6

CONSTANT AND VARIABLE PRODUCTIVITY ADJUSTMENTS FOR PRICE-CAP REGULATION

Ferenc Kiss

It has been suggested by several studies[1] that price-cap regulation with an adjustment formula for inflation and productivity is capable of offering incentives for regulated firms to improve their efficiency by minimizing costs under existing technologies as well as by improving technologies via increased innovation. It has also been suggested that the allocative efficiency of prices may improve under the price-cap regime, and that considerable social welfare improvement may result from it and from the greater efficiency of the production processes of regulated firms. Whether and to what extent the FCC's now emerging price-cap regulation for U.S. telecommunications will deliver these benefits depends, among other things, on how well its design and execution are adapted to the behavioral characteristics of input prices and productivity in the regulated sector of the U.S. telecommunications industry. Utilizing a broad range of information from empirical evidence on the productivity performances of regulated telecommunications carriers in the past to econometric productivity analysis, this article explores what is, paradoxically, perhaps the most important as well as the least understood aspect of price-cap regulation: productivity adjustments.[2]

The subject allows for certain simplifying assumptions with regard to the features of the regulatory regime.[3] A single aggregate price cap is assumed for all regulated outputs. The price cap is subject to contract negotiations between the regulatory agency and the regulated firm for specified contract periods. The contract periods are several years long. Once agreed upon, the price cap does not necessarily remain fixed until the next round of negotiations but becomes subject to intercontractual adjustments. The negotiating parties agree in advance on the precise method of adjustments and also set up both trigger mechanisms and

M. A. Einhorn (ed.), *Price Caps and Incentive Regulation in Telecommunications*

emergency procedures, in case significant unforeseen cost changes render the agreed upon formula inoperable. The negotiating parties agree that the adjustments are to reflect changes in the cost of production during contract periods. Cost changes are reflected in index numbers of input price, productivity, and specified exogenous cost items. The actual regulatory process consists of (1) negotiations and the resulting input price and productivity adjustment formulae; (2) post factum verification of compliance, say, once a year;[4] (3) punitive price adjustment if the regulated firm fails to comply; (4) re-negotiation at the end of the contract period.

The article consists of eight sections. Section 1 describes the general form of the price-cap formula and discusses the various ways in which the price-cap formula may offer incentives to the regulated firm to minimize its production costs. Section 2 deals with the importance of the accuracy of price adjustments in price-cap regulation. Section 3 discusses incentive compatibility for price adjustment formulae. Section 4 investigates the empirical performance of four productivity indexes which may be considered for the price-cap formula. Section 5 shows how price index adjustments would have performed for pre-divestiture AT&T during the 1970's. Since neither the investigated productivity indexes nor the hypothetical AT&T price cap are satisfactory, Section 6 takes an analytic approach and uses causal variables of productivity gains to construct a variable productivity adjustment formula, which can be expected to perform better than existing and proposed formulae. Section 7 details the author's recommendations for variable productivity adjustments, while Section 8 summarizes the recommended formula and evaluates its foreseeable weaknesses and advantages.

1. Output Price Adjustments and Incentive Factors

As several sources—for example, Vogelsang (1988)—suggest, the longer the contractual period the greater the incentive most versions of price-cap regulation offer to the regulated firm to increase its productivity. For this reason, periods are expected to be several years long. This article analyzes cost characteristics for four-year periods. It is likely that unit costs change significantly, in both foreseen and unforeseen ways, during longer periods of time. Hence the need to adjust price caps during contract periods. The price cap allows price adjustments, whenever unit costs change to such a degree that the regulated firm would earn either too much or too little profit under constant prices, relative to the terms of the existing regulatory contract. The objective of the adjustments is to keep the profits of the regulated firm within reasonable limits.

There are three generic sources of cost increases in regulatory contract periods with which price adjustments should concern themselves. These are:

1. Cost inflation; i.e., exogenous inflationary increases in the purchase prices of factor inputs which the firm uses in order to produce its output of telecommunications services.[5]

2. Productivity gains; i.e., factor input volume, and thus cost, savings relative to the volume of output produced due to the improved efficiency of the production process of the firm.
3. Exogenous changes in "prescribed" costs. A cost item is prescribed if its magnitude is determined by rules set by outside (government and other) agencies. The examples include taxes, access line charges for interexchange carriers, or changes in separations procedures for local exchange carriers. A cost change is exogenous if, and to the extent, it is due to changes in the rules that govern the cost items; e.g., changes in tax rates or the tax base. Bell Communications Research (1988) itemizes and estimates the magnitude of ten exogenous changes in prescribed cost items for local exchange carriers.

In addition to cost changes, imperfections in the price-cap formula, non-compliance with the price cap, errors in managerial decision making, and exogenous events may lead to changes in the economic profits of the regulated firm by creating either inadequate or excessive profits, or changes in their magnitudes. Changes in economic profits constitute the fourth component of the price-cap formula. In the telecommunications industry, it is not the price itself but the temporal price index of regulated services that is capped. Price-cap regulation is synonymous with limiting (capping) aggregate output price changes over time. Its most general form is:

$$\dot{p} = c_w \dot{w} + c_y \dot{y} - \dot{\varphi} - \dot{\pi}\,; \tag{1}$$

where $\dot{p}$ denotes the proportionate (percentage) change in the price cap over time (e.g., annually); $\dot{w}$ denotes the proportionate change in the price index of the non-prescribed inputs of the regulated firm; $\dot{y}$ refers to the proportionate change in prescribed costs; c_w and c_y are cost shares for the non-prescribed and prescribed factor inputs, respectively; $\dot{\varphi}$ is the annual total factor productivity (TFP) gain; and $\dot{\pi}$ is the proportionate change over time in the economic profit of the regulated firm. Applied post factum, this formula suggests that if inflationary increases in the purchase prices of productive factors alone have increased, say, 80 percent of costs by 4.75 percent ($\dot{w} = 0.0475$), and prescribed costs, amounting to 20 percent of the total production cost of the firm have undergone an exogenous increase of 8 percent ($\dot{y} = 0.08$), while the productivity of the firm has improved by 2.1 percent ($\dot{\varphi} = 0.021$) and there is no economic profit; then a 2.3 percent price increase ($\dot{p} = 0.023$) would leave the firm approximately in the same financial position as it enjoyed before the changes took place.

The components of the price adjustment formula of equation (1) may be endogenous or exogenous to the regulated firm. For example, $\dot{y}$ is exogenous by definition. While some components may be clearly one or the other, most contain both endogenous and exogenous elements. Both $\dot{w}$ and $\dot{\varphi}$ represent such mixed components, although $\dot{w}$ is typically mostly exogenous. Regulatory incentives are to be provided in order to influence the firm's decision regarding the endogenous factors of the price adjustment formula in a socially favorable manner. The incentives are built into the price adjustment formula as explicit or implicit incentive factors which allow the firm to have a share in the profits that result from realizing favorable changes in the endogenous components, and force the firm to carry part of the burden that results from unfavorable changes. Contract negotiations formulate expectations for the following contract period regarding each component of the price cap. A system of incentive factors is also agreed upon and explicitly or implicitly included in the adjustment formula. The general form of the price adjustment formula with a full set of explicit incentive factors is:

$$\dot{p} = c_w(\alpha_e \dot{w}_e + \alpha_\Delta \Delta\dot{w}_1 + \Delta\dot{w}_2) + c_y\,\dot{y} - (\beta_e \dot{\varphi}_e + \beta_\Delta \Delta\dot{\varphi}_1 + \Delta\dot{\varphi}_2)\ ; \qquad (2)$$

where subscript e refers to expectations as formulated in the regulatory contract, and the deviation between expectation and the subsequent actual value of both input price change and productivity gain is broken down into endogenous ($\Delta\dot{w}_1$ and $\Delta\dot{\varphi}_1$, respectively) and exogenous ($\Delta\dot{w}_2$ and $\Delta\dot{\varphi}_2$, respectively) parts. The price cap is fully adjusted for the changes in both exogenous deviations ($\Delta\dot{w}_2$ and $\Delta\dot{\varphi}_2$), while the benefits from both endogenous deviations ($\Delta\dot{w}_1$ and $\Delta\dot{\varphi}_1$) are shared between the firm and its customers according to the incentive factors α_Δ and β_Δ, respectively. The input price expectation $\dot{w}_e$ and the productivity expectation $\dot{\varphi}_e$ are also shared because they contain endogenous as well as exogenous elements.

Unfortunately, it is often impossible to distinguish between the endogenous and exogenous elements of changes in input prices and productivity. The information necessary to disentangle endogenous and exogenous changes is seldom available. Furthermore, in many cases, the endogenous and exogenous aspects co-exist and are intertwined to such a degree that it is not possible to "allocate" them, even in the presence of extensive information. Only the full deviations of the actual input price and productivity from their respective expected values are known and can be incorporated into the adjustment formula in the following manner:

$$\dot{p} = c_w(\alpha_e \dot{w}_e + \alpha_\Delta \Delta\dot{w}) + c_y\,\dot{y} - (\beta_e \dot{\varphi}_e + \beta_\Delta \Delta\dot{\varphi})\ . \qquad (3)$$

Their inability to separate endogenous and exogenous factors certainly hampers the ability of price-cap formulae to offer clear and strong cost minimizing incentives. While this weakness is widely regarded as incurable, a careful selection of incentive factors may help reduce its negative impact. For example, the value of α_Δ would normally be close to one because most deviations from expected changes in the input price index may be characterized as exogenous. Other considerations are mentioned below.

