q_{i,t} + q_{i,s}) p^r_{i,s}}{\sum_{i=1}^{N} (q_{i,t} + q_{i,s}) p^r_{i,s}} \right\}^{1/r}$$

Now whenever $a_i(t) \neq a_i(s)$, then both $J^r_{t,s}(t)$ and $J^r_{t,s}(s)$ must be part of $K^r_{t,s}$, hence

$$K^r_{t,s} = \{J^r_{t,s}(t) \cdot J^r_{t,s}(s)\}^{\frac{1}{2}}$$

i.e. $K^r_{t,s}$ is a simple geometric average of the 2-simple index with the (-r)-simple one. In general

$$K^r_{t,s} = \left\{ \frac{\sum_{i=1}^{N} w^u_{i,t} I^r_{t,s}(x_i)}{\sum_{i=1}^{N} w^u_{i,s} I^{-r}_{t,s}(x_i)} \right\}^{1/2r}$$

Given $u = 1$, this is exactly Diewert's "Quadratic Mean of Order (2r) Indices".

Alternatively, by using different sets of weights, one would have

$$K^r_{t,s} = \left[J^r_{t,s}(s)\ J^{-r}_{t,s}(t)\right]^{1/2}$$

$$K^r_{t,s} = \left[J^r_{t,s}(t,s)\ J^{-r}_{t,s}\ (t,s)\right]^{1/2}$$

depending upon the choice between Laspeyres, Paasche weights.

In the eventuality $r = u = 1$, then the numerator of the first $K^r_{t,s}$ considered will be a Laspeyres index, the denominator being the inverse of a Paasche index, and the geometric mean of the two is Fisher's ideal index.

However, the fact or reversal test trivally requires $u = r$ and given $u = 1$ only if $r = 1$ will the factor reversal test be fulfilled, i.e. given P_s the index obtained for $x_{i,t} = P_{i,t}$ and Q_s that generated by $x_{i,t} = q_{i,t}$, will it be true that

$$P_s \cdot Q_s = \frac{\sum_{i=1}^{N} p_{i,s}\ q_{i,s}}{\sum_{i=1}^{N} p_{i,t}\ q_{i,t}}$$

It is of interest, at this stage, to see how a class of indices which also meet the factor reversal test can be created. Earlier a class of indices of a r-mean of the component variables was introduced:

$$J^r_{t,s}(x) = \left(\frac{\sum_{i=1}^{N} a_{i,t,s}\, x^r_{i,s}}{\sum_{i=1}^{N} a_{i,t,s}\, x^r_{i,t}} \right)^{1/r}.$$

Provided $a_{i,t,s}$ be either independent of t and s or symmetrical with respect to t and s, i.e. $a_{i,t,s} = a_{i,s,t}$, the index is elementary and as such it meets the time reversal condition. Whenever $a_{i,t,s}$ does not meet this criterion, an index which will meet the time reversal test is constructed as $K^r_{t,s}(x)$ where

$$K^r_{t,s}(x) = \left(\frac{\left[\sum_{i=1}^{N} a_i(t)\; x^r_{i,s}\right] \left[\sum_{i=1}^{N} a_i(s)\; x^r_{i,s}\right]}{\left[\sum_{i=1}^{N} a_i(t)\; x^r_{i,t}\right] \left[\sum_{i=1}^{N} a_i(s)\; x^r_{i,t}\right]} \right)^{1/2r}$$

The factor reversal test implies that an index of quantity times the index of price is equal to the ratio of total expenditures between periods t and s, i.e.

$$K^r_{t,s}(q)\; K^r_{s,t}(p) = \frac{\sum_{i=1}^{N} p_{i,s}\, q_{i,s}}{\sum_{i=1}^{N} p_{i,t}\, q_{i,t}}$$

Then a possible approach is to start from indices of p and q, $J^r_{t,s}(x)$ where $a_{i,t,s}$ is alternatively $q^r_{i,t}$, $q^r_{i,s}$, $p^r_{i,t}$ and $p^r_{i,s}$, and to weight this index by the ratio of the r-th root of the inner product of the $\underset{\sim}{p}^r$ and $\underset{\sim}{q}^r$ vectors by the inner product of the $\underset{\sim}{p}$ and $\underset{\sim}{q}$ vectors, i.e. indices such as

$$J^r_{t,s}(q) = \left\{ \lambda^r_t \frac{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,s}}{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,t}} \right\}^{1/r} = \frac{\{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,s}\}^{1/r}}{\sum_{i=1}^{N} p_{i,t}\, q_{i,t}}$$

where

$$\lambda = \frac{\{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,t}\}^{1/r}}{\sum_{i=1}^{N} p_{i,t}\, q_{i,t}}$$

A general class of ideal indices, i.e. indices which meet both the time reversal test and the factor reversal test, can now be generated:

$$K^r_{t,s}(q) = \left\{ \lambda^r_t \frac{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,s}}{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,t}} \cdot \lambda^{-r}_s \frac{\sum_{i=1}^{N} p^r_{i,s}\, q^r_{i,s}}{\sum_{i=1}^{N} p^r_{i,s}\, q^r_{i,t}} \right\}$$

where

$$\lambda^{-1}_s = \frac{\sum_{i=1}^{N} p_{i,s}\, q_{i,s}}{\{\sum_{i=1}^{N} p^r_{i,s}\, q^r_{i,s}\}^{1/r}}$$

that is

$$K^r_{t,s}(q) = \left\{ \left(\frac{\sum_{i=1}^{N} p_{i,s} q_{i,s}}{\sum_{i=1}^{N} p_{i,t} q_{i,t}} \right) \left(\frac{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,s}}{\sum_{i=1}^{N} p^r_{i,s}\, q^r_{i,t}} \right)^{1/r} \right\}^{1/2}$$

$$K^r_{t,s}(p) = \left\{ \left(\frac{\sum_{i=1}^{N} p_{i,s} q_{i,s}}{\sum_{i=1}^{N} p_{i,t} q_{i,t}} \right) \left(\frac{\sum_{i=1}^{N} p^r_{i,s}\, q^r_{i,t}}{\sum_{i=1}^{N} p^r_{i,t}\, q^r_{i,s}} \right)^{1/r} \right\}^{1/2}$$

V.7 Divisia Indices

When considering a productive process, given the total revenue, R_t, one may ask whether there exists an index Q_t of outputs and an index P_t of the prices of outputs such that for some constant λ

$$Q_t \cdot P_t = \lambda R_t$$

i.e. after differentiation with respect to time,

$$\frac{\dot{Q}_t}{Q_t} + \frac{\dot{P}_t}{P_t} = \lambda \frac{\dot{R}_t}{R_t}$$

where $\dot{x}_t / x_t = d \ln x_t/dt$

However, by definition, given a set of outputs $q_{i,t}$ priced at $p_{i,t}$,

$$R_t = \sum_{i=1}^{N} p_{i,t}\, q_{i,t}$$

$$\frac{\dot{R}_t}{R_t} = \frac{\sum_{i=1}^{N} \{p_{i,t}\, \dot{q}_{i,t} + q_{i,t}\, \dot{p}_{i,t}\}}{\sum_{i=1}^{N} p_{i,t}\, q_{i,t}}$$

$$= \sum_{i=1}^{N} \left(\frac{p_{i,t} q_{i,t}}{\sum_{i=1}^{N} p_{i,t} q_{i,t}} \right) \left(\frac{\dot{q}_{i,t}}{q_{i,t}} \right) + \sum_{i=1}^{N} \left(\frac{p_{i,t} q_{i,t}}{\sum_{i=1}^{N} p_{i,t} q_{i,t}} \right) \left(\frac{\dot{p}_{i,t}}{p_{i,t}} \right)$$

However then it can be noted that each sum on the right-hand side is a weighted sum of elementary quantity and price indices $\dfrac{\dot{q}_{i,t}}{q_{i,t}}$

and $\frac{\dot{p}_{i,t}}{p_{i,t}}$, the weight being the share of the revenue generated by commodity i, hence it is natural to set

$$\left(\frac{\dot{P}_t}{P_t}\right) = \sum_{i=1}^{N} \left(\frac{p_{i,t}q_{i,t}}{\sum_{i=1}^{N} p_{i,t}q_{i,t}}\right) \left(\frac{\dot{p}_{i,t}}{p_{i,t}}\right)$$

$$\left(\frac{\dot{Q}_t}{Q_t}\right) = \sum_{i=1}^{N} \left(\frac{p_{i,t}\, q_{i,t}}{\sum_{i=1}^{N} p_{i,t}q_{i,t}}\right) \left(\frac{\mathring{q}_{i,t}}{q_{i,t}}\right)$$

Price and quantity play a symmetrical and these indices meet Fisher's factor reversal test.

