**Alternative Approaches to Analysis of
Total Factor Productivity at the Plant Level**

Robert H. Schnorbus and Philip R. Israilevich

March 1987

Working paper series on

**Regional
Economic
Issues**

Federal Reserve Bank of Chicago

Alternative Approaches to Analysis of Total Factor Productivity at the Plant Level

Robert H. Schnorbus and Philip R. Israilevich*

I) Productivity Analysis and the Firm

The production function is the foundation of any economic analysis of productivity. In theory, the production function models the explicit relationship between a given output and all its inputs, or factors of production. Productivity, as a measure of economic performance, is often interpreted as an indicator of the relative efficiency with which those inputs are converted into outputs. While one of the most basic tools of economists, productivity analysis should also be a useful tool for managers to identify problem areas in their firms' production process and to formulate specific actions that improve profitability at the plant or divisional levels of the firm. Yet surprisingly, productivity analysis has not been integrated effectively into the decision-making process of the firm.

Managers are often skeptical of productivity measures, viewing them as biased, inaccurate, or analytically meaningless. Labor productivity (or, Q/L, where Q is output and L is man-hours), for example, is by far the most widely used measure of productivity.[1] Being a partial-factor productivity measure, however, labor productivity may cause misleading analyses. Improvements in labor productivity, for example, could be due to capital substitution, changes in scale economies, quality improvements in labor, better discipline, new work rules, improved management, and a host of other reasons unrelated to simply the more efficient use of labor (i.e., increased effort by labor directly). Moreover, changes in the use of one input, as might be induced by technological change, require a whole series of adjustments in the production process that will alter the most efficient combination of all factors of production. In fact, a technological change intended to increase labor productivity conceivably could reduce profitability by requiring costly adjustments in other inputs (e.g., higher quality requirements of materials). If so, what value are managers likely to place in the limited finding that labor productivity has increased?

Managers are concerned ultimately with the profitability of their investment and the necessary adjustments among factors of production that will have to be made in the whole production process in order to maximize that profit. They must be aware of the impact that technological change through their investment will have on the entire economic structure of their firm. As a result, they need a more comprehensive approach to productivity analysis than provided by simple comparisons of labor productivity.

An alternative to partial-factor productivity analysis is analysis of the combined productivity of all inputs, or total factor productivity (TFP). TFP can be derived from the firm's production function by taking the average physical product of all inputs that have somehow been aggregated. While introducing the problem of how to aggregate diverse types of inputs, TFP analysis avoids the problem of factor substitution that is inherent in partial productivity measures. In addition, it can be expanded to deal with scale economies, technological bias, and other important aspects of production theory of interest to managers. Indeed, it can be expanded through input-output analysis to capture indirect productivity effects that may occur in the production of inputs (i.e., a firm may produce its own machinery or some other intermediate input that are used in its production process), as well as direct effects in the production of the firm's final product.

Unfortunately, estimating production functions and deriving productivity meas ures at the plant level has usually required rather strict assumptions that managers dealing with the complexities of the real world often find unpalatable. Indeed, although it has now been thirty years since Solow pioneered major advances over partial productivity analysis, the controversy over how to measure productivity at the microeconomic level has—if anything—intensified.[2]

The basic issue clouding productivity analysis, therefore, is whether a production function can realistically be measured at the plant or firm level, much less be aggregated in any meaningful way to represent an industry production function. As an extreme example, consider a recent quote by a leading authority on productivity: "Most economists fail to realize that managers have no need for production functions at the plant level in the case of single production operations and regard them as meaningless in the case of multi-product operations."[3] If production functions are useless or meaningless at the plant level, what possible meaning can productivity measures derived from them have in the analysis of a multi-regional firm or a regional industry? Without the aid of an empirical production function, managers need an economic framework to evaluate all the partial productivities and an array of other economic variables within an economic context that interrelates all the variables—a network of productivity, cost, and profitability relationships.

With TFP analysis gaining much wider acceptance for analysis at the national level, the focus of current research has shifted to applications at the regional and plant levels. The techniques for measuring productivity that have been developed at the macroeconomic level are conceptually applicable at the firm or regional plant level, which would allow detailed interregional and interplant comparisons. For example, using TFP, knowledge of a regional plant's productivity performance relative to other plants in the firm or to the "average plant" in the industry could be used to identify areas in the production process where profitability of the firm could be improved by raising the plant's efficiency to some firm-wide or industry-wide standard.