Several price-cap formulae exist or have been proposed in Great Britain and in the U.S. None of these formulae contains but they all imply a full set of incentive factors. For instance, $\dot{w}$ is not formally separated into $\dot{w}_e$ and $\Delta\dot{w}$. The input price index expectation is typically formulated by making it equal to some external price index such as the Retail Price Index (RPI) in Britain or the GNP deflator (PGNP) in the U.S. This solution implies that α_e=1 and α_Δ=0; i.e., the firm is allowed to pass on to its customers, in the form of higher output prices, the costs of all expected input price increases but must absorb the cost impact of any, exogenous as well as endogenous, deviations from the expectation. Very strong incentive is provided if the deviations are endogenous but the formula merely generates high risk instead of strong incentive if, and to the extent, the deviations are exogenous. This distinction is important because the influence on input prices of exogenous variables is normally much greater than that of endogenous variables. The chief danger in using external price indexes is that they may not be capable of reflecting accurately the input price index of the regulated firm and thereby they may increase the exogenous deviations. It is generally understood that exogenous deviations may be large in any given year, but it is hoped that the annual deviations largely cancel each other and do not materially affect the financial well-being of the regulated firm over regulatory periods of several years. The past performance of AT&T suggests that such hopes are not very well founded. Table 1 indicates that major exogenous price index deviations did occur and, by extension and in the absence of arguments to the contrary, may occur in the future. For AT&T between the late 1940's and the late 1970's, the reduction of price index deviations to comfortable levels would typically have required at least seven- or eight-year regulatory contract periods — too long to consider for regulatory contract periods.

Table 1: Absolute Values of Percentage Point Deviations Between Four-Year Moving Averages of the GNP Deflator (PGNP) and AT&T's Input Price Index

Period	Contemporaneous PGNP	Lagged PGNP
1968-71	1.9	1.1
1970-73	1.3	1.1
1972-75	0.1	0.9
1974-77	1.7	1.8
1976-79	0.9	0.7
Average	1.2	1.1

PGNP is lagged by one year; i.e., the input price change of AT&T in 1968-71 is compared to the change in PGNP in 1967-70, etc.

For the productivity component, β_e and β_Δ identify the customers' reward and $1 - \beta_e$ and $1-\beta_\Delta$ represent the regulated firm's reward and, thus, incentive. $(1-\beta_e)$ $(1-\beta_\Delta)$ will function well if the deviation is largely endogenous, and the opposite relationship is advisable if the deviations are expected to be largely exogenous. Failure to distinguish in an explicit fashion between expectation and deviation (i.e., the setting of a single expected productivity gain such as the constant percentages of British Telecom and AT&T) implies that β_e=1 and β_Δ=0. The interpretation of these values is that (1) the regulated firm is not given a share of profits from expected productivity gains; (2) the customers do not share with the regulated firm the profits or losses from unexpectedly high or low productivity gains. The expected productivity is implicitly assumed to be fully exogenous and the subsequent deviations from it fully endogenous. Since deviations from expected productivity tend to be very large and also tend to be influenced rather strongly by exogenous developments, this is an extreme case of a high risk and high reward formula for the regulated firm. Risk and reward are further increased if the productivity expectation is not allowed to change from year to year in response to changes in external conditions but must remain constant for the entire contract period. Table 2 contains deviations of annual productivity gains from four-year average gains for a number of firms. The table indicates that the annual deviations may be very high, and that even their average absolute values tend to be uncomfortably high. Naturally, the deviations further increase on average if the average annual productivity gain for the four-year contract period is not accurately foreseen.

Table 2: Average Absolute Values of Percentage Point Deviations of Four-Year Average Productivity Gains from Their Annual Components

Period	AT&T	AGT	BELL	BCT
1968-71	1.76	n.a	1.38	n.a
1970-73	1.65	2.04	1.74	n.a
1972-75	0.45	3.29	1.07	n.a
1974-77	0.43	4.26	2.66	3.86
1976-79	0.35	3.99	0.60	2.98
1978-81	n.a	1.26	0.54	1.99

AGT: Alberta Government Telephones, BELL: Bell Canada, BCT: British Columbia Telephone.

2. The Importance of Accuracy

The price-cap formula with incentives excludes the $\dot{\pi}$ term of equation (1). Regulatory contract negotiations are supposed to arrive at an adjustment formula which contains the perfectly foreseen future annual input price and total factor productivity changes of the regulated firm and, thus, generates neither too low nor too high profits for the regulated firm, so that $\dot{\pi}$=0. Non-zero $\dot{\pi}$ values may nevertheless occur either because of non-compliance with the price cap or because of error (inaccuracy) in the adjustment formula. Ad hoc (non-formula-based) punitive action may apply for the former case. While a correction to the profit level would be desirable in the case of inaccuracy, formula-based routine corrective action would probably cause more harm than benefit, as a non-zero $\dot{\pi}$ would undo the incentive for cost minimization by removing the profits due to unexpectedly favorable endogenous factors and rewarding unexpectedly poor endogenous factors by compensating them for their mistakes. Changes in economic profits that are due to imperfect foresight are best left unaddressed until the re-negotiation of the regulatory contract. However, even during re-negotiation, the regulators must handle the issue of economic profits with care, since the removal of profits, whether due to exogenous factors or unexpectedly good performance by management and employees, would trigger attempts by the regulated firm to hide productivity improvements by generating temporary input price and volume increases during the last year of the regulatory contract period. Forward shifts over time in deliveries of, and payments for, factor inputs would achieve very much the same effect as simple cyclical goldplating. The amount of thus created "cost reserve" would depend on the amount of expected profit adjustment in the re-negotiated regulatory contract.

Inaccuracy may ultimately defeat the regulatory reform itself. A formula with a low degree of accuracy would make it necessary for the regulator and the regulated firm alike to keep a close eye on profits and rates of return. Rates of return would have to be reported, analyzed, and discussed in public, and corrections due to inaccurate adjustments would have to be designed in order to return profits and rates of return to their respective acceptable ranges. Inaccurate price-cap regulation would revert to rate-of-return regulation.

Accuracy is largely a matter of the flexibility of the formula (its ability to react to a variety of unforeseen cost changes during contract periods) and the appropriateness of the component index numbers (their ability to reflect the nature and magnitude of the relevant cost changes). The adjustment formula may be flexible to various degrees.

The most rigid adjustment formula prescribes the annual price adjustment as a constant for each year of the contract period. Such a formula assumes perfect foresight of cost inflation as well as productivity. For example, if perfectly foreseen annual inflation were 3.5 percent ($\dot{w} = 0.035$), "prescribed" costs were not to change ($\dot{y} = 0$), and the also perfectly foreseen annual productivity gain were 2 percent per year ($\dot{\varphi} = 0.02$), and if there were compelling reasons to believe that the regulated firm should keep the profits of one percentage point's worth of productivity improvement, it would follow that the price cap should be adjusted upward by 2.5 percent every year until the next contract ($\dot{p} = 0.025$).

A somewhat less inflexible formula would allow a measure of cost inflation to influence the price-cap adjustment, but would pre-set the expected productivity performance. The required annual productivity adjustment may be constant for the entire regulatory period. The current British Telecom and AT&T formulae belong to this category. The pre-set productivity adjustment may also be variable over the years of the regulatory period.

As an alternative to pre-setting productivity, the allowable cost inflation could be pre-set for the contract period and the adjustment could be made sensitive to some measure of the firm's productivity gain. This would be more logical and safer to introduce because inflation is measured, analyzed, and forecast by a large number of forecasters; thus, it is likely to be more accurately foreseen than the largely unknown productivity performance of telecommunications carriers. For this reason, it is somewhat surprising that no proponent of price cap regulation has recommended a, say, "4-TFP" formula, in which expected inflation is 4 percent per year and the adjustment is sensitive to some measure of gain in total factor productivity (TFP). The reluctance of regulators and regulated firms to adopt such a semi-flexible solution can be explained to a large degree by the complex and expensive nature of total factor productivity measures and analyses.

Finally, the most flexible formula would allow the price cap to change according to index measures of both cost inflation and productivity. Accuracy would be best served by formulae of this class.

3. Incentive Compatibility

Apart from the obvious requirement of being able to accommodate the various desirable incentive factors, the incentive compatibility of the price adjustment formula also requires (1) accuracy in forecasting $\dot{w}$, $\dot{\varphi}$ and $\dot{y}$; (2) independence of the index measures from the negotiating parties, so that the relevant information cannot be manipulated; (3) simplicity both in the formula and in the processes of verification of compliance; (4) stability in the rules governing the formula and the verification of compliance.

Inaccuracy is not incentive compatible because it increases exogenous deviations from expectations, and it generates changes in economic profit which are undesirable and cannot be remedied without removing incentives and encouraging harmful strategic behavior by the regulated firm.

Independence is both desirable and controversial. It is desirable because it helps avoid data manipulation by the negotiating parties, and "minus-productivity" type price adjustments. The danger of manipulation is considerable. Productivity data can be manipulated in a number of subtle ways with significantly distorted end results. A minus-productivity adjustment situation[6] may develop, when lagged actual productivity gains are used in the price cap formula to adjust regulated prices in the future, or when the regulated firms collectively have a very large weight in the industry-wide measure that is used to adjust the price cap. Such arrangements may virtually eliminate the cost minimization incentive either by rewarding the regulated firm for all of its past cost inefficiencies through building those into future output prices, or by building their collective inefficiencies into their collective adjusted price cap.[7] The independence requirement is also controversial. Since the information that is necessary for the calculation of accurate indexes is in the possession of the regulated firm, and not available (or partially and insufficiently available) to disinterested third parties, there is a conflict between the requirements of accuracy and independence.

Simplicity is generally regarded as a highly desirable requirement. Indeed, several important practical considerations suggest that adjustable price caps can function successfully as incentives only if the adjustment formula is transparent and its components are readily understood by the negotiating parties as well as by the general public. A further advantage of simple formulae is that they offer an opportunity for easy verification. While the requirements of independence and simplicity seem to go hand-in-hand, there is an obvious conflict between the requirements of accuracy and simplicity. Since both input price and productivity changes are complex phenomena, simpler formulae tend to be less accurate.