Similarly, if all revenues are return to a factor of production, then it may be asked whether there exists a price index W_t and a quantity index L_t such that

$$W_t \cdot L_t = R_t$$

The same procedure as the one followed for the outputs yields:

$$\left(\frac{\dot{L}_t}{L_t}\right) = \sum_{j=1}^{M} \left(\frac{w_{j,t}l_{j,t}}{\sum_{j=1}^{M} w_{j,t}l_{j,t}}\right) \left(\frac{\dot{l}_{j,t}}{l_{j,t}}\right)$$

$$\left(\frac{\dot{W}_t}{W_t}\right) = \sum_{j=1}^{M} \left(\frac{w_{j,t}\, l_{j,t}}{\sum_{j=1}^{M} w_{j,t}\, l_{j,t}}\right) \left(\frac{\dot{w}_{j,t}}{w_{j,t}}\right)$$

If total productivity is defined as the ratio of the index of outputs with respect to the index of inputs, then a duality between the quantity indices and the price indices is obtained. If TFP_t denotes total factor productivity, then

$$\frac{\dot{TFP}_t}{TFP_t} = \frac{\dot{Q}_t}{Q_t} - \frac{\dot{L}_t}{L_t} = \frac{\dot{P}_t}{P_t} - \frac{\dot{W}_t}{W_t} .$$

This approach to developing indices starts from the factor reversal property; hence answers originally that the index considered meets that property. However, in that case an index which meets this property would reduce over indefinitely small changes of time, to the Divisia index. Thus, if one starts with Fisher's ideal index taken in terms of a period s infinitely closed to the base period t, then the proportional change in, say, the quantity index Q_t will be, given

$$Q_t = \left[\frac{\sum_{i=1}^{N} p_{i,t} q_{i,s} \sum_{i=1}^{N} p_{i,s} q_{i,s}}{\sum_{i=1}^{N} p_{i,t} q_{i,t} \sum_{i=1}^{N} p_{i,s} q_{i,t}} \right]^{1/2}$$

$$\frac{\dot{Q}_t}{Q_t} = (1/2) \sum_{i=1}^{N} \left[\frac{p_{i,t} q_{i,s}}{\sum_{i=1}^{N} p_{i,t} q_{i,s}} + \frac{p_{i,s} q_{i,s}}{\sum_{i=1}^{N} p_{i,s} q_{i,s}} \right] \frac{\dot{q}_{i,t}}{q_{i,t}}$$

$$+ \sum_{i=1}^{N} \left[\frac{p_{i,s} q_{i,s}}{\sum_{i=1}^{N} p_{i,s} q_{i,s}} - \frac{p_{i,s} q_{i,t}}{\sum_{i=1}^{N} p_{i,s} q_{i,t}} \right] \frac{\dot{p}_{i,t}}{p_{i,t}}$$

Since then $p_{i,s} = p_{i,t}$ and $q_{i,s} = q_{i,t}$,

$$\frac{\dot{Q}_t}{Q_t} = \sum_{i=1}^{N} \frac{p_{i,t}q_{i,t}}{\sum_{i=1}^{N} p_{i,t}q_{i,t}} \quad \frac{\dot{q}_{i,t}}{q_{i,t}}$$

i.e. the Fisher ideal index reduces to a Divisia index.

If the commodities over which the index is measured are partitioned in different classes, then the Divisia index shares with the Laspeyres and the Paasche indices the fact that the Divisia index, over the partition, of individual classes Divisia indices is itself the Divisia index over all commodities.

If the weights $(p_{i,t}q_{i,t} / \sum_{i=1}^{N} p_{i,t}q_{i,t})$ can be assumed to be constant over the time period considered, then the Divisia index will also be equivalent to the share-weighted O-mean index:

$$\ln\left(\frac{Q_s}{Q_t}\right) = \sum_{i=1}^{N} \left(\frac{p_{i,t}\, q_{i,t}}{\sum_{i=1}^{N} p_{i,t}q_{i,t}}\right) \ln\left(\frac{q_{i,s}}{q_{i,t}}\right)$$

In practice, the shares will not be constant and a common approach has been to use Tornqvist-type weights, i.e. the arithmetic mean of the weights in period t and those measured in period s. This has the obvious advantage to produce an index which meets both the time and the factor reversal test, i.e. an ideal index:

$$\ln\left(\frac{Q_s}{Q_t}\right) \simeq 1/2 \sum_{i=1}^{N} \left[\left(\frac{p_{i,t}q_{i,t}}{\sum_{i=1}^{N} p_{i,t}q_{i,t}}\right) + \left(\frac{p_{i,s}q_{i,s}}{\sum_{i=1}^{N} p_{i,s}q_{i,s}}\right)\right] \ln\left(\frac{q_{i,s}}{q_{i,t}}\right)$$

Unfortunately when the shares change from period t to period s, it will not be true anymore that this approximation of the Divisia index of approximated Divisia indices will equal the approximation of the Divisia index of the components.

Nevertheless, Star and Hall (1976), using this approximation to the Divisia index, have shown that, "a reasonable approximation ... can be calculated using data from only the beginning and end of a long period of time."

V.8 The Economic Analysis of Index Numbers: a Diagrammatic Approach

In the preceding sections, index numbers have been considered from mainly an axiomatic or/and a statistical point of view. Given elementary indices as the relative values taken by a variable in two distinct situations, indices were introduced as averages of elementary indices. The index problems were on the one hand that of selecting the desired average and the proper weighting schemes and on the other hand that of the development of a set of properties an ideal index would be expected to meet.

In Economics, the consumers will maximize utility and producers will minimize cost or maximize profits, and through such optimization determine the combination of commodities demanded or produced in terms of prices. Most of the index literature has centered on the consumer, the problem being to determine the cost-of-living index, however, in terms of the producer, one can similarly determine a revenue deflator. Both of these alternatives have been considered by Fisher and Shell (1972).

In determining cost-of-living indices, the basic building block is the "basket of goods". Barring satiation and provided the preference map meets certain regularity conditions (see, for instance, Diewert (1979), in terms of the consumer two baskets will be seen as indistinguishable if they yield the same utility, i.e. if E_α and E_β are two vectors of quantities of commodities and if U( ) is the utility function,

$$U(E_\alpha) = U(E_\beta)$$

implies that E_α and E_β represent the same aggregate quantity of goods and services.. On the other hand, if

$$U(E_\beta) > U(E_\alpha)$$

i.e. if E_β is preferred to E_α, even though it is possible that some of goods and services are available in a smaller quantity, it will be said that E_β correspond to a quantity greater on the whole than E_α. In other words, the indifference curve is monotonically related to the quantity index.

In terms of the producer, his reference will be the production possibility frontier (PPF) determined, given inputs, by an aggregator F of outputs.

Facing prices $p_{1,\alpha}$ and $p_{2,\alpha}$, given a utility level u, consumers will minimize expenditures, and demand quantities x_1 and x_2 of each commodity. Through this operation, one can determine the expenditure function

$$G(u,p_1,p_2) = \min_{x_1,x_2} \{y = p_1x_1 + p_2x_2; \ U(x_1,x_2) \geq u, \ x_1, x_2 \geq 0\}$$

Given a price change from $(p_{1,\alpha}, p_{2,\alpha})$ to $(p_{1,\beta}, p_{2,\beta})$, a price index can be defined as the ratio of what it costs, given situation β to maintain the level of utility u in terms of what it costs under situation α. Such an index, which Diewert (1979) calls a Könus (1924) price index, can be denoted by $P_k(\underset{\sim}{x}, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta)$

$$P_k(\underset{\sim}{x},\underset{\sim}{p}_\alpha,\underset{\sim}{p}_\beta) = \frac{G(u,\underset{\sim}{p}_\beta)}{G(u,\underset{\sim}{p}_\alpha)}$$

It can be immediately observed that, as a result of the definition, given two vectors $\underset{\sim}{x}_\gamma$ and $\underset{\sim}{x}_\lambda$ such that

$$U(\underset{\sim}{x}_\gamma) = U(\underset{\sim}{x}_\lambda)$$

then

$$P_k(\underset{\sim}{x}_\gamma, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta) = P_k(\underset{\sim}{x}_\lambda, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta)$$

hence it is useful to first consider combinations of commodities yielding the same utility.

Had we considered the producer, it is a revenue function which we would have defined as

$$R(z, p_1, p_2) = \max_{x_1, x_2} \{y = p_1 x_1 + p_2 x_2 ;\ F(x_1), x_2) \leq z,\ x_1,\ x_2 \geq 0\}$$

where F is an output aggregator function determined by the technology.

A revenue deflator corresponding to the Könus price index can be defined as

$$D_k(\underset{\sim}{x}, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta) = \frac{R(z, \underset{\sim}{p}_\beta)}{R(z, \underset{\sim}{p}_\alpha)}$$

As with the cost-of-living index, if $\underset{\sim}{x}_\gamma$ and $\underset{\sim}{x}_\lambda$ are on the same PPF, i.e. if