Microeconomic applications of TFP analysis are just beginning to be refined, but few managers are aware of the alternative approaches available to them. This *Working Paper* examines the concept of the production function and alternative TFP measures for the purpose of evaluating their relative strengths and weaknesses. Section II presents the concept of productivity in the context of the theory of production. In sections III through V, three major approaches to TFP analysis—the network of productivity, cost, and profitability relationships; neoclassical productivity analysis; and input-output analysis—are described and evaluated with respect to their usefulness for managers. The final section summarizes the relative strengths and weaknesses of the three approaches from the perspective of the manager.

II) Production Functions and the Concept of Productivity

A firm's production function in its purest theoretical form relates the maximum amount of physical output, Q, that can be produced with the technologically efficient combinations of the relevant set of physical factor inputs. A generalized form of the production function would be:

II-1 $$Q = F(X_1, X_2, ...X_n, T)$$

where every conceivable factor used in the production of Q is assigned to a vector of n aggregated inputs, X_i , which are characterized by close substitutability within the X_i vector and strong separability between the X_i and X_j vectors.[4] The variable T for time is a proxy for technology A, or, in some sense, efficiency, and allows for any kind of shift in the production function that may occur over time. Technological change is only one—albeit the most interesting—of the sources of productivity growth that are shifting the production function outward.

To facilitate the description of the most common properties assumed by a production function, a simple two-input model is used:

II-2 $$Q = F(K, L, T)$$

where Q is output measured by value added, K represents the capital stock (or more appropriately, the services derived from capital) and L represents labor services. This is the most common form of production functions used in empirical studies. Its properties must be understood before alternative productivity measures can be properly compared.

Within the economically meaningful range of a well-behaved production function, average and marginal physical products of each input (i.e., Q/K and dQ/dK, respectively) are assumed to be positive, but at some point the rates of change in the marginal products are assumed to be negative (the law of diminishing returns). Holding output constant, the marginal rate at which labor can be substituted for capital (i.e., dK/dL) is assumed to be negative and, for convenience, the rate of change in substitution is negative (the principle of diminishing marginal rates of technical substitution).[5]

A measure of overall productive efficiency would combine all average physical products into a single index number that could then be used to measure the level of TFP (i.e., $TFP_t = Q_t/X_t$, where Q is aggregated output, and X is aggregated inputs at a given time period t). Since the absolute level of an index number is meaningless, TFP can be presented as a proportional growth rate (i.e., taking the derivative with respect to time and dividing by the value of the respective variable to get

II-3 $$dTFP/dt(1/TFP) = dQ/dt(1/Q) - dX/dt(1/X)$$

or, for simplicity,

II-4 $$T\dot{F}P = \dot{Q} - \dot{X}$$

where ($\cdot$) designates a proportional rate of growth). According to Kendrick, TFP "...indicates the relationship between real product in a given year to the real product that would have been produced (real factor cost) if the productive efficiency of the factors had been the same in the given year as in the base year."[6]

The inclusion of all inputs in the measure of productivity growth in effect controls for the factor-substitution problem. Each input's growth is weighted by its relative importance to total cost (i.e., $v_L = wL/(wL + rK)$ and $v_K = rK/(wL + rK)$, where w is the wage rate, r is the rate of return on capital, and v is factor share in a two input model that assumes constant returns to scale and producer equilibrium). In other words, the contribution to output growth, $\dot{Q}$, made just by increasing the amounts of all inputs, $\dot{X}$, is $v_L\dot{L} + v_K\dot{K}$, and the TFP equation can be re-written as:

II-5 $$TFP = \dot{Q} - v_L\dot{L} - v_K\dot{K}.$$

Now it no longer matters whether labor is growing faster than capital or, in effect, is being substituted for capital. It therefore follows that changes in total input, $\dot{X}$, capture movement *along* the production function and the difference between total output growth, $\dot{Q}$, and total input growth, $\dot{X}$, captures the *shift* in the production function or the change in productivity, *TFP*.