One-sided solutions to this problem, favoring extremely simple formulae at the detriment of accuracy, represent probably the greatest source of danger to the success of price-cap regulation.

Of course, simple formulae are not necessarily as simple as they appear. Expectations are formed and negotiations are conducted based on lengthy and complicated analyses of input prices and productivity. Once the expectations are agreed upon, the negotiating parties may decide to use a "simple formula" to express them for public consumption. The question is whether the simplicity of the formula matters at all. While it does not seem to matter from the point of view of forming expectations (as long as the formula expresses them faithfully), it matters very much when the expectations fail to materialize. A semi-rigid formula such as British Telecom's is bound to succeed only when the productivity expectation turns out to be correct, but fails otherwise. Only slightly more complicated flexible formulae, on the other hand, that could be generated by the same negotiating process with an approximately equal degree of difficulty, would be able to reflect a range of unexpected developments and, as a result, would perform significantly better in the presence of such developments. Sections 6 and 7 describe such flexible formulae.

The stability of the rules of price adjustments is important from the point of view of incentives in the sense that the rules must be independent from the endogenous variables that influence the components of the price adjustment formula. At the same time, the rules should be flexible enough to be influenced by exogenous variables, so that the financial health of the regulated firm is not altered by the impact of changes in circumstances beyond its control. Since it is not possible in practice to distinguish between endogenous and exogenous changes in the price adjustment components, the stability of rules is simply extended by making them insensitive to exogenous as well as endogenous variables. Some sources go as far as suggesting that the formation of a constant productivity expectation is necessary to create incentive compatible stability in the price-cap formula. While a constant expectation is undoubtedly a simpler target than a variable one, its lesser accuracy alone makes it less incentive compatible than a variable expectation which reflects changes in the economic environment of the regulated firm.

4. Productivity Indexes

The task of choosing productivity indexes for price-cap adjustment is made very difficult by the nearly complete lack of productivity measures for U.S. telecommunications. Only a few agencies are engaged in the systematic and regular measurement of productivity gains in the U.S. economy, and only one, the American Productivity Center in Houston, Texas, offers sufficient disaggregation of data to allow for the measurement of the aggregate productivity gains of the telecommunications industry.[8] It is rather unfortunate that, in the absence of detail and background data, their results are practically unanalyzable. Other sources, including the FCC's own effort to construct a "dual" (price-index-based) measure of the TFP gains of the telecommunications industry, appear to lack adequate information for reliable productivity measurement. Productivity is not measured at all for individual federally regulated telecommunications carriers or for classes of such carriers (e.g., local and interexchange carriers).

Four possible productivity indexes are examined in this section. Due to the lack of empirical evidence, the evaluation of their strengths and weaknesses is supported by less and older empirical material than would be desirable.

4.1 Total Factor Productivity of the Regulated Carrier

The idea of using the firm's own measured and forecast TFP index may appear attractive at first. It is, after all, the firm's own productivity gain that is to be foreseen. The ideal measure is TFP, because only TFP is capable of reflecting the totality of factor input and, thus, cost savings, while partial productivity indexes miss or misrepresent savings in all omitted inputs. Furthermore, the measurement of the regulated firm's TFP would generate an in-depth understanding of productivity performance, and this understanding would likely improve the accuracy of productivity expectations.

The most important argument against the use of the firm's own TFP is that it is not incentive compatible because the disaggregated input and output data are under the control of the regulated firm and, thus, are subject to manipulations. Reliance on each regulated carrier's TFP measure would also increase, rather significantly, the direct cost of price-cap regulation. Finally, and paradoxically, there are several reasons to believe that the accuracy of $\dot{\varphi}$ that could be forecast by relying on the firm's own TFP measure is also suspect.

First, due to the joint use of many factor inputs between regulated and unregulated services, it is not possible to measure the TFP of regulated services alone. The TFP of all services, on the other hand, systematically overestimates (underestimates) the TFP of regulated services if unregulated services grow faster (slower) and are subject to a higher (lower) rate of technological change. The bias in all-service TFP surrogates would normally work to the disadvantage of the regulated firm.

Second, it is an even greater concern that a sensible application of past TFP gains to a future time period may seldom be possible. Perhaps the most striking characteristic of the annual TFP gains of telecommunications carriers is their great variation both in time and in space (among carriers). Table 3 contains annual TFP gains which show that this was indeed the case during the 1970's and early 1980's. Great variation in time makes it difficult or impossible to choose an appropriate historic TFP gain for the adjustment formula. Lagged historic annual gains are clearly inoperable. Average gains may be attractive at first sight, but closer inspection reveals some important problems. Average gains for short and intermediate periods (two to seven years) still tend to vary considerably, due to the presence in the average of extremely high or low values as well as to the existence of some cyclical movement in productivity performance. This can be illustrated by comparing the four-year average productivity gains of the four telecommunications carriers whose annual productivity gains appear in table 3. In the 1947 to 1979 period (not shown), the highest four-year average productivity gain of AT&T was 2.5 times as high as its lowest four-year average productivity gain. Several 4-year averages were very high or very low. For variable periods of time (depending on data availability), the similarly calculated multiplier is 2.6 times for British Columbia Telephone and 2.7 times for both Bell Canada and Alberta Government Telephones. The data reveal uncomfortably high degrees of variation and show that telecommunications carriers of radically different sizes, operating under considerably different conditions and having significantly different productivity improvements over time, have been remarkably similar with respect to the degree of variation in their four-year average annual productivity gains. There seems no reason to believe that the observed variation of annual TFP gains will not continue in the future. The available data also reveal that only much longer (eight- to 12-year) averages would reduce the variation in average annual productivity gains to comfortable levels. It is unlikely that such long periods can be considered for regulatory contracts.

Table 3. Annual Total Factor Productivity Gains of Four Carriers

Year	AT&T	AGT	BELL	BCT
1973	0.043	0.069	0.053	0.046
1974	0.038	0.135	0.058	0.090
1975	0.027	0.016	0.088	0.077
1976	0.045	-0.014	0.024	0.041
1977	0.037	0.038	0.011	-0.033
1978	0.049	0.090	0.024	0.041
1979	0.043	0.102	0.035	0.056
1980	n.a	0.099	0.040	0.112
1981	n.a	0.061	0.029	0.063

The problem of variation is especially serious under conditions which make TFP accelerate or decelerate for three- or four-year periods. When such conditions prevail, the use of long-term average measures of past TFP gains would have a tendency to rob the needy and reward the undeserving. Suppose that, due to uncontrollable reasons, the productivity gain of the firm decelerates (or simply decreases from its historic average level) in the contract period. If a long moving average of historical TFP gains is used to adjust it, the price cap will be adjusted to decelerating productivity very slowly and partially (the average is, say, 10 years long, while the contract period is only four), thereby prolonging the underestimation of the price adjustment. The use of a fixed average creates an even greater problem because it does not allow any correction at all. The underestimation may reach serious proportions and, because it is both large and prolonged, it may threaten the financial viability of the regulated firm. The process works in a similar fashion to the disadvantage of the customer if productivity gains increase or accelerate due to exogenous causes. Under accelerating or decelerating productivity, a 10-year average TFP gain in the price-cap formula is capable of causing far more harm or undeserved profit advantage than lagged annual measures—precisely because it eliminates the quick short-term fluctuation in productivity gains.

Productivity forecasting is a complex and involved process. It can be done with success only if the forecasting effort is extended to include detailed and regular economic and econometric analyses of the entire production process as well as a thorough understanding of, and deep involvement in, the planning and budgeting processes. The mere availability of historical input, output, and productivity data is not sufficient to guarantee its success.

4.2 Labor Productivity of the Regulated Carrier

Labor productivity is often substituted for TFP when the former is available and the latter is not. Labor productivity is easy to measure under price-cap regulation. Detailed revenue data are normally readily available and detailed price information is required for the verification of compliance with the price cap. Thus, direct output price indexes and indirect output volume indexes are fairly easy to obtain. Labor data do not create serious difficulties either, since the number and classification of employees and annual labor costs are readily accessible.

Table 4: Percentage Point Deviations Between the Labor Productivity Gains and Total Factor Productivity Gains of Four Carriers

Year	AT&T	AGT	BELL	BCT
1973	3.0	-4.3	0.2	1.1
1974	2.4	-2.5	0.6	0.9
1975	3.2	-6.1	5.1	9.3
1976	4.4	8.0	1.1	10.3
1977	0.2	-3.9	1.6	6.2
1978	0.1	2.4	-1.6	-5.2
1979	0.7	4.2	-0.2	-4.7
1980	n.a	-9.0	2.2	0.4
1981	n.a	2.3	-1.8	2.5

The accuracy of the labor productivity proxy is examined for four telecommunications carriers in the 1973 to 1981 period in table 4. The table contains the annual percentage point deviations between each carrier's TFP and labor productivity gains. Even though labor was a large input (30 to 40 percent of the total production cost was typically associated with labor during the late 1970's), the labor productivity gains appear to have been decidedly poor surrogates for TFP gains. There are large deviations between TFP and labor productivity gains for all companies in most years. Some deviations are gigantic. The table also shows that the labor productivity gain differs from the TFP gain in the long run. The main reason is capital-labor substitution and a historically rapidly increasing capital-labor ratio. Since labor does not normally increase as fast as total input, labor productivity would systematically overestimate the productivity gain if used in price-cap formulae. By doing so, it would create a downward bias in price adjustments—to the detriment of the regulated firm.

The use of labor productivity measures as surrogates for TFP is not incentive compatible. It would provide the firm with a powerful incentive to increase the prices of its regulated output by substituting labor for capital, so that the labor productivity gains would be lower. Non-optimal factor proportions and at least some retardation in technological progress would be the result of such a practice.