$$F(\underset{\sim}{x}_\gamma) = F(\underset{\sim}{x}_\lambda)$$

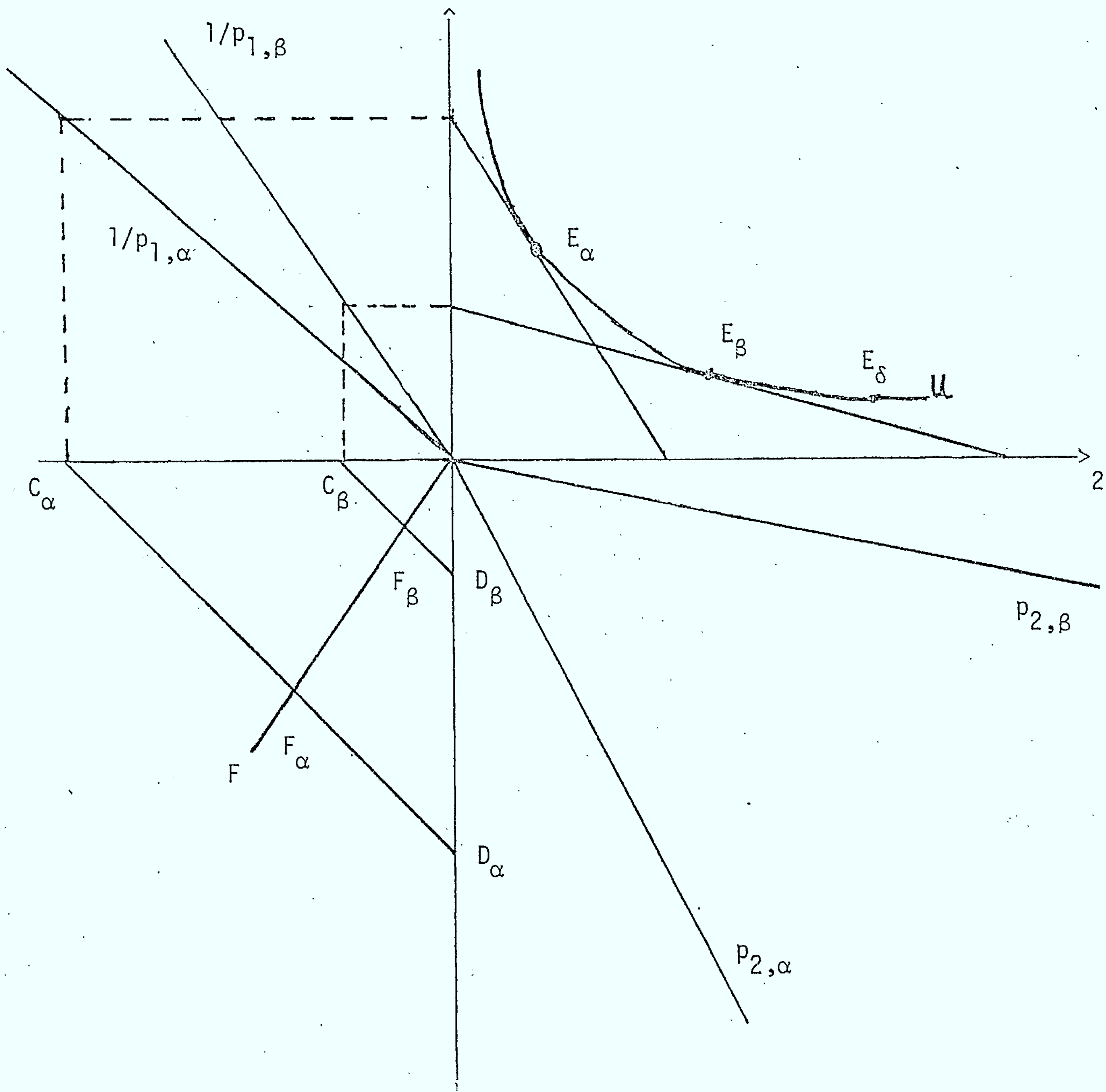
then

$$D_k(\underset{\sim}{x}_\gamma, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta) = D_k(\underset{\sim}{x}_\lambda, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta)$$

To illustrate these indices, we will consider the cost of living index, using the diagrammatic approach introduced earlier.

Given some vector E_δ of commodities 1 and 2 yielding under expenditure minimization u , the problem consists in determining the price index $P_k(E_\delta, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta)$. Given expenditure minimization, to $\underset{\sim}{p}_\alpha$ will correspond E_α and to $\underset{\sim}{p}_\beta$, E_β , such that

$$\underset{\sim}{U}(E_\alpha) = U(E_\beta) = U(E_\delta)$$



To E_α, given $p_{1,\alpha}$ and $p_{2,\alpha}$ - such that $(p_{2,\alpha}/p_{1,\alpha})$ is the slope of the utility function at E_α -, corresponds the budget line $\overline{C_\alpha D_\alpha}$. Similarly, given fall in prices to $p_{1,\beta}$, and the new expenditure minimization in E_β, given u, a new budget line $\overline{C_\beta D_\beta}$ is derived. The price index P_k is given by the ratio (OF_β/OF_α) given any line OF in the third quadrant, i.e. by

$$P_k(E_\delta,\underset{\sim}{p}_\alpha,\underset{\sim}{p}_\beta) = \frac{OF_\beta}{OF_\alpha} = \frac{OC_\beta}{OC_\alpha} = \frac{OD_\beta}{OD_\alpha}$$

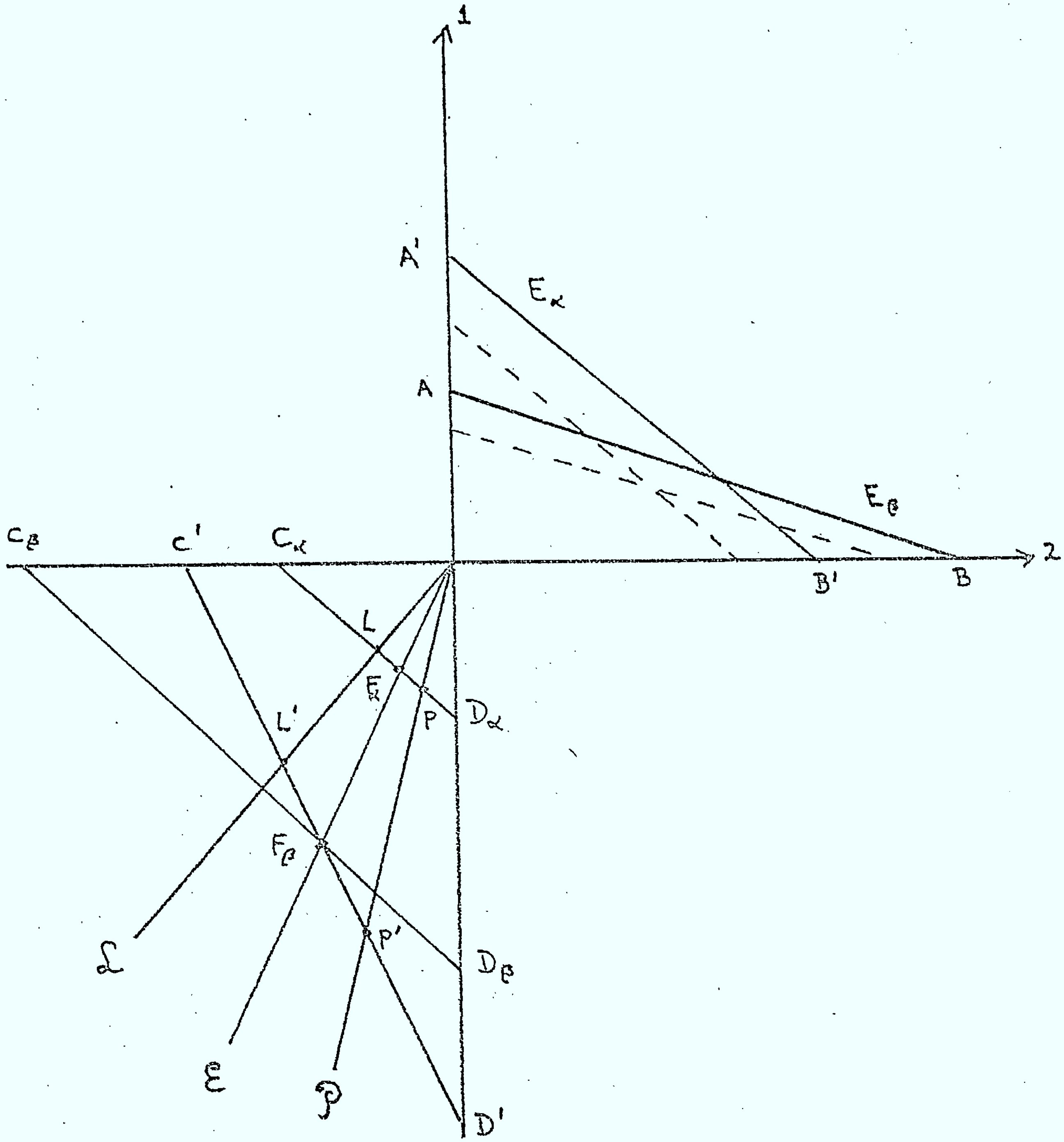
Given α as the base situation, Diewert (1979) defines the Laspeyres-Könus price index as $P_k(E_\alpha,\underset{\sim}{p}_\alpha,\underset{\sim}{p}_\beta)$ and the Paasche-Könus price index as $P_k(E_\beta,\underset{\sim}{p}_\alpha,\underset{\sim}{p}_\beta)$. Given $U(E_\alpha) = U(E_\beta)$, then

$$P_k(E_\alpha,\underset{\sim}{p}_\alpha,\underset{\sim}{p}_\beta) = P_k(E_\beta,\underset{\sim}{p}_\alpha,\underset{\sim}{p}_\beta)$$

Furthermore, we may compare this index to the Paasche and Laspeyres indices.

Let's assume that E_α and E_β are both on the same transformation curve, then from AB, the budget line $\overline{C_\alpha D_\alpha}$ is obtained, while from $A'B'$, the budget line $\overline{C_\beta D_\beta}$ is derived. The Könus price index will be determined along the line $\mathcal{E}$, as (OF_β/OF_α). The Laspeyres index, based on the original price set $(p_{1,\alpha},p_{2,\alpha})$ and output composition given by E_α, as seen earlier, will be given along the line $\mathcal{L}$ as (OL'/OL). Similarly, given the new allocation E_β, the Paasche index will be determined along the line $\mathcal{P}$ as (OP'/OP).

The following sets of inequalities follows:


1
A'
E_α
A
E_β
C_β
C'
C_α
2
B'
B
L
E_α
P
D_α
L'
F_β
P'
D_β
D'

(i) Given a cost-of-living index (i.e. an index in terms of the consumer maximizing utility), and given a proportionately greater increase in p_2 than in p_1

$$(p_{2,\beta}/p_{2,\alpha}) = (OD'/OD_\alpha) > \text{Laspeyres index} = (OL'/OL) >$$

$$> \text{Könus price index} = (OF_\beta/OF_\alpha) >$$

$$> \text{Paasche index} = (OP'/OP) >$$

$$> (p_{1,\beta}/p_{1,\alpha}) = (OC'/OC_\alpha)$$

(ii) Given a revenue deflator (i.e. a price index in terms of the producer maximizing profit) and given a proportionately greater increase in p_2 than in p_1

$$(p_{1,\alpha}/p_{1,\alpha}) = (OC'/OC_\alpha) < \text{Laspeyres index} = (OL'/OL) <$$

$$< \text{Deflator index} = (OF_\beta/OF_\alpha) <$$

$$< \text{Paasche index} = (OP'/OP) <$$

$$< (p_{2,\beta}/p_{2,\alpha}) = (OD'/OD_\alpha)$$

If relative prices do not change, E_α and E_β will be one and the same and the price effect will be a pure inflationary effect, with C'C' parallel to CD and the above inequalities becoming equalities.