The classical illustration of how a shift in the production function is separated from movements along a production function can be summarized as follows.[7] In order to capture two inputs and one output on a two-dimensional chart, production functions for two different time periods (i.e., t and $t+1$) are plotted on a Q/L by K/L plane, where $1 = Q/L$ (labor productivity) and $k = K/L$ (capital-to-labor ratio, or capital intensity). (See chart 1.) Thus, the production function, Q_t, can be viewed in either of two ways: holding labor constant, output increases as more capital is added; or, labor productivity increases along the production function as capital intensity is increased. Rewriting the TFP equation in terms of the ratios that are relevant to chart 1 reduces the expression to

II-6 $$\dot{q} = TFP + v_K(\dot{k})$$

where $\dot{q} = \dot{Q} - \dot{L}$ and $\dot{k} = \dot{K} - \dot{L}$.

To demonstrate how a change in output is decomposed into shift and substitution components, assume that output moves from point M at time t to point N at time $t+1$. If capital intensity had not changed between time periods (i.e., $\dot{k} = 0$), then all growth in output would have been due to productivity (i.e., $\dot{q} = TFP$, which theoretically is the difference between points M and O). The problem is that O can not be observed, and O is the point on the production function, Q_{t+1}, with the same capital intensity, k_t , as the initial production function, Q_t . All that can be observed is N, which has a different capital intensity, k_{t+1} . Point N is used to estimate the production function, Q_{t+1}, and solve for k_t , in order to find O. But, a function can not be estimated from a single point. The solution proposed by Solow (i.e., forward estimation) was to draw a line tangent to Q_{t+1} at point N, which is a first-order approximation of Q_{t+1}. This tangent line can then be used to identify O*, which is the estimated value of O on the new production function with the same capital intensity as the initial production function.[8] The movement from M to O*, therefore, would represent the change in TFP and the movement from O* to N would represent factor substitution. (In fact, the distance between O* and N overestimates actual TFP growth by O* − O., which represents an inherent measurement error.)

CHART 1
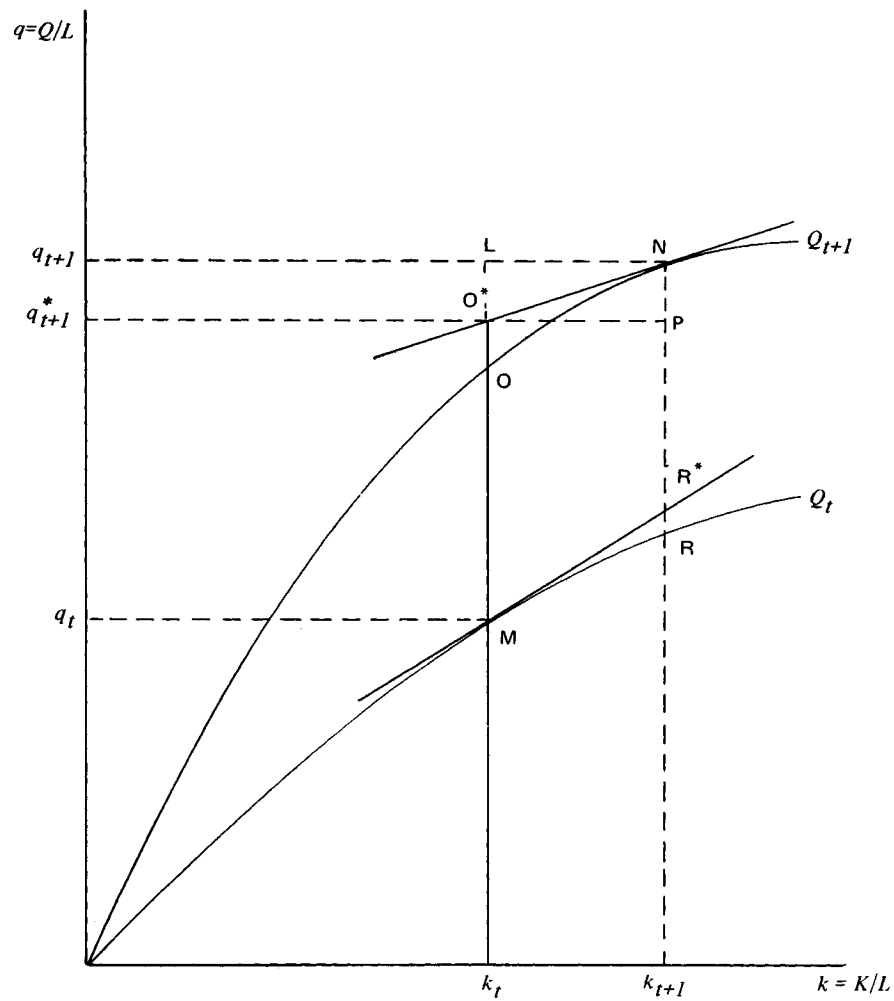
SOURCE: Adapted from Solow, 1957, p. 313