4.3 Total Factor Productivity of the Telecommunications Industry

The appropriateness of industry-wide TFP measures is mainly a matter of the degree of variation of annual TFP gains in space; i.e., among carriers. The annual TFP gains of four carriers in table 3 reveal that the use of industry-wide productivity measures would create two serious problems for price-cap formulae. The lesser problem is that the annual productivity gains vary greatly among firms in most years; thus, there would be large deviations from any average value as well. The deviations are similarly large between annual firm-specific gains and lagged average values. The result of using any kind of industry-wide measure would be a quick succession of large windfall profits and losses. The second problem is that, depending mostly if not exclusively on the economic conditions under which they operate, some firms perform consistently below the industry average, while others perform, with comparable consistency, above it. This phenomenon raises the possibility of permanently punishing some and rewarding other regulated firms for the exogenous characteristics of their economic environment. Both the reward and the punishment may be large. To mention an extreme Canadian example, during the last five years shown in table 3, the productivity growth of AGT was nearly three times as high as that of Bell Canada and more than 60 percent higher than that of British Columbia Telephone. As figure 1 demonstrates, the productivity gains of AGT are exceptionally strongly correlated with their output growth rates, determined largely by increases in demand which, in turn, were related primarily to the oil boom in the area served by AGT. Had a price-cap formula with industry-wide productivity been in existence, it would have given a great opportunity for AGT to increase prices and profits, while punishing other carriers for the very high productivity gains of AGT. Similar cases may easily occur among the potentially more greatly different geographic areas of the United States. It is unfortunate that the deviations of firm level productivity gains from the average productivity gain of the federally regulated part of the American telecommunications industry are unknown and the FCC is unable to analyze this potentially very serious problem.

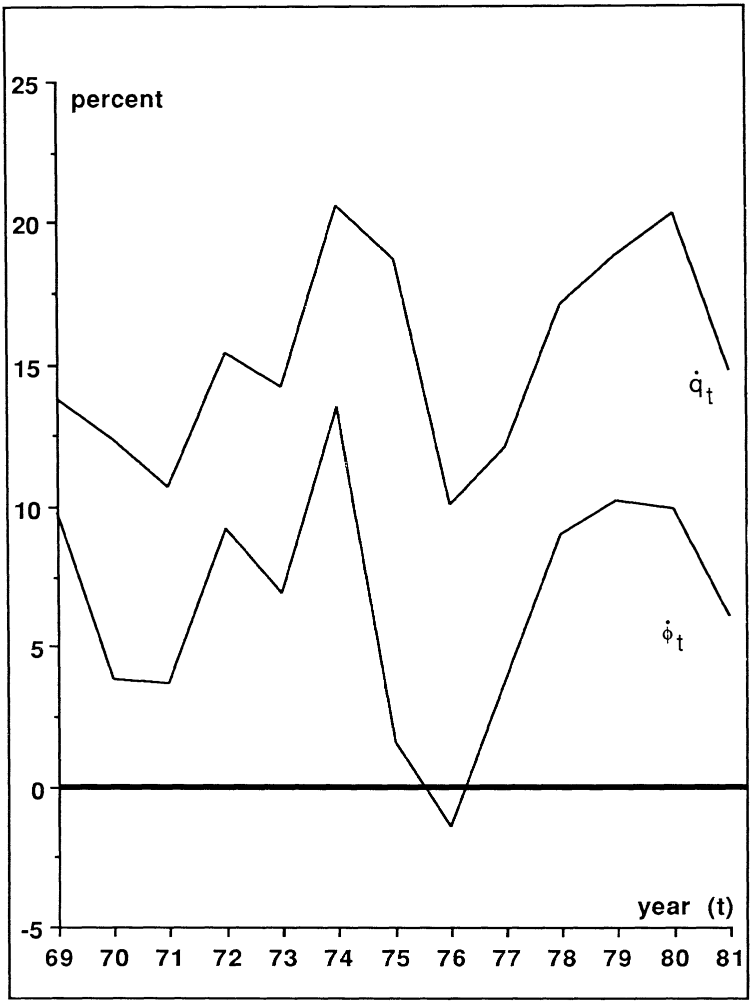

Figure 1:
Annual growth rates of output ($\dot{q}$) and productivity ($\dot{\phi}$)
for Alberta Government Telephones

4.4 Indirect (Dual) Productivity Measures

If the economic profit of the firm does not change ($\dot{\pi} = 0$), the productivity gain of the firm can be expressed as the difference between the change in its input price level and the change in its output price level:

$$\dot{\varphi} = \dot{w} - \dot{p} \,. \tag{4}$$

This alternative expression of the productivity gain, in which the prices (and not the volumes) of output and input appear, is sometimes referred to as the "dual" measure of productivity. Since $\dot{p}$ is the ultimate unknown variable of the price-cap formula, the firm's own forecast data cannot be used to calculate the productivity gain expectation. Past productivity gains may be calculated if the $\dot{\pi} = 0$ assumption is approximately valid and major "prescribed" cost changes do not disturb the index relations. However, the difficulties of using past TFP gains to form expectations for contract periods, discussed above in Section 4.1, would apply.

Another option is to find external surrogates for $\dot{w}$ and $\dot{\varphi}$. If sufficiently accurate and incentive compatible price index surrogates for the outputs and the inputs of telecommunications carriers can be found the dual measure offers further possibilities in designing price-cap formulae.

Both the Consumer Price Index (CPI) and the Producer Price Index (PPI) of the Bureau of Labor Statistics (BLS) contain telephone price components. The CPI has sub-indexes for "telephone service," "local service," "intrastate toll," and "interstate toll," while the PPI has sub-indexes for "local service," "toll service" (broken down into intrastate, interstate, international, and WATS), "interstate private lines," and "directory advertising."[9] For a detailed description and extensive analysis, see Lande and Wynns (1987). General inflation, as measured by the CPI or PPI, may be considered a surrogate for the input price indexes of regulated telecommunications carriers. If the telephone components of the CPI or PPI can also be used to measure the output price indexes of the same firms, then, at least in concept, they also yield a measure of the productivity gains according to the equation above. Thus, it may be possible to derive productivity measures for use in price-cap formulae from either the CPI or the PPI of the BLS.

Lande and Wynns (1987) note that the CPI has increased eight-fold, while its telephone service component approximately doubled since 1935. This translates to a roughly 5 percent annual input price increase and a 1.7 percent annual output price increase. The implied annual productivity gain corresponds to the primal (volume-based) measure of the long-term performance of AT&T. It seems that for very long periods of time the CPI may offer a sufficiently accurate measure of the productivity gain of the telecommunications industry. Unfortunately, the intermediate-and the short-term performances of the CPI are unacceptable.

The intermediate-term performance can be investigated during the 1978 to 1986 period. When average annual changes of the CPI and its telephone components are calculated in a number of alternative ways,[10] the results indicate average annual productivity changes anywhere between a loss of 2.1 percent and a 1.6 percent gain for the aggregate telephone service category. For local services, the average annual productivity change is a loss which ranges from 1.0 to 5.1 percent. Average annual gains between 0.5 and 4.0 percent are calculated for intrastate toll, while interstate toll is shown to have had a higher (4.3 to 6.5 percent) average annual gain in productivity. The derived values are obviously unrealistic and inconsistent. The average productivity gain of the local and two toll categories is inconsistent with the productivity gain that is directly calculated for the total telephone service category. Only unrealistically high local and low interstate toll weights could yield equivalents to the direct result.

The annual calculations offer an extremely distorted picture. For the aggregate telephone service category, very high productivity gains are shown for 1978 to 1980 and productivity losses appear in every year from 1981 to 1986 (with the exception of 1983, when a gain of 0.2 percent is registered). Local service productivity gains are parallel to those of the aggregate category, but the estimated TFP changes are even more unrealistic. For example, a 4 percent increase in the all-item CPI and a 17.1 percent increase in the local service CPI yield an indication of a local service productivity loss of 13.1 percent during the year 1984. Further losses of 5.1 and 6.0 percent are shown for 1985 and 1986, respectively. Intrastate toll productivity is shown to have remained virtually unchanged from 1981 to 1986.

What explains the failure of the dual measure? The all-item CPI is not the source of the problems. The average annual change in the CPI was identical to that in the input price index of AT&T for the period 1970 to 1979 and it is likely that the gradually declining annual changes in the CPI more or less corresponded to the gradually declining inflationary pressures on the costs of telecommunications carriers during the 1980's. The distortions seem to originate from the telephone price components. The first problem is the underestimation of the industry's productivity improvement. In this regard, Lande and Wynns (1987) note that telephone prices are lagged by one to three years behind the all-item CPI. In a period of rapidly decelerating inflation, such a lag would result in relatively high telephone price increases and can considerably underestimate the dual measure of productivity improvement. Such an underestimation took place after 1981. During periods of accelerating inflation, the lag in telephone prices achieves the opposite effect and the dual productivity is overestimated. This happened at the end of the 1970s. The second problem is a multitude of distortions in the local and toll price changes. The underestimation of local service productivity can be attributed to upward distortions of changes in the local telephone prices. Lande and Wynns (1987) analyze this problem in great detail and mention weight problems in the CPI as well as some specific events which caused overestimation. The outcome of their analysis is that the use of outdated expenditure patterns (based on data from the early 1970s) overstated the local price increase. The weights were updated in 1987; however, it seems that even the new weights, which rely on expenditures in the early 1980s, are outdated in that they do not reflect the continuing rapid change in the structure of expenditures in favor of interstate toll services. Among the special events, Lande and Wynns mention the effect of detariffing CPE, the appearance of separate charges for inside wire maintenance, federal subscriber line charges, and changes in federal excise taxes. The most important single event, however, is the general movement towards cost-based prices.