If now E_α is on the ordinate, then L' and C' will be one and the same and

$$[(p_{1,\beta}/p_{1,\alpha}) = (OC'/OC_\alpha)] = [\text{Laspeyres index} = (OL'/OL)]$$

Similarly, if E_β is on the abscissa, then P' and D' will be the same and

$$[\text{Paasche index} = (OP'/OP)] = [(p_{2,\beta}/p_{2,\alpha}) = (OD'/OD_\alpha)]$$

Now this new index is equivalently the statistical index of the output combination E_i , as developed earlier, where E_i is the intersection of AB and A'B' . Since AB and A'B' are both negatively sloped, E_i and E_α will be the same if and only if E_α and E_β are the same since E_i is necessarily between E_α and E_β . To have $E_\alpha = E_\beta$ implies that the transformation curve has a kink at E_α . In fact if this result holds for all possible relative prices, i.e. for all (p_2/p_1) , then the transformation curve must be a right angle at E_α and the technology must be of the Leontief-type illustrating a result in Diewert (1979). In terms of a cost-of-living index, E_α would be to the right of E_β and the exact index would correspond to the index through E_0 rather than through E_i .

Until now the results have been based on E_δ , E_α and E_β being on the same indifference curve. Assuming that U is homothetic, i.e.

$$\frac{U_1(x_1,x_2)}{U_2(x_1,x_2)} = \frac{U_1(\lambda x_1,\lambda x_2)}{U_2(\lambda x_1,\lambda x_2)} \qquad \lambda \geq 0$$

where

$$U_i(x_1,x_2) = \frac{\partial}{\partial x_i}[U(x_1,x_2)]$$

if, given the price vectors $\underset{\sim}{p}_\alpha$ and $\underset{\sim}{p}_\beta$, E_α and E_β are the corresponding vectors of goods demanded, then for all $\lambda \geq 0$, λE_α and λE_β will be the vectors of goods demanded for a new utility level u_λ . In the third quadrant, passing from E_α and E_β to λE_α and λE_β, all the values will be multiplied by λ, yielding λOC_α, $\lambda OC'$, λOC_β, λOD_α, λOD_β, $\lambda OD'$, ...

The new indices will be

$$\mathcal{P}_k(\lambda \underset{\sim}{x}, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta) = \frac{\lambda OF_\beta}{\lambda OF_\alpha} = \frac{OF_\beta}{OF_\alpha} = P_k(\underset{\sim}{x}, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta)$$

$$\mathcal{L}_k(\lambda E_\alpha, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta) = \frac{\lambda OL'}{\lambda OL} = \frac{OL'}{OL} = (E_\alpha, \underset{\sim}{p}_\alpha, \underset{\sim}{p}_\beta)$$

...

Hence, given homotheticity, the earlier results generalize to points which are not on the same indifference curve (PPF), and the answers depend solely on the prices.

V.9 The Statistical Index and Economic Analysis

It is useful to evaluate the statistical approach to index numbers developed earlier in the context of the economic analysis. The basic statistical price index was a function of a vector of quantities and two price vectors; its aim was to measure, given that quantity vector, the average price change. Given the pure statistical approach, there exists no criterion to compare two output combinations, say E_α and E_β hence as soon as the price change leads to a move from E_α to E_β, the index number problem appears and to it there is no unambiguous solution, especially since, as we have seen the desirable properties one may want an index to fulfill are not all simultaneously consistent (Eichhorn and). The economic analysis does enable one to say something about two commodity vectors since those commodities are conceived as the result of either a production process describable by a production function or a utility generation process describable by a utility function. Such an analysis does put constraint on the commodity space, indicating that as long as two commodity vectors either require the same technology and the same aggregate of factors of production as input or produce the same satisfaction, they are in the context of the problem, indistinguishable.

Then, whereas the statistician can only ask how much would the given commodity vector cost under two alternative price regimes and obtain

The question of the relative values of the statistical index and of the economic index cannot be answered without further specifications since the statistical index is specified in terms of an arbitrary vector of goods and services. Three possible specifications have already been considered. Two specifications are determined by whether the vector of goods and services denoted by E corresponds to the economic equilibrium (i) prior to the price change yielding the Laspeyres price index, (ii) following the price change yielding the Paasche price index, the third corresponds to the value the statistical index's vector of goods and services must take to equal the economic index.

Let's return to the first specification, and consider now simultaneously the Könus cost of living index P_k and the deflator D_k. The point E_α corresponds to the vector of goods and services produced and demanded in the base situation α, as determined by the price vector $\underset{\sim}{p}_\alpha$. Through E_α

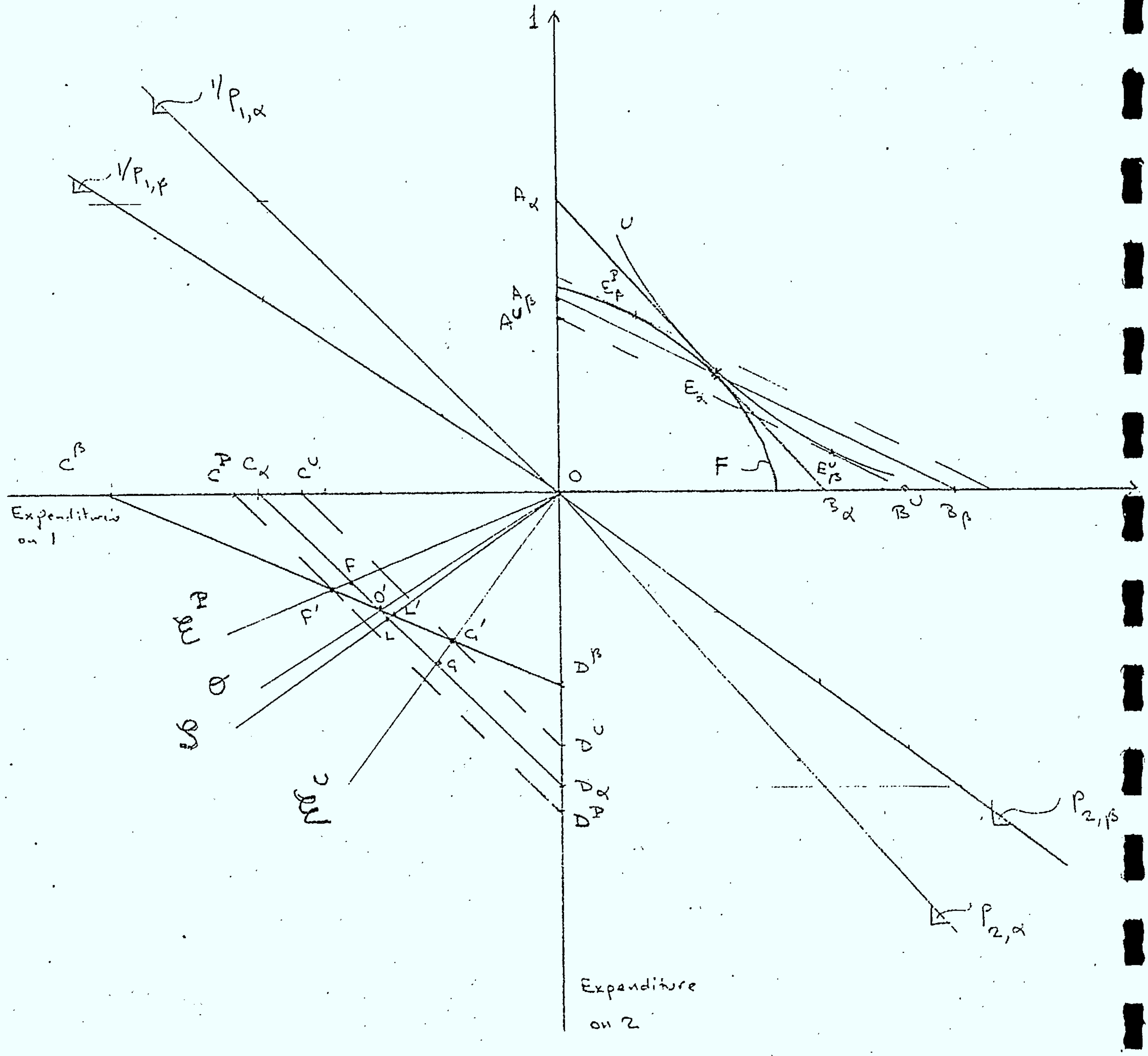
passes a utility function $U(E_\alpha)$, a PPF $F(E_\alpha)$ and the situation α budget constraint $\overline{A_\alpha B_\alpha}$. Since E_α was an equilibrium point, at E_α the utility function, the PPF and the budget line are all tangent to each other. Given a new price vector $\underset{\sim}{p}_\beta$, the statistical index, which will correspond to the Laspeyres index, will be determined by a budget line $\overline{A_\beta B_\beta}$ which will also pass by E_α . In the attached diagram, since p_1 increases while p_2 falls, indices are bounded from above by (OC_β/OC_α) and from below by (OD_β/OD_α) , where

$$(OC_\beta/OC_\alpha) > 1 > (OD_\beta/OD_\alpha)$$

Given E_α such that $x_{1,\alpha} = 3$ and $x_{2,\alpha} = 4$, the statistical index is determined along the line $0\mathcal{S}$ as (OL'/OL) . It is unsatisfactory from an economic point of view because it neglects a simple economic fact; the consumer facing the new price vector $\underset{\sim}{p}$ will find it advantageous to ask less of commodity 1 and to make up for the loss by demanding more of commodity 2 - unless, as we have seen earlier his utility function, being of the Leontief type, exludes substitution. In practice, the consumer will move from E_α to E^u_β such that $x^u_{1,\beta} < x_{1,\alpha}$ and $x^u_{2,\beta} > x_{2,\alpha}$, where $E_\beta = (x_{1,\beta}, x_{2,\beta})$. The budget line is now $\overline{A^u B^u}$, the constant quantity being represented by the same utility level as in E_α , i.e.,

$$U(E^u_\beta) = U(E_\alpha)$$

Then the Konüs cost of living index is read on $\overline{0\mathcal{E}^u}$ by (OG'/OG) . Denoting the statistical index by $\mathcal{S}(E_\alpha)$ and the Konüs cost of living index by $\mathcal{E}^u(E_\alpha)$, it follows that the curvature of the utility function yields, as we have seen



1
$1/p_{1,\alpha}$
A_α
U
F
O
C^β
C_α
C^U
Expenditure on 1
B_α
B^U
B_β
D^β
D^U
D_α
$p_{2,\beta}$
$p_{2,\alpha}$
Expenditure on 2

$$\mathcal{E}^u(E_\alpha) \leq \mathcal{S}(E_\alpha)$$

the equality holding only if the utility function is a Leontief function.

Simultaneously, reading the revenue deflator along the line $\overline{0\mathcal{E}^P}$ as (OF'/OF) , the curvature of the PPF gives us as we saw earlier

$$\mathcal{E}^P(E_\alpha) \geq \mathcal{S}(E_\alpha)$$

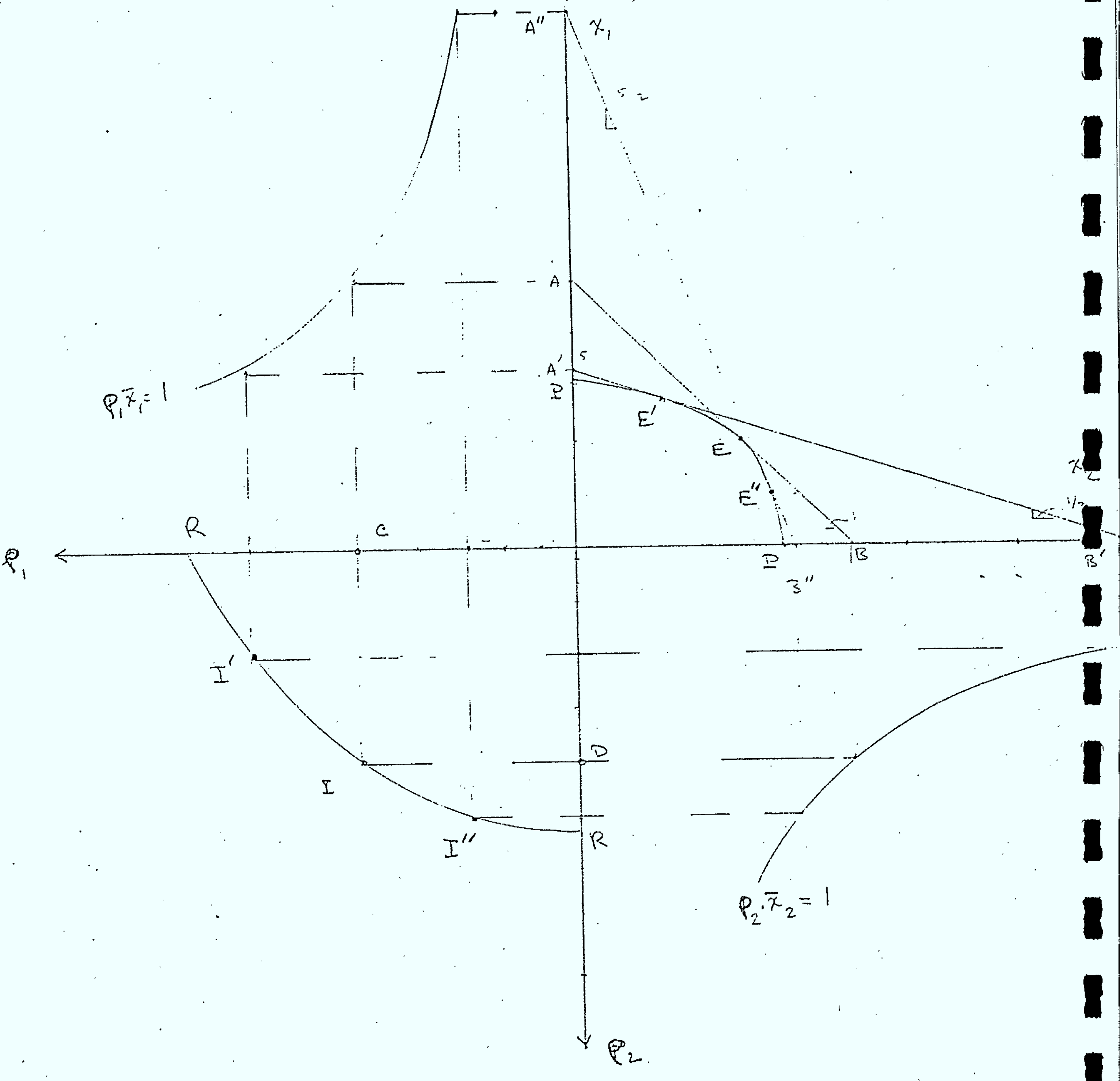
It follows that statistical indices, Laspeyres and Paasche indices all fail to account for the substitution effect due to a price change, the cost of living being systematically overestimated and the revenue deflator being systematically underestimated by the Laspeyres index, the opposite holding in terms of the Paasche index. In practice, however without international trade, the consumer cannot move to E^u_β while the producer moves to E^P_β , rather, through market forces, both will move to E_β which could be E^P_β or E^u_β or any other vector in the commodity space.

V.10 Cost Functions and Price Indices: a Diagrammatic Analysis

Another approach to the construction of cost of living index and price deflator can be presented using the duality between prices and quantities.

First we begin by following Darrough and Southey's (1977) method to map the indirect PPF function for a given value of the PPF, $F(x)= F_{\tilde{o}}$, denoted in the diagram by PE'EE"P . The indirect PPF function may be defined as $G(\tilde{\rho})$ where

$$G(\tilde{\rho}) = \max_{x} \{F(\tilde{x}) : \tilde{\rho}^T\tilde{x} \leq 1 , x \geq 0\}$$


A''
x_1
A
A'
E'
E
E''
x_2
R
C
ρ_1
B
B'
B''
I'
D
I
I''
R
$\rho_1 \cdot \bar{x}_1 = 1$
$\rho_2 \cdot \bar{x}_2 = 1$
ρ_2

$$\underset{\sim}{\rho} = (\frac{1}{y})\ \underset{\sim}{p}$$

Let's consider E . If the budget is expressed in terms of commodity 1, it would correspond to $\bar{x}_1$, in A , which through the normalized budget constraint, $\rho_1 \bar{x}_1 = 1$, in the form of a unit hyperbole, yields the price ρ_1 . ρ_2 is obtained the same way. Repeating the operation for E' , E" ,..., one determines I' , I" and the whole indirect PPF $G(\underset{\sim}{\rho})$.

Now, we may also define the cost function as

$$C(p) = \min_{x} \{\underset{\sim}{p}^T \cdot \underset{\sim}{x} ; F(\underset{\sim}{x}) \geq F ; x \geq 0\}$$

and note that, given $\underset{\sim}{p} = \underset{\sim}{\rho}$ and $F = F_o$,

$$G(\underset{\sim}{\rho}) = F_o \cdot C(\underset{\sim}{\rho})$$

In other words, we have obtained a mapping of the cost function but for a constant. It is now simple to map the other cost iso curves by replacing the unit hyperbole by hyperbole of the form $p_i \bar{x}_i = (\underset{\sim}{p}^T \cdot \underset{\sim}{x}) = y$.

It should be noted that a move along the cost curve say from I to I' represents neither a price change nor a quantity change since, as far as the optimizing producer is concerned, I and I' are indistinguishable. They are not quantity changes since, by hypothesis F(E) = F(E') = F_o . The relative price increase of commodity 1 has been exactly compensated through an increase in the quantity produced of that commodity at the expense of commodity 2 .

The Könus Price Deflator is now easily obtained since, given two price vectors $\underset{\sim}{p}_\alpha$ and $\underset{\sim}{p}_\beta$, and a PPF F_o , two cost functions are derived and these functions are positive linear homogeneous (PLH).

V.11 Quantity Indices

Until now we have constructed price deflators (or alternatively cost of living indices) on the presumption that output was aggregated through an aggregator function F and that this aggregator function took the value F_0 . In a more general situation, there is no reason why E_β , the new equilibrium in the output space, should be on the same PPF as E_α , the original equilibrium.

If F is homothetic however, i.e. if given some positive number λ there exists some positive number γ such that

$$F(\lambda \underset{\sim}{x}) = \gamma F(\underset{\sim}{x})$$

then, denoting by $F_i(\underset{\sim}{x})$ the partial derivative of F with respect to x_i ,

$$\frac{F_i(\underset{\sim}{x})}{F_j(\underset{\sim}{x})} = \frac{F_i(\lambda \underset{\sim}{x})}{F_j(\lambda \underset{\sim}{x})} \qquad i,j = 1,2,\ldots N$$

this implies that, given a price vector $\underset{\sim}{p}$, the output composition is independent of scale.