Lande and Wynns find that for a long period of time, movements in local and toll service prices were very similar. Their similarity seems justified from the point of view of cost inflation, since it is unlikely that the inputs of local and toll services were subject to significantly different inflationary pressures. On the other hand, the observed similarity of price changes is contrary to the perceived productivity movements. It is generally accepted that toll services achieved considerably higher productivity improvements than local services, on account of both their faster growth rate and their faster introduction of new cost-saving technologies. As the differences in productivity gains were not reflected, the prices of telecommunications carriers grew increasingly less cost based. As Lande and Wynns observe, the 1980s have witnessed a major correction in prices. Local rates increased not only because, and not to the extent, inflationary cost increases exceeded the productivity gain of local carriers, but mainly because they were judged to have been too low relative to the cost characteristics of their production process. Rule changes and new rules concerning the method of cost recovery,[11] mentioned by Lande and Wynns, added further elements of distortion to the picture. The counterpart of the downward distortion of local service productivity is the upward distortion of the productivity of toll, especially interstate, services. This is particularly serious if one considers that the very large capacity additions by the OCCs during the 1980s have not been well utilized and, thus, considerably reduced the productivity of the interstate service industry as a whole, and that this reduction has not manifested itself in higher prices because the competitive pressure of AT&T's price reductions has forced the OCCs to lower, instead of increasing, their prices. As a result, the OCCs failed to earn adequate profits.

In sum, the dual measure of productivity can be useful only if the absence of abnormally high or low profits can be ensured, and if the rules and regulations are stable and do not distort price movements in significant ways. In the present transitory state of the U.S. telecommunications industry, the prevalence of the conditions that would guarantee the satisfactory functioning of dual measures cannot be ensured and is not likely to occur. The adjustments of reported prices and price indexes that would be required to eliminate the effect of price distortions would be numerous, difficult, costly, and controversial.

4.5 The "Double-Dual" Method of the FCC

In an effort to overcome measurement difficulties caused by the absence of productivity data in telecommunications, the FCC has constructed a method which may be called "double-dual" because it is based not on volume information but on the comparison of input and output prices for the economy and for the telecommunications industry. FCC (1988) observes that PGNP "reflects certain productivity gains in the economy," then suggests that equal output price changes imply equal productivity gains and, furthermore, that the absolute values of output price change differentials and productivity gain differentials are equal. Observing an approximately 2.5 percentage point differential between changes in PGNP and telecommunications service prices, FCC (1988) comes to the conclusion that on average the TFP gains of the telecommunications industry have been and, by extension, can be reasonably expected to continue to be, about 2.5 percentage points higher than the TFP gains of the economy in general.

Denoting variables of the economy by capital letters and those of the telecommunications industry by small letters, the FCC's suggestion may be expressed as:

$$\dot{P} - \dot{p} = \dot{\varphi} - \dot{\Phi} = 0.025\,. \tag{5}$$

In addition to the already discussed difficulties of applying past long-run average annual productivity gains as expectations for future four-year periods (see Section 4.1), there are three major difficulties in this approach. First of all, it is not consistent with the use of PGNP in the price-cap formula. Given that $\dot{P} = \dot{W} - \dot{\Phi} - \dot{\Pi}$ and $\dot{p} = \dot{w} - \dot{\varphi} - \dot{\pi}$, the equality of the price and productivity differential holds only if $\dot{W} = \dot{w}$ and $\dot{\Pi} = \dot{\pi}$. However, the use of PGNP (which is, by definition, the output price index of the economy) in the price-cap formula as a proxy for the input price index of the regulated telecommunications industry implies that $\dot{P} = \dot{w}$ which, in turn, implies that $\dot{P} = \dot{W}$ and therefore $\dot{\Phi} = 0$ (if $\dot{\Pi} = 0$). In words, it is implied that the economy cannot have a productivity improvement. Consistent with this result is the conclusion that $\dot{P} - \dot{p} = \dot{w} - \dot{p} = \dot{\varphi}$; i.e., the productivity gain measured by the price differential is not the deviation from the economy's productivity gain but the entire productivity gain of the telecommunications industry.

The second difficulty is that the FCC formula cannot offer valid information on productivity unless its two implicit assumptions ($\dot{W} = \dot{w}$ and $\dot{\Pi} = \dot{\pi}$) are validated. Unfortunately, the measurement of the input price index of the economy as a whole raises difficulties not only because of unavailability of data but also because the identification of the desired factor inputs is by no means a simple task. Even greater difficulties are presented by the rate of change in economic profits. It seems that in order to overcome some of the complexities of productivity measurement, the double-dual method of the FCC has to deal with and resolve even greater complexities.

The third difficulty stems from the strong possibility of distorted price changes over time in the telecommunications industry as well as in the economy in general. Several distortions of telecommunications service price changes have been discussed in some detail in connection with the dual method in the previous sub-section. The double-dual method is further influenced by disequilibria and price distortions in the economy in general. For this reason, its use is even riskier than the use of the simple dual method.

5. How Price Index Adjustments Would Have Performed for AT&T

The behavior of input prices and productivity for a wide range of telecommunications carriers indicates sufficient similarity both in space and in time to suggest that an examination of how well price index adjustments would have performed for pre-divestiture AT&T during the 1970s may be instructive. In table 5, hypothetical price index adjustments are calculated for general inflation (PGNP) and productivity. The latter is based on the perfectly foreseen long-run (10-year average) productivity gain of AT&T. Bias in the price index adjustment is calculated by subtracting the proportionate change in PGNP from that in the AT&T input price index (inflation bias), and by subtracting the annual TFP gain from the 10-year average TFP gain (productivity bias). The total bias is the sum of the inflation and productivity biases.

Table 5. Short-run Percentage Point Biases in Hypothetical Price Adjustments for AT&T Between 1970 and 1979

Year	Inflation	Productivity	Total
1970	1.7	-2.8	-1.1
1971	1.4	-2.3	-0.9
1972	-3.6	0.8	-2.8
1973	5.9	0.9	6.8
1974	3.1	0.4	3.5
1975	-5.5	-0.7	-6.2
1976	-4.8	1.1	-3.7
1977	0.4	0.3	0.7
1978	-0.6	1.5	0.9
1979	1.5	0.9	2.4

Positive values denote overestimation and negative values denote underestimation of actual changes.

The annual biases are generally large. Out of seven possible four-year subperiods between 1970 and 1979, the period-end price level, the consequence of cumulative price adjustments during the four-year period would have been greatly overadjusted in one (1971-74), greatly underadjusted in two (1974-77, 1975-78), moderately overadjusted in two (1970-73, 1972-75), and accurately adjusted in two (1973-76, 1976-79). However, overall accuracy at the end of the latter periods would not have resulted from accurate inflation and productivity adjustments, but rather from large offsetting inflation and productivity biases.[12]

It is unlikely that price cap regulation in U.S. telecommunications in the 1990s will be as accurate as the AT&T example in table 5. First, AT&T's TFP showed less variation during the 1970s than in any other period. Variation of TFP gains may be higher in the 1990s. Second, larger firm size tends to reduce TFP gain variation. This is important because the telecommunications carriers of the 1990s will be considerably smaller than pre-divestiture AT&T. Third, the long-run productivity gain will not be perfectly foreseen. In contrast to these three reasons for poorer performance, price-cap regulation would improve to a considerable degree by the utilization of a variable productivity adjustment formula, recommended in the next two sections.

6. An Analytic Approach

The alternative to simple applications of any of a variety of available or possible measures of past productivity performances to the future is an "analytic approach" which relies on an analysis and understanding of the causes of productivity improvements, and uses estimated causal relationships to form expectations. Such an analytic approach has proved to be operable and useful in forming expectations of factor inputs, costs, and productivity under alternative sets of exogenous variables in at least one carrier's internal strategic planning and budgeting processes.[13] It is not a perfect solution in that, as explained below, part of the inaccuracy remains in the predictions but it demonstrably improves the accuracy of adjustments. It is also incentive compatible; i.e., it accommodates incentive factors, it is independent from the regulated carrier, it is simple, and it gives rise to sufficiently stable rules. Productivity predictions can be obtained and used to satisfy the requirements of price-cap regulation if (1) one can identify the main causal variables; (2) one can quantify, at least approximately, the relationship between causal variables and productivity; (3) the identified causal variables explain a sufficiently large portion of the changes in productivity; (4) one can predict the causal variables. The latter is not necessary if regulatory activity during intercontractual periods is restricted to the post factum verification of compliance.

The most important causal factors of productivity improvements are identified in numerous studies as (1) growth of output in the presence of economies of scale, and (2) technological changes. In addition, it has been suggested that non-optimal input proportions (deviations from the cost minimizing capital-labor ratio and other input ratios) have an influence on productivity. When the output prices are distorted (i.e., relative to the marginal costs of outputs), the output proportions change and productivity is influenced. Productivity is affected by the firm's inability to adjust its inputs to changing demand and other conditions instantaneously and without incurring significant costs of adjustment. Productivity is also influenced by a host of individual events and circumstances. Despite the existence and the occasionally great importance of these additional causal variables, the following discussion is restricted to growth and technology. Two major reasons explain the restriction. First, in numerous econometric models of telecommunications carriers, growth and technological changes successfully explain large parts of the productivity performance, while efforts to establish the relationship between productivity and other causal factors have not produced usable empirical results (even though considerable progress has been made). Second, the measurement and analysis of the additional causal variables represent a much higher level of complexity in the economic analysis and prediction of productivity gains.

A simple two-factor productivity decomposition is capable of estimating how much productivity gain is due to:

1. output growth in the presence of economies of scale, where a measure of overall economies of scale is allowed to express the relationship between total

input and total output, which relationship is influenced by capacity utilization, economies of scale, economies of scope, and economies of density;
2. the cost-saving effect of newly introduced technologies and technological improvements which, in turn, depend on how fast new technologies and improvements are introduced, and how much cost saving they are capable of generating.