Now, given the equilibrium price vector $\underset{\sim}{p}_\alpha$ and the corresponding commodity equilibrium vectors E_α and E_β , then a ray from the origin passing through E_β will intersect E_α's PPF at E_α . Denoting the distance from the origin to $F(E_\alpha)$ as a ratio to that distance from the origin to E_α by λ, F_α and F_β the values taken by the aggregator functions passing through E_α and E_β respectively, then there exists some positive number γ such that

$$F_\alpha = \gamma F_\beta$$

i.e.

$$F(\lambda E_\beta) = \gamma F(E_\beta)$$

Homotheticity implies that, in terms of E_α, the constant λ^{-1} applied to the output vector will take us on the PPF $F(E_\beta)$, i.e.

$$F(\lambda^{-1} E_\alpha) = \gamma^{-1} F(E_\alpha)$$

where

$$F(\lambda^{-1} E_\alpha) = F(E_\beta)$$

and we conclude that

$$F(E_\beta) = F(\lambda E_\alpha) = \gamma F(E_\alpha)$$

The quantity index is defined as γ, i.e. that scalar by which we need to expand E_α so as to reach E_β's PPF. More formally, given $F(E)$ the output aggregator function,

$$Q = \frac{F(E_\beta)}{F(E_\alpha)}$$

However, once again, one can use duality theory to construct the quantity index. Given a price vector $\underset{\sim}{p}_\alpha$, we can determine a cost curve in the price space, C_0, such that any point in that curve is indistinguishable from any other point. To each situation there corresponds a budget, hence we move from budget y_α to budget y_β, i.e. in quadrant II and IV from one hyperbole to another through which we determine the PPF's F_α and F_β. The index thus constructed is both the Allen quantity index and the Könus implicit quantity index since C and F are assumed to be homothetic.

Price and quantity indices may now be represented simultaneously Given a change from price vector $\underset{\sim}{p}_\beta$ and given the corresponding move from E_α to E_β in the commodity space, the move from the cost curve associated with $\underset{\sim}{p}_\alpha$ to that associated with $\underset{\sim}{p}_\beta$ is the price index and the move from F_α to F_β, where F_α and F_β denote the PPF, is the quantity index. In fact homotheticity insures that

$$\frac{C(\underset{\sim}{p}_\beta, F_\beta)}{C(\underset{\sim}{p}_\alpha, F_\alpha)} = \frac{c(\underset{\sim}{p}_\beta)}{c(\underset{\sim}{p}_\alpha)} \cdot \frac{F(E_\beta)}{F(E_\alpha)}$$

$$= P \cdot Q$$

where c is the unit cost function

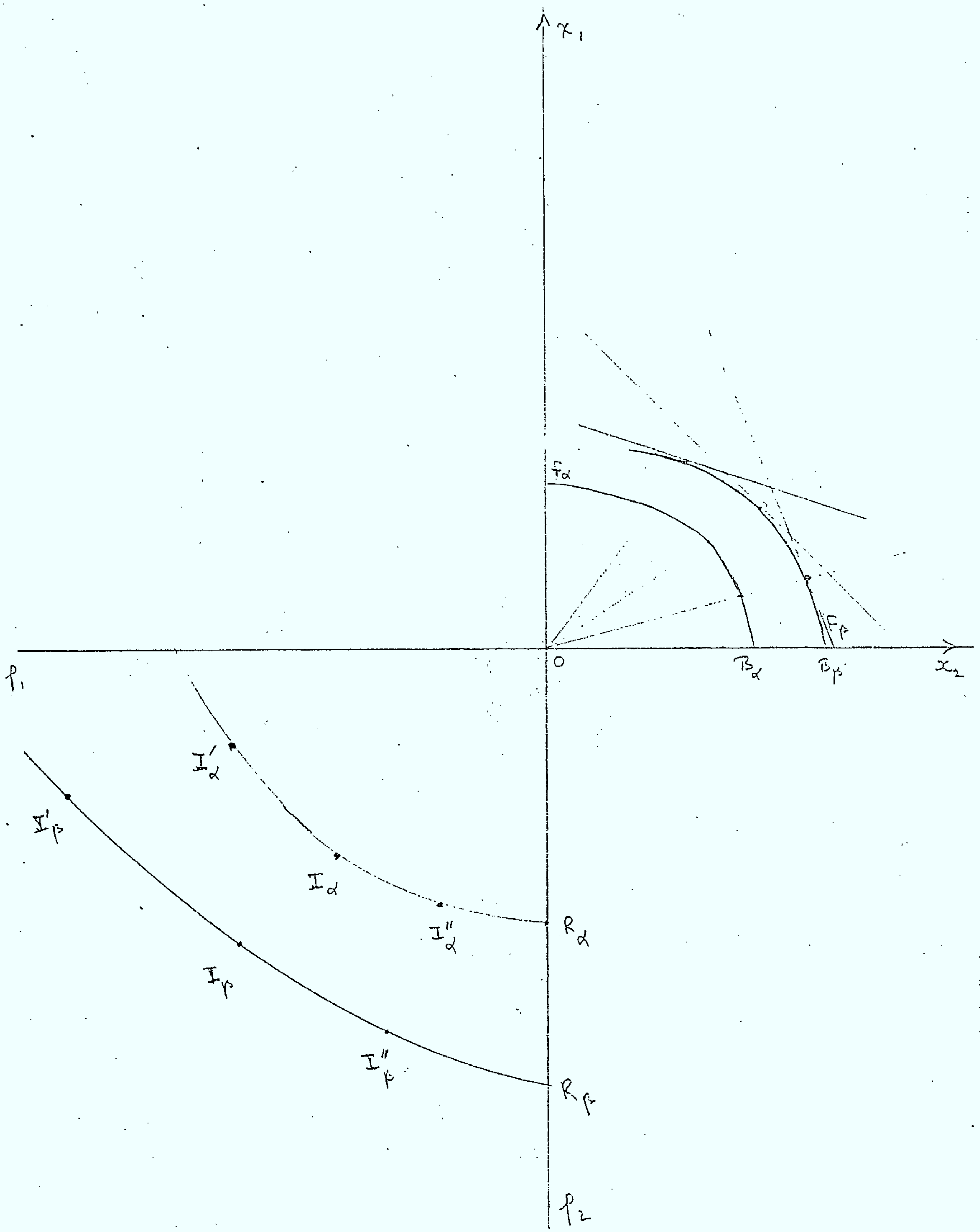
P and Q are respectively the price and quantity index

In the attached diagram, the quantity index can be read on any ray from the origin in the first quadrant, say

$$Q = \frac{OB_\beta}{OB_\alpha}$$

while the price index can be read on any ray from the origin in the third quadrant, say

$$P = \frac{OR_\beta}{OR_\alpha}$$



x_1
F_α
F_β
O
B_α
B_β
x_2
p_1
I'_α
I'_β
I_α
I''_α
R_α
I_β
I''_β
R_β
p_2

V.12 Non-Homothetic Functions

Assuming that F is not homothetic, then the distance from $F(E_\alpha)$ to $F(E_\beta)$, in relation to that from the origin to $F(E_\alpha)$, is not unique and independent of the output combination. In terms of the cost function, the unit cost function associated with E_α will not be homothetic with respect to that associated with E_β.

In this context there does not exist a unique strategy to construct a price index in the sense that all indices are sensitive to the hypotheses.

i) Könus price index and implicit quantity index

It is possible to select an equilibrium in the commodity space. This can be E_α, E_β or some combination of them. Denoting the selected point in the commodity space by E_0, with it a PPF F_0 can be associated. With F_0, a cost function can be mapped in the price space. Then the price index is defined as

$$D_k = \frac{C(p_\alpha, E_0)}{C(p_\beta, E_0)}$$

Given the two budgets y_α and y_β, where $y = \sum_{i=1}^{N} p_i x_i$, a quantity index may be defined as

$$\tilde{Q}_k = \left(\frac{y_\beta}{y_\alpha}\right) \Big/ D_k$$

i.e. an implicit Könus quantity index.

ii) Allen's quantity and implicit price index

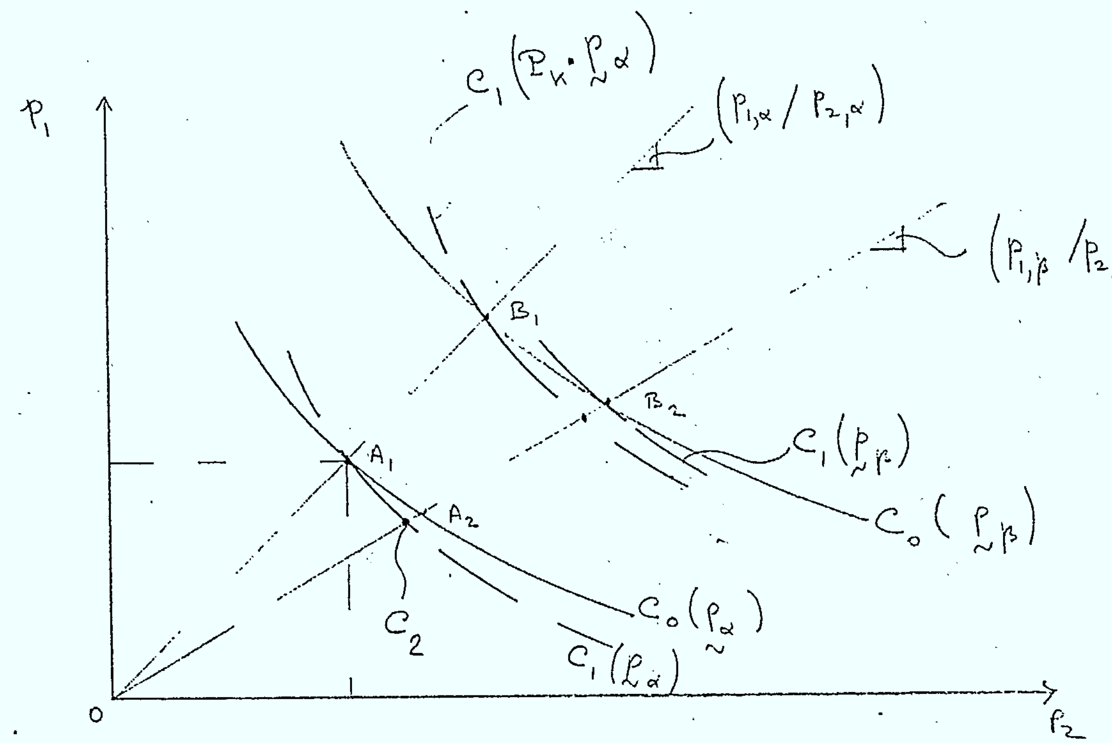
Using duality, some point in the price space could be selected, say $\underset{\sim}{p}_0$, to which will correspond the cost curve C_0. Then using the procedure introduced earlier, given y_β and y_α, a pseudo-production function $(\tilde{F})$ can be mapped in the commodity space which is positive linear homogeneous, and a quantity index is defined as the ratio of the distance to $\tilde{F}_\beta$ to that to $\tilde{F}_\alpha$. This index will always correspond to the Allen quantity index which is defined as

$$Q_A = \frac{C[F(E_\beta), \underset{\sim}{p}_o]}{C[F(E_\alpha), \underset{\sim}{p}_o]}$$

A price index $\tilde{p}_A$ can be defined implicitly so that

$$\tilde{P}_A \cdot Q_A = (y_\beta / y_\alpha)$$

Könus price index and Allen's quantity index, from that approach are comparable in that where, to develop the Könus index a reference PPF F_o is taken to determine an homothetic (and linear homogeneous) mapping in the price space, denoted in the diagram by C_o.