All other positive and negative events and circumstances that may influence productivity are combined in the stochastic error term of the underlying econometric model.

The available extensive empirical evidence indicates that the relationship between the growth rates of total input and total output remains quite stable for longer periods of time and is quite similar among telecommunications carriers. Given output growth expectations or measured actual output growth rates (as in post factum applications of price-cap formulae) this (estimated or observed) relationship can be used with a considerable degree of safety in assessing the rate of productivity improvement which is generated by growth alone. Furthermore, the available empirical evidence suggests that the annual productivity improvements that are generated by technological progress are also fairly stable over time and similar among telecommunications carriers; therefore, their past magnitude can be quite safely predicted to hold for future regulatory contract periods of reasonable length. The existing empirical evidence also suggests that it may well be sufficient to predict productivity gains due to growth and technological change. The sum of these two effects explains a sufficiently large part of productivity improvements. Several sources contain productivity gain decompositions. For detailed empirical results, the reader is referred to the Bell Canada study of Kiss (1983).

The productivity gains of the adjustment formula may be composed as:

$$\dot{\varphi} = \mathrm{E}\,\dot{q} + \dot{T}\,; \tag{6}$$

where E is an elasticity which measures the degree of dependence of productivity gain on the growth rate of output, $\dot{q}$ is the percentage change in the volume of regulated output, and $\dot{T}$ is the annual productivity improvement due to cost savings generated by newly introduced technological improvements.

7. Recommendations

7.1 Coefficient Values

Currently available evidence on the causal components of the productivity gains of telecommunications carriers suggests that certain numerical values can be assigned to E and $\dot{T}$. Based on an exhaustive review of evidence, the following formula can be recommended for the productivity gain component of the price-cap formula for federally regulated telecommunications carriers in the United States for the period 1989 to 1992:

$$\dot{\varphi} = \beta(0.3\dot{q} + 0.01)\,. \tag{7}$$

β is a single productivity incentive factor which will be discussed at length in the next sub-section. E = 0.3 signifies that each percentage increase in the output of the firm is expected to generate a corresponding productivity gain of 0.3 percent, which is tantamount to saying that each percentage change in total input generates a 1.43 percent increase in total output. The latter relationship is often referred to as scale elasticity or the degree of overall economies of scale (ε). The scale elasticity is the inverse of the output elasticity of productivity; i.e., $E = 1 - 1/\varepsilon$.[14]

The value of E may be estimated with the aid of econometric production or cost models or, in some cases, directly observed and calculated from the input, output, and productivity measures of regulated firms and industries.

A large number of econometric cost models of AT&T and Bell Canada are surveyed by Kiss and Lefebvre (1987). Results for 16 representative econometric models are included in table 6. It is consistent with these econometric models to expect the degree of overall economies of scale in the 1.4 to 1.5 range.[15]

The degree of overall economies of scale is one of the most fundamental economic characteristics of the production process. It changes relatively slowly over time as (1) growth in the scale of production gradually exhausts the existing economies of scale, and (2) new economies of scale are generated, also normally gradually in the telecommunications industry, by the introduction of technological improvements. Kiss (1983, 91-96) describes the slowly changing nature of the

Table 6. Summary of Estimated Economies of Scale (ε) and Rates of Technical Change ($\dot{T}$)

Outputs	Author(s)	ε	$\dot{T}$
One	Nadiri-Shankerman	1.75	0.0120
	Christensen et al.	1.50 - 1.90	n/a
	Smith-Corbo	1.22	n/a
	Denny et al.	1.58	0.0068
	Kiss et al.	1.75	0.0083
	Kiss-Lefebvre 1.	1.73	0.0075
	Kiss-Lefebvre 2.	1.67	0.0084
Two	Smith-Corbo	1.20	n/a
	Kiss et al.	1.62	0.0130
	Evans-Heckman	1.39	n/a
	Charnes et al.	1.39	n/a
	Kiss-Lefebvre	1.38	0.0063
Three	Denny et al.	1.46	0.0057 - 0.0080
	Kiss et al.	1.43	0.0094
	Breslaw-Smith	1.60	n/a
	Fuss-Waverman	0.94	n/a

degree of economies of scale and show the influence on it of growth and technological progress. It seems that the value of E can be kept constant for four-year regulatory contract periods with a high degree of safety.

Sometimes it is easy to do non-econometric calculations. For example, figure 1 demonstrates that during the 1970's the productivity gains of AGT depended very strongly on its output growth — and on very little else. It is directly visible to the "naked eye" that the productivity gain was 30 to 50 percent of the output growth rate in most years, and about 40 percent on average. Hence E = 0.4. When, on the other hand, output growth is not very fast and other factors also play important roles, the value of E cannot be observed for individual years. However, the long-term relationships are still approachable. AT&T is a good example. Christensen (1981) reports that AT&T's long-run average annual output growth was $\dot{q} = 0.074$ for the period 1947 to 1979.[16] If E = 0.3 and $\dot{T} = 0.01$ the formula yields a productivity gain of $\dot{\varphi} = 0.032$, which is equal to the actual long-run average productivity gain of AT&T, reported by Christensen for the same period. Since the output and productivity growth rates ($\dot{q}$ and $\dot{\varphi}$, respectively) are available for some foreign telecommunications carriers, the numerical values of E and the corresponding degree of economies of scale can be calculated under the assumption that technological improvements generate a 1 percent improvement in productivity per year. The results are displayed in table 7. The output elasticity of productivity is calculated as $E = (\dot{\varphi} - 1)/\dot{q}$, and the degree of economies of scale is $\varepsilon = 1/(1 - E)$.

It is interesting to observe that the calculated E is very close to the recommended value of 0.3 not only for AT&T but also for five foreign (one German, one French, and three Canadian) carriers. This is strong evidence that the most fundamental economic characteristics of technologically sophisticated telecommunications carriers tend to be similar under a wide range of circumstances (such as geographic

Table 7. Computed Output Elasticities of Productivity (E) and Degrees of Economies of Scale (ε)

Carrier	Period	$\dot{q}$	$\dot{\varphi}$	E	ε
AT&T	1970-79	0.0735	0.0335	0.320	1.47
AT&T	1960-69	0.0769	0.0330	0.299	1.43
AT&T	1950-59	0.0712	0.0324	0.315	1.46
AGT	1969-81	0.1419	0.0634	0.376	1.60
BELL	1968-81	0.0818	0.0388	0.352	1.54
BCT	1973-81	0.1138	0.0524	0.372	1.59
DB	1970-79	0.0851	0.0493	0.303	1.44
DGT	1975-80	0.1400	0.0600	0.357	1.55

DB: Deutsche Bundespost, Germany; DGT: Direction Generale des Telecommunications, France.

location, climate, customer density, size, ownership, regulation, national characteristics, etc.).

The need to calculate $\dot{q}$ does not represent additional regulatory burden. Regardless of the form of the price cap, a price index must be calculated for the regulated output in order to verify compliance with the adjustment formula. Since the price index exists and revenue indexes can be easily calculated from readily available revenue information, the derivation of an implicit volume index as the ratio of the revenue index and the price index for regulated output is a trivial task.

The constant $\dot{T}$ term of the decomposed productivity gain expresses the effect on productivity of cost savings generated by improvements in technology in each year of the contract period. In reality, this term is not constant but variable over time. Its value is determined by the rate of introduction of new technologies and the cost elasticity with respect to technological changes. It has been simplified into a constant over time because the survey of Kiss and Lefebvre (1987) concludes that its value was quite stable over time as well as among firms in the past. As table 6 indicates, nearly every one of the reported econometric models estimated it to be close to $\dot{T} = 0.008$. Two further considerations are important when forming an estimate of $\dot{T}$. First, the econometric models that yielded $\dot{T} = 0.008$ were estimated with data for carriers with a high (around 50 percent) revenue share of long-distance services. Since the rate of technical change was demonstrably faster for long-distance than for local services, a somewhat lower value for local exchange carriers and a correspondingly higher value for interexchange carriers would be appropriate. Second, the rate of technical progress is expected to accelerate in the future as a result of the incentives for innovation to be provided by price-cap regulation. The $\dot{T}$ of the future will be greater than the $\dot{T}$ of the past. While it cannot be reasonably expected that additional incentives would double efficiency improvements due to innovation (even a 50 percent increase would be radically overstated, especially in the long run),[17] it may not be unrealistic to assume that price-cap regulation has the promise to increase the rate of the already very fast technical change by about 25 percent. Thus, the desirable value can be relatively safely fixed at $\dot{T} = 0.01$. This value means that the regulatory agency will expect the regulated carrier to achieve at least a 1 percent productivity improvement from newly introduced technologies and improvements in each year.[18]

7.2 Incentive Factor Values

The productivity sharing incentive factor β must be small enough to generate sufficiently large profits (according to $1-\beta$) for the employees of the regulated firm to make their extra efforts to increase efficiency undoubtedly worthwhile. Cost minimization is by no means costless. The firm incurs tangible costs through its management activities to promote and organize efforts aimed at increasing productivity. In addition to the tangible costs thus incurred, there is a probably much larger intangible cost, which can be expressed in terms of increased personal sacrifices and risks of employees. Sacrifices are associated with self-adaptation to higher rates of change in the work environment (philosophy, style, organization, human relations, technical requirements), higher labor intensity, increased voluntary and unpaid overtime, increased willingness to take work and related problems home, etc. As the pace of change quickens, career risks also tend to increase in corporate organizations. Promotional opportunities may be threatened by changing performance evaluation norms. Entire career paths may be jeopardized if certain existing types of human capital become obsolete as a result of faster technological change. The incentive to be provided by the β factor must not only be greater than the sum of tangible and intangible costs, as perceived by the management and non-management employees of the regulated firm, but must exceed it by a wide enough margin to demonstrate in a convincing manner to the employees of the regulated firm that their efforts will be well rewarded.

There is a category of the cost of regulation which particularly strongly influences the desirable level of the β factor. This is the cost of short-run "emergency boosts" of productivity in unfavorable years. Whenever the exogenous economic conditions become unfavorable to such a degree that the "normal" productivity gain that can be expected under those conditions is less than the gain that is allowed for price-cap adjustment purposes, the regulated firm will encounter economically unjustifiable and undesirable profit reductions. Since inadequate profits have a number of harmful effects even in the short run, the firm's management has a strong incentive to "boost up" its productivity gain by saving factor inputs which may be dispensable for some time, but not in the long run. Labor and material savings are the best examples of such short-run savings, but capital may be "saved" as well by the postponement of some elements of the construction budget. Such emergency boosts are costly and reduce productivity gains in the long run. (For example, short-run lay-offs lower labor productivity and increase recruiting and training costs in later years.) They are contrary to the long-term cost minimizing behavior of the regulated firm. They are arbitrary—the harmful artifacts of the regulatory regime.

Both the frequency and the size of emergency boosts are reduced if the number of years in which the allowed price adjustment is inadequate for exogenous reasons is reduced. The β incentive factor achieves such a reduction by lowering the productivity component of the price-cap formula. Figure 2(a) depicts a regulated firm with an increasing trend of productivity improvements over time and rather strong variation among the annual productivity gains. Assuming constant productivity adjustment and perfect foresight, the expected productivity gain is equal to the period average. Out of a total of 14 years, the annual gains are above the expectation in seven and below it in the other seven years. By reducing the productivity adjustment in the price cap, $\beta = 0.7$ would eliminate profit reduction, and with it the need for harmful emergency boosts, in four out of seven years. Due to the chosen numerical value of the β factor, the regulated firm encounters financial difficulties only in 21 percent, and not 50 percent, of the time. For four-year regulatory contract periods, the expectation of one year of moderate financial hardship (25 percent) may well be acceptable to the regulated firm but it is difficult to see its reason for embracing price-cap regulation if it means that it can expect financial problems, sometimes harsh ones, in two years out of four; i.e., 50 percent of the time.

Variable productivity adjustment further improves the situation because it reduces the deviations between the expected and the subsequent actual annual productivity gains. Figure 2(b) depicts the above described regulated firm with variable productivity adjustment. While the period-average productivity expectation and adjustment remain unchanged, the number of years of financial hardship is reduced from three to one (equivalent to 7 percent of the period's total time), and the extent of hardship is considerably reduced in the remaining one year. Alternatively, the value of β may be increased. In the example, the incentive factor could be increased to $\beta = 0.82$ and the "hardship percentage" would still be lower at 14 percent than in the case of constant productivity adjustment. Variable adjustment allows the setting of the incentive factor at a higher value, thereby benefitting the consumers of the products of the regulated firm.

What is a desirable value for β? An investigation of AT&T's productivity characteristics in the 1947 to 1979 period has yielded some interesting results. Using data from Christensen (1981), the 32-year period was segmented into four-year sub-periods. A total of 116 such sub-periods were distinguished. Assuming constant productivity adjustment and perfect foresight, the period-average productivity gain was calculated for each period, and the deviations of each year's actual gains from the average were taken. Hardship was indicated for 48 percent of the total time. The period-average productivity gains were multiplied by various values for the beta factor and the deviations were re-calculated for each year of each of the possible 116 sub-periods. With $\beta = 0.7$, the hardship percentage was reduced to 25.

There are reasons to believe that β should be lower than 0.7, when used in a scheme of constant productivity adjustments. First, it may well be desirable for the regulated firm to lower the hardship percentage somewhat below 25 percent. Second, hardship percentages may be higher in the future on account of smaller firm size and, thus, greater variation in the annual productivity gains. Third, the productivity expectations of the future will not be perfect. Imperfect foresight increases the risk of hardship. Based on a comparison of constant and variable productivity adjustment schemes, it has been concluded above that the value of beta may be higher if variable adjustment is used. However, the advantage of variable adjustment is overstated in figures 2(a) and 2(b). Variable adjustment yields less improvement over constant adjustment if the period of constancy is shorter. A constant adjustment for four years is considerably more "variable" than a constant adjustment for 14 years. Ultimately, the value $\beta = 0.7$ appears the most reasonable choice.

8. Summary and Evaluation of the Recommended Adjustment Formula

Setting the incentive factor at $\beta = 0.7$, the recommended price-cap formula becomes:

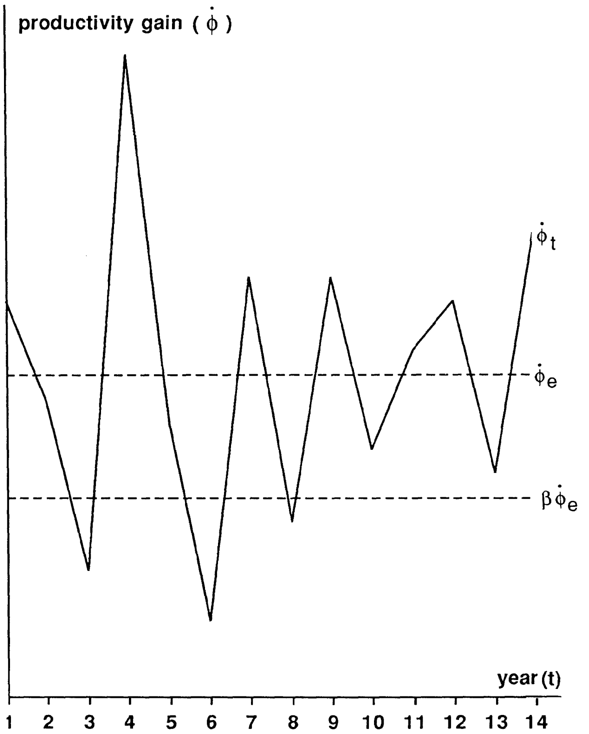

Figure 2/a:
Annual productivity gains ($\dot{\phi}_t$) and
constant productivity adjustments ($\dot{\phi}_e$ and $\beta\,\dot{\phi}_e$)
for a hypothetical telecommunications carrier

$$\dot{p} = \dot{w} - 0.7(0.3\dot{q} + 1.0)\,; \tag{8}$$

where, as suggested in FCC (1988), some lagged percentage change in PGNP may represent $\dot{w}$. In the case of 4 percent expected general inflation, no change in "prescribed" cost items, and a 4 percent growth in regulated output in a given year of the contract period, the post factum verification of compliance with the price cap would indicate that any aggregate price increase not in excess of 2.46 percent would be acceptable because $0.04 - 0.7\ (0.3 \times 0.04 + 0.01) = 0.0246$.[19]

The recommended formula is not nearly perfectly accurate. The decomposition of productivity does not explain the entire annual productivity gain. In addition to the growth effect and the technology effect, there is a residual, combining the impact on productivity of all other causal factors. While the residual is zero in the long run, its annual values may be fairly large. Therefore, the recommended adjustment formula does not eliminate—but it reduces—the problem of short-term

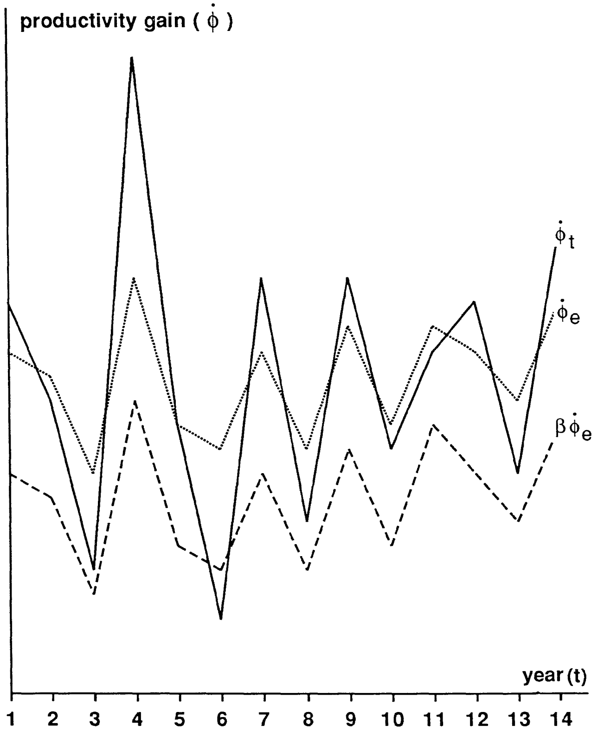

Figure 2/b:
Annual productivity gains ($\dot{\phi}_t$) and
variable productivity adjustments ($\dot{\phi}_e$ and $\beta\,\dot{\phi}_e$)
for a hypothetical telecommunications carrier

fluctuation in productivity gains. The remaining difficulty notwithstanding, the formula represents a significant improvement over alternative treatments of the productivity adjustment. The only telecommunications carrier for which detailed productivity decomposition is available is Bell Canada. It is evident from Kiss (1983, 92) that the actual annual gains are considerably closer to the "calculated"[20] than to the actual average annual gain. For the period 1970 to 1980, their average (absolute) deviation was 1.81 from the average gain, while it was 1.34 from the "decomposed" gain.[21]

It may also be perceived as a difficulty that the recommended formula necessitates an explicit agreement between the negotiating parties regarding the approximate degree of economies of scale (including scale, scope and density, and capacity utilization) and the rate of technical change. Such an agreement may be difficult to achieve, even if the pay-offs are considerable, because the results of econometric productivity analysis are imperfect, and its terms are not necessarily sufficiently familiar to the negotiating parties and to the general public. Nevertheless, the difficulties should not be overstated. Productivity adjustments (even in simple inflexible forms such as British Telecom's 3 and 4.5 percents) require that the negotiators develop some kind of a consensus regarding the causal variables of productivity gains and their expected impact in the future. Without analyzing the nature and role of causal factors, it is not possible to form a valid expectation regarding the magnitude of future productivity gains. The difference is not whether the causes are considered but whether they receive explicit or implicit consideration.