The reference PPF is crucial since, had we selected F_1 different from F_o , given the lack of homotheticity, another map C_1 would have been derived.

In the chart,

$$P_k(F_o) = \frac{OB_1}{OA_1} = \frac{OB_2}{OA_2} = \frac{C[\underset{\sim}{p}_\beta, F_o]}{C[\underset{\sim}{p}_\alpha, F_o]}$$

$$P_k(F_1) = \frac{OB_2}{OC_2} = \frac{C[\underset{\sim}{p}_\beta, F_1]}{C[\underset{\sim}{p}_\alpha, F_1]} \neq \frac{OB_1}{OA_1} = \frac{C[P_k(F_o)\underset{\sim}{p}_\alpha, F_o]}{C[\underset{\sim}{p}_\alpha, F_1]}$$

and

$$P_k = (F_o) \neq P_k(F_1)$$

Allen's quantity index can be represented equivable in the commodity space in terms of the PLH pseudo-production mapping, given a cost function.

iii) Malmquist indices

In the approaches earlier one starts alternatively from the production or the cost function, specified at some arbitrary level to develop through duality a correspondant PLH mapping in the dual space. The alternative associated with Malmquist is to start in either space, select, as previously, a reference curve (either production or cost) and then define a distance function between that curve and those which correspond to the base and current situations. The quantity index is analysed in Diewert (1979) and the price index is analysed, independently, by Blackorby, Primont and Russell (1978). The distance function is the transformation function:

$$D[F(E), E_0] = \max_{\lambda} \{\lambda : F(E_0/\lambda) \geq F(E), \lambda > 0\}$$

or the transformation function derived from the indirect production function.

$$D[F(E), \underset{\sim}{p}_0] = \min_{\lambda} \{\lambda : V(\underset{\sim}{p}/\lambda) \geq V(\underset{\sim}{p}_0), \lambda > 0\}$$

Then, depending upon which distance function is used, either a quantity index Q_M or a price index p_M will be derived. The other, $\tilde{p}_M$ or $\tilde{Q}_M$, will be derived implicitly.

Neither the Könus-Allen approach nor the Malinquist approach resolve the problem unambiguously in the presence of non-homotheticity. The problem of homotheticity may not be all that serious - as long as one utilizes chain index - since then the homotheticity condition will generally be approximately met between two consecutive periods even if it does not hold even approximately over two periods further apart. For instance, even though Breslau, Corbo and Smith (1978) have shown that homotheticity did not hold for Bell Canada, as long as one can find a series of homothetic production (cost) functions G_t which approximates it well over $(t-1, t)$, the requirement needed for the index construction will be fulfilled.

It is not clear that to any index, and in particular those thus far considered there should correspond production functions, and a subset of indices is defined to include all those indices to which corresponds production functions (or cost functions); indices belonging to the subset will be said to be exact. If, in addition, these production (or cost) functions are flexible; i.e., if they can be considered to be second order approximations then these indices have been defined by Diewert to be superlative. Following Christensen, let us consider the implicit production function

$$f(y_1, \ldots , y_m; x_1, \ldots x_n; t) = 0$$

where the y_j denote the outputs, the x_i the inputs and t denotes time. It has been shown by McFadden that if this production function has a strictly convex input structure, then there exists a unique cost function which is its dual:

$$C = g(y_1, \ldots y_m ; w_1, \ldots , w_n ; t)$$

where the w_i are the prices at which the x_i are purchased. C is the total cost, that is

$$C = \sum_{i=1}^{N} w_i x_i$$

The cost function is linear homogeneous and, by Sheppard's lemma,

$$\frac{\partial \ln g}{\partial \ln w_i} = \frac{w_i x_i}{\sum_{i=1}^{N} w_i x_i}$$

Total differentiation with respect to time yields both

$$\frac{\partial \ln C}{\partial t} = \sum_{j=1}^{M} \frac{\partial \ln g}{\partial \ln y_j} \frac{\partial \ln y_j}{\partial t} + \sum_{i=1}^{N} \frac{\partial \ln g}{\partial \ln w_i} \cdot \frac{\partial \ln w_i}{\partial t} + \frac{\partial \ln g}{\partial t}$$

and

$$\frac{\partial \ln C}{\partial t} = \sum_{i=1}^{N} \left(\frac{w_i x_i}{\sum_{i=1}^{N} w_i x_i} \right) \left(\frac{\partial \ln \dot{w}_i}{\partial t} + \frac{\partial \ln \dot{x}_i}{\partial t} \right)$$

As $(\partial \ln g / \partial \ln y_j)$ is the cost elasticity of the jth output, given constant return to scale in production and marginal cost pricing, then

$$\frac{\partial \ln g}{\partial \ln y_j} = \frac{p_j y_j}{\sum_{j=1}^{M} p_j y_j}$$

i.e.,

$$-\frac{\partial \ln g}{\partial t} = \sum_{j=1}^{M} \left(\frac{p_{j,t} y_{j,t}}{\sum_{j=1}^{M} p_{j,t} y_{j,t}} \right) \left(\frac{\dot{y}_{j,t}}{y_{j,t}} \right) - \sum_{i=1}^{N} \left(\frac{w_{i,t} x_{i,t}}{\sum_{i=1}^{N} w_{i,t} x_{i,t}} \right) \left(\frac{\dot{x}_{i,t}}{x_{i,t}} \right)$$

The terms on the right-hand side are Divisia indices of outputs and inputs used in Jorgenson and Griliches' total factor productivity measure. In as much as the underlying assumptions do not hold, say if the output prices do not reflect marginal costs and/or if there is increasing or decreasing return to scale, then one can use cost elasticities with respect to outputs.

As in the preceding section, the major stumbling block is the fact that one does not have continuous observations, and the solution is to use an approximation to the Divisia index. The most commonly used approximation is to average the shares of the two consecutive periods. This has been shown by Lau to imply as an underlying production function a translogarithmic function; i.e., that the index of, say, inputs, x_t :

$$\ln \left(\frac{x_t}{x_{t-1}} \right) \simeq \frac{1}{2} \sum_{i=1}^{N} \left(\frac{w_{i,t} x_{i,t}}{\sum_{i=1}^{N} w_{i,t} x_{i,t}} + \frac{w_{i,t-1} x_{i,t-1}}{\sum_{i=1}^{N} w_{i,t-1} x_{i,t-1}} \right) \ln \left(\frac{x_{i,t}}{x_{i,t-1}} \right)$$

is exact and superlative with respect to the aggregator function

$$\ln x_t = \alpha_o + \sum_{i=1}^{N} \alpha_i \ln x_{i,t} + \frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} \beta_{i,j} \ln x_{i,t} \ln x_{j,t}$$

Had the Fisher ideal index been used as an approximation, to the Divisia index, then it would have been exact and superlative for that function which is the square root of a homogenous quadratic function.

It should be noted that Diewert's comment to the extent that the fact that the Tornquist approximation and, in general, superlative approximation to the Divisia index are not factor reversible can be associated with the observation that while one index follows from the cost function, the other follows from the production function and with the result that a second order flexible cost function will not generally be the dual of a second order flexible production function cannot be substantiated since, while Fisher ideal indices are superlative indices hence while the quantity and price index forms correspond to production and cost functions which are not dual of each other, such indices do meet the factor reversal test. This may follow, as Diewert noted subsequently, from the fact that the relationships which define exactness and superlativeness are not due to one; i.e., that, while more than one index may be exact in terms of a functional form, conversely an index can be exact in terms of more than one functional form.

Diewert has developed another result which appears crucial to the index problem in the practical context of the total factor productivity study. It was indicated, in the context of Laspeyres and Paasche indices that, in view of their linearity, an index of indices was the index of the components used to build those indices. This property is not shared by Fisher ideal index, nor, for the matter, with any index constructed as a single geometric average of indices such as, say, the quadratic mean of order r indices; this problem is well known, for instance, by those who use the Tornquist approximation to the Divisia index. Diewert's result shows that in fact, locally, in the neighbourhood of the base period t, any superlative index is a second order approximation to a Vartia index which meets the desired property.

The last problem remaining is not trivial, however it may be only of second order importance: given a set of possible superlative indices, which one should be chosen? Its importance is of second order since any superlative index is exact with respect to a flexible form which is itself a second order approximation to either any production function or any cost function which meet certain general criterion; i.e., two distinct superlative index will be exact with respect to two distinct flexible functions which will themselves be second order approximations of the same underlying form. Nevertheless, with some procedure to select among alternative second order approximations, one would have to follow an iterative procedure between building indices for aggregated inputs and outputs and evaluating the various flexible forms since the aggregated inputs and outputs would have to be exact with respect to the flexible functions to be evaluated.

APPENDIX. MEASURES OF CENTRAL TENDENCY

In spite of the more common formulation of the Laspeyres and Paasche indices as a ratio of two budgets, to study the general class of indices, it is best to consider them as weighted arithmetic or harmonic means of elementary indices.

The arithmetic mean is well known; it is the weighted sum of the component variables. It is written as

$$\overline{X}_A = \sum_{i=1}^{N} w_i x_i$$

where the weights are w_i .

It was shown that the Paasche index was the inverse of the arithmetic mean of the inverse of the elementary indices; i.e., a harmonic mean:

$$\overline{X}_H = \left(\sum_{i=1}^{N} w_i x_i^{-1} \right)^{-1}$$

In fact those two forms support a generalization, the r-mean

Definition. The r-mean is the rth root of the arithmetic mean of the variables x_i at the rth power, $r \neq o$

$$\overline{X}_r = \left(\sum_{i=1}^{N} w_i x_i^{r} \right)^{1/r}$$

When $r = 1$, this is the arithmetic mean while, when $r = -1$, it is the harmonic mean.

The geometric mean is in fact the o-mean ; i.e., the limit of r-mean as r tends toward zero:

$$\left(\overline{x}_r\right)^r = \left(\sum_{i=1}^{N} w_i x_i^{\,r}\right)$$

and, since $\sum^{N} w_i = 1$,

$$\left\{\frac{(\overline{x}_r)^r - 1}{r}\right\} = \sum_{i=1}^{N} w_i \left(\frac{x_i^{\,r} - 1}{r}\right)$$

and, as r approaches zero

$$\left(\frac{x_i^{\,r} - 1}{r}\right) \rightarrow \ln x_i$$

i.e. $(\overline{x}_r)$ approaches $(\overline{x}_o)$ where

$$\ln \overline{x}_o = \sum_{i=1}^{N} w_i \ln x_i$$

$$\overline{x}_o = \prod_{i=1}^{N} x_i^{\,w_i}$$

It can be shown that given $s > r$, for all s and r , $(\overline{x}_s) > (\overline{x}_r)$

hence that the harmonic mean is smaller than the geometric mean, itself smaller than the arithmetic mean, ... In fact $(\overline{x}_r)$ will vary in a continuous fashion between $(\min_i x_i)$ and $(\max_i x_i)$ as r varies between -s and +s . In other words any value in the interval between $(\min_i x_i)$ and $(\max_i x_i)$ is itself r-mean for some r .

In fact the concept of an average can be further generalized by specifying that, given x_i and w_i , $i = 1,2, \ldots , N$, and given any vector space in which vector addition is denoted by α and scalar multiplication by β , the Q-mean of x_i , $\overline{x}_Q$, is given by

$$\overline{x}_Q = (w_1 \ \beta \ X_1) \ \alpha \ (w_2 \ \beta \ x_2) \ \alpha \ \ldots \ \alpha \ (w_N \ \beta \ x_N)$$

Another way to represent the Q-mean of x_i , since by definition it is isomorphic to the vector space with single addition and vector addition, is

$$\overline{x}_Q = Q^{-1}\left\{\sum_{i=1}^{N} w_i Q(x_i)\right\}$$

If one considers all subsets of x_i obtained in successive selections with replacement of the x_i variables among the N variables, and any mean taken over each of these subsets, one can further take any mean of those means. If the means of the subsets and the mean of the means are defined in terms of the same vector space, then the mean of the means will be none other than that mean of the x_i ; however, this need not be the case. The general formulation will be

$$\overline{x}_{Q,\psi} = Q^{-1}\left\{\sum_{\ell} w_\ell \ Q\left[\psi^{-1}\left(\sum_{k}^{N} v_k \ \psi(x_k)\right)\right]\right\}$$

A particular form of the (Q,ψ) mean is the quadratic mean of order r , where $n=z$, $\psi(x_k) = \ln x_k$, $v_k = \frac{1}{n}$, and finally the Q-mean is any r-mean:

$$\overline{x}_{r,o} = \left\{\sum_{\ell=1}^{N} \sum_{k=1}^{N} w_{\ell,k} \ x_\ell^{r/2} \ x_k^{r/2}\right\}^{1/r}$$

A last generalization of averages is also utilized -- it is the Q-mean of a set of different ψ_k-means, $k=1,2, \ldots , H$.

To conclude this section, it must be noted that one can conceive of as many means as one wants, hence that the choice must be made in terms of some external criterion. This is in fact the approach Fisher had followed in building price indices.

Footnotes

1. The following listed items may not be self-explanatory.

    (7) "Local PL" are private lines within an EAS area.
    (8) "PL Radio" are non-operator handled radio calls.
    (12) "Net Toll PL" are monthly rental of toll-free lines.
    (14) "Semi-Public Coin" are coin telephones on private premises.

2. As in footnote 1, the following notes may help explain certain items.

    (II.4) "EFRC" is extended flat rate calling which is put in place only after positive response referenda by the affected customer.
    (III.4) "Special Assemblies" include items such as custom built terminals.
    (VII.3) "Custom Work" includes items such as the moving of poles.

3. A centre established by private industry for the study of productivity.

4. See Puiseux and Bernard for a good description of the underlying theory.

5. From "Rate Adjustment Formula, An Overview and Assessment", Department of Communications, Economic Policy and Statistics Branch, July 1975.

References

Artle, R. and C. Averous (1973), The Telephone System as a Public Good: Static and Dynamic Aspects", Bell Journal of Economics and Management Science, 4, 89-101.

Atkinson, S. and R. Halvorsen (1980), "A Test of Relative and Absolute Price Efficiency in Regulated Utilities", Review of Economics and Statistics, LXII, 1, 81-88.

Averch, H. and L. Johnson (1962), "Behavior of the Firm Under Regulatory Constraint", American Economic Review, 52, 1052-1069.

Baude, J., G. Cohen et. al. (eds.) Perspectives on Local Measured Service, Telecommunications Industry Workshop, Kansas City, Missouri.

Bernard, P., et. al., "La Productivite de l'Industrie Electrique", Union Internationale de Producteurs et Distributeurs d'Energie Electrique, Congres de la Haye, Aout, 1978.

Bernard, P. and L. Puiseux, "Essai de Mesure de la Productivite Globale de Facteurs a l'Ectricite de France", La Revie Francaise de L'Energie, no. 180, Mai, 1976.

Box, G. and D. Cox (1964), "An Analysis of Transformations", Journal of the Royal Statistical Society, Series B, 26, 211-243.

Cairne, M. and B. Kirln, "Canadian for Hire Trucking and the Effects of Regulation: A Cost Structure Analysis", CTC, Nov., 1979.

Caves, D.W. and L.R. Christensen, "Productivity In Canadian Railroads, 1956-1975", CTC, August, 1978.

D'Amous, A. and G. Godbent, "Measures de Productivite: Le Cas de Bell Canada", Universitie de Sherbrooke, Aout, 1974.

Denny, M. (1974), "Conceptual Problems in the Measurement of Service Outputs", Institute for Policy Analysis, Toronto.

Denny, M., M. Fuss and L. Waverman (1979), "Productivity, Employment and Technical Change in Canadian Telecommunications: The Case of Bell Canada, Report to Department of Communications, Ottawa.

Denny, M. and J.D. May, "On the Sensitivity of Productivity Measurement in Canadian Manufacturing to Some Conceptual Issues", Paper presented at the CEA meetings, June 1977, Fredericton.

Dhruvarajan, P.S. and R.F. Harris, "A Productivity Study of the Canadian Airline Industry", CTC, March 1978.

Diewert, W.E., (1976), "Exact and Superlative Index Numbers", Journal of Econometrics, 4, 115-145.

Diewert, W.E. (1978), "Superlative Index Numbers and Consistency in Aggregation", Econometrica, 46, 883-900.

Diewert, W.E. (1979), "The Economic Theory of Index Numbers: A Survey", Discussion Paper 79-09, University of British Columbia, Vancouver.

Diewert, W.E. (1979), "The Theory of Total Factor Productivity Measurement in Regulated Industries", University of British Columbia, Vancouver.

Economic Policy Division, "Rate Adjustment Formula, An Overview and Assessment", DPC, July 1975.

Electricité de France, "Le Surplus de Productivite Comme Critere de Gestion", Internal Paper, Circa 1970.

Electricité de France, "Les Progres de la Productivite Globale des Fractures a Electricite de France en 1975", Commission de l'Exploitation du 24 juin 1976.

Griliches, Z. (1964), "Notes on the Measurement of Price and Quantity Changes" in Models of Income Determination, Studies in Income and Wealth, NBER, Princeton University Press.

Hulten, C.R. (1973), "Divisia Index Numbers", Econometrica, 41, 1017-1026.

Hulten, C.R. (1975), "Technical Change and the Reproducibility of Capital", American Economic Review, 65, 956-965.

Jorgenson, D.W. and Z. Griliches (1967), "The Explanation of Productivity Change", Review of Economic Studies, 34, 349-83.

Jorgenson, D. and K. Nishimizu (1978), "U.S. and Japanese Economic Growth, 1952-74: An International Comparison", Economic Journal, 88, 707-726.

Richter, M. (1966), "Invariance Axioms and Economic Indexes", Econometrica, 34, 739-55.

Squire, L. (1973), "Some Aspects of Optimal Pricing for Telecommunications", Bell Journal of Economics and Management Science, 4, 515-26.

Treadway, A.B., "What is Output? Problems of Concept and Measurement", NBER Conference on Productivity, Ottawa, Ontario, 1967.

Usher, D., "The Suitability of the Divisia Index for the Measurement of Economic Aggregates", Review of Income and Wealth, Sept. 1974.

Werner, M., "Productivity Based Planning Model for Teleglobe Canada", Proceedings of the Ninth International Teletraffic Congress, Vol. 1, Spain, 1979.

Werner, M. and J. Routledge, "Total Factor Productivity, Symposium 2, Quantification", Proceedings of a Symposium on Productivity Measurement, Vancouver, Nov. 1977.