The recommended formula of variable productivity adjustment has several important advantages over its alternatives. First, by utilizing the firm's own output growth rates, the productivity adjustment becomes firm-specific, so that punishment and reward for different exogenous economic conditions are avoided. Second, because the adjustment depends on the growth of regulated output only, it is specific to the regulated output of the firm. This is important for two reasons. On the one hand, it is expected that with the advances of competition the number of regulated services will decline over time in the future. The recommended formula adjusts to the changing number of regulated services with ease. On the other hand, unregulated services are expected to grow faster than regulated services and, for this reason, to generate higher productivity gains. Higher productivity gains are also expected from the faster rate of introduction of new technologies into the production process of unregulated services. By not distinguishing between the productivity gains of regulated and unregulated services, price-cap formulae would overstate the expected productivity gain from regulated services, understate the necessary increases in the price cap, and thereby harm the regulated firm. Third, the recommended formula exhibits a great degree of flexibility. By being proportional to the growth rate of regulated services, it makes the price adjustment sensitive to the largest and most important determinant of productivity improvement. Furthermore, the flexible formula can be altered in simple ways. The proposed $\beta(E\dot{q} + \dot{T})$ can be changed into

- $\beta E\dot{q} + \dot{T}$, which would increase the incentive to innovate because the firm could keep the profits of all technology-related cost savings above the required minimum $\dot{T}$;
- $E\dot{q} + \beta\dot{T}$, which would increase the regulated firm's incentive to increase its output (presumably by lowering prices) because the profits of economies of scale would be kept by the firm in their entirety;
- $\beta_q\dot{q} + \beta_T\dot{T}$, which would allow the regulatory contract to balance the two incentives of the regulated firm, depending on specific considerations that may exist at the time of the contract negotiations.[22]

Fourth, even though the productivity adjustment is firm-specific, the ability of the firm to manipulate the calculations of output growth rate is minimal or non-existent. The regulatory agency has nearly perfect revenue and price information, enabling it to verify the calculation of the output growth rate or to carry out the calculations. This way, compliance with the independence requirement is fully ensured. Fifth, the formula is simple and represents no additional burden either on the regulator or on the firm. Sixth, the formula does not require that the results of productivity studies of either the firm or the industry be directly applied in the process of regulation. However, the ongoing study of productivity performance for a better understanding of the sources of productivity gains and the characteristics of productivity performance (and, thus, of the entire production process) becomes important from the point of view of the successful re-negotiation of the formula.

Notes

1. For example, NTIA (1987), Vogelsang (1988), FCC (1987, 1988).

2. The first manuscript of this article was written as an immediate reaction to the FCC's Note of Proposed Rulemaking in August 1987. An improved and extended version, identical in contents but somewhat different in appearance from the present article, was presented at the Massachusetts Institute of Technology in July 1988.

3. For the sake of the brevity and clarity of analysis, several otherwise important issues are consistently ignored. No discussion is offered on capped vs. uncapped outputs or possible reasons for treating classes of capped outputs differently. Problems associated with new products are not discussed either. Distinctions between local exchange and interexchange carriers, the relationship between federal and state regulation, and the problems associated with market domination are also ignored.

4. The alternative idea of the regulated firm submitting preestimates of rate revisions to the regulatory agency, presumably accompanied with proof of compliance, has several major shortcomings (cost, delay, rigidity, difficulties due to the simultaneity of price and volume changes, etc.) and, therefore, it is not embraced here.

5. Not all input price changes are exogenous. Allowance is given below for the endogenous nature of some input price changes.

6. "Minus-productivity" is the incentive-incompatible counterpart of "cost-plus" rate-of-return regulation.

7. In both cases, the dependence of the productivity adjustment on the regulated firm's productivity appears to be the cause of incentive-incompatibility. This appearance, however, is not necessarily true. The danger of providing a counter-incentive is avoided if the lagged or collective productivity measures are not used mechanically but serve instead as tools for an analysis and understanding of the productivity performance of the regulated firm and industry.

8. The Bureau of Labor Statistics (BLS) of the U.S. Department of Commerce also plans further disaggregation of input and output data.

9. These components of the PPI are not aggregated into a telephone price sub-index.

10. For periods of six to nine years, based on both year-end and average annual values of the indexes.

11. "Prescribed" cost transfers among major service categories.

12. Large and often clustered positive and negative biases would have generated great variation in consecutive period-end cumulative biases. For example, a one-year postponement of the start of the regulatory period from 1973 to 1974 would have turned accuracy into great underadjustment. A similar delay two years later (1976 instead of 1975) would have created the opposite effect. While AT&T's profits would have changed considerably by the end of most regulatory periods, making incentives inefficient and re-negotiations difficult, there seems to have been a single favorable choice of regulatory periods—namely, 1972-75 and 1976-79—which would have resulted in sufficiently accurate price adjustments and therefore smooth re-negotiations.

13. This statement is made with reference to my presentation entitled "Factor input and productivity forecasting for strategic planning and budgeting" at the Bellcore Economic Cost Modeling Forum in Atlantic City in September 1985.

14. The degree of economies of scale is the ratio between the output growth rate and the input growth rate; i.e., $\varepsilon = \dot{q}/\dot{x}$. Hence, if the degree of overall economies of scale is $\varepsilon = 1.43$, then a 1 percent increase in total input ($\dot{x} = 0.01$) will generate a 1.43 percent increase in total output ($\dot{q} = \varepsilon \dot{x} = 0.0143$). The

productivity gain will be $\dot{\phi} = \dot{q} - \dot{x} = 0.0043$, and E, which is the ratio between the productivity change and the output change, will become $E = \dot{\phi}/\dot{q} = 0.0043/0.0143 = 0.30$.

15. Keeping in mind that the indicator accounts not only for economies of scale but also for some forms of capacity utilization changes, economies of scope and economies of density.

16. The average year-to-year output volume index was 1.077. In traditional percentage terms, this is referred to as a 7.7 percent increase. The reader is reminded that this article uses logarithmic proportionate changes; thus, $\dot{q} = \log 1.077 = 0.074$.

17. More improvement may be expected in the first year or two of price-cap regulation if the regulated carrier has "reserves." However, in view of the very large budget cuts by American carriers throughout the 1980's, such reserves may be severely reduced or, more likely, non-existent at the beginning of the first regulatory contract period.

18. In addition to this direct impact, technological improvements also tend to improve productivity gains in the long run by increasing the degree of economies of scale. Without this indirect impact, the value of E could not be kept constant over time.

19. This would ensure an at least 40 percent absorption rate of inflation; i.e., at least 40 percent of cost inflation would be "absorbed" by cost savings due to productivity improvement and no more than 60 percent would be passed on to the customers of the firm in the form of output price increases.

20. The sum of the growth effect and the technology effect.

21. The unpublished results of Kiss and Lefebvre (1984) show a greater reduction of the average absolute deviation for Alberta Government Telephones.

22. These are only the "first order" flexibilities. It is possible to increase the flexibility of the formula further by introducing "second order" flexibilities, whose most general form is

$$\beta_{q1}E\dot{q}_1 + \beta_{q2}E\dot{q}_2 + \beta_{T1}\dot{T}_1 + \beta_{T2}\dot{T}_2,$$

where $\dot{q} = \dot{q}_1 + \dot{q}_2$ and $\dot{T} = \dot{T}_1 + \dot{T}_2$. In this formula, there is a minimum required output growth ($\dot{q}_1$) and technical change ($\dot{T}_1$), whose benefits are passed on to the customer partially or in their entirety (depending on the selected values of β_{q1} and β_{T1}), while the profits of additional output growth and technological progress are shared between the firm and its customers according to β_{q2} and β_{T2}.

References

Bell Communications Research. 1988. "The Impact of Federal Price Cap Regulation on Interstate Toll Customers." Manuscript. Livingston, NJ (March 17).

Christensen, L.R. 1981. "Testimony in United States v. AT&T." C.A. No. 74-1698.

FCC. 1987. "Notice of Proposed Rulemaking." CC Docket No. 87-313, Federal Communications Commission, Washington, DC (August 21).

FCC. 1988. "Further Notice of Proposed Rulemaking." CC Docket No. 87-313, Federal Communications Commission, Washington, DC (May 23).

Kiss, F. 1983. "Productivity Gains in Bell Canada." In *Economic Analysis of Telecommunications: Theory and Application*, edited by L. Courville., A. de Fontenay, and R. Dobell. New York: North-Holland, pp. 85-113.

Kiss, F., and B. Lefebvre. 1984. "Comparative Analysis and Econometric Forecasting of Factor Inputs and Productivity: Some Empirical Results in Canadian Telecômmunications." Fourth International Symposium of Forecasting, London (July).

Kiss, F., and B. Lefebvre. 1987. "Econometric Models of Telecommunications Firms: A Survey." *Revue Economique*, 38 (no. 2, March): 307-373.

Lande, J.L., and P.L. Wynns. 1987. "Primer and Sourcebook on Telephone Price Indexes and Rate Levels." Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission, Washington, DC.

NTIA. 1987. "Regulatory Alternatives Report." National Telecommunications & Information Administration, U.S. Department of Commerce, Washington, DC (July).

Olley, R.E., and C.D. Le. 1984. "Total Factor Productivity of Canadian Telecommunications Carriers." Project report to the Department of Communications and to members of the Canadian Telecommunications Carriers Association, Ottawa (February).
Vogelsang, I. 1988. "Price Cap Regulation of Telecommunications Services: A Long-Run Approach." The Rand Corporation, Note N-2704-MF, Santa Monica, CA (February).