A second way to measure *TFP*, or backward estimation, would be to use linear approximation of Q_t, which would have a capital intensity of k_{t+1}. In this case, *TFP* would be the distance between R* and N, which is obviously too small by R* − R. One way to minimize the estimation error is to compute both measures of *TFP* and average them together. By averaging, the errors will tend to cancel each other out. Since the difference between using t and $t+1$ is essentially a choice between v_K at t and $t+1$, some researchers choose a second way to reduce the error, which is simply to use the average values of v_K.

Decomposing output growth into a shift in a production function and movement along a production function is relatively easy to conceptualize, but there are many other sources of growth that can be modeled in a production function. More complex properties of production functions are empirically difficult to incorporate into a production function, but should not be ignored if the intent is to isolate an unbiased measure of technological change. Among the most important of these properties that can also affect output growth are scale of production, substitutability of inputs, and technological bias.

Evaluating the effects of increasing output on productivity of inputs requires assumptions about economies of scale. For example, the most common assumption is that the production function is homogeneous to the first degree, which means that doubling all inputs will exactly double output—or simply, constant returns to scale. If homogeneity is greater than unity, the production function will exhibit increasing returns to scale (i.e., doubling all inputs more than doubles output). If homogeneity is less than unity, the production function will exhibit decreasing returns to scale (i.e., doubling inputs less than doubles output).[9] Normally, returns to scale are assumed to be evenly distributed over all inputs (i.e., the production function is homothetic). If the production function is assumed to be both homothetic and Hicks-neutral (i.e., to be holothetic), the impact of technological change becomes indistinguishable from the scale effect and impossible to measure separately. Only under different assumptions of the production function can technological change be separated from scale effects, although not without difficulty.[10] Returns to scale, for example, could disproportionately favor some inputs over others (i.e., the production function could be nonhomothetic), which would subsequently affect the optimal combinations of inputs.

A linearly homogeneous production function has the advantage of holding average and marginal products constant as output increases, because the ratios of the inputs are held constant. In other words, productivity is constant, because no substitution of inputs occurs. Another advantage of linear homogeneity is that, by being able to apply Euler's Theorem, total output can now be distributed over the inputs according to each input's

share of total cost (amount of input times its marginal cost or price). Thus, the assumption of linear homogeneity is empirically appealing and makes production functions easier to work with.

The implications of fixed- and variable-coefficient production functions on productivity of inputs is equally straightforward. In the fixed-coefficient case, productivity of an input remains constant for a given level of output as the other input increases, but the other input's productivity falls sharply. In the variable case, productivity of one input rises as more of the other input is used, again holding output constant. By far, the most commonly used production function for empirical work is the Cobb-Douglas model, which is linear in the logarithms of its variables. As a result, it is assumed to have fixed-coefficients with both constant and unitary elasticity of substitution (i.e., constant returns to scale).

Another aspect related to substitutability is the influence of movements in relative factor prices on productivity. Since inputs depend on their relative prices to find the least-cost combination at a given level of output, changes in factor prices induce movements along the isoquants (i.e., contours of equal output). And, as we have seen, productivity changes occur as factor substitution takes place (unless substitution is restricted, as in the Cobb-Douglas case). The effectiveness of factor price changes depends on the degree of elasticity of substitution between inputs (i.e., if the elasticity of substitution equals one, relative price changes have substantial effects on input productivity; if the elasticity of substitution equals zero, they have no effect).[11]

Finally, when all the possible qualifying adjustments to TFP have been made, the remaining growth in productivity is due to technological change (A). To be sure, technological change would still be a nebulous term that is measured as a residual, but this residual represents the least-biased measure practically attainable. Having isolated technological change, however, there is still more to be learned because there are different types of technological change. For example, technology can enter the production function either as embodied in or disembodied from the inputs. In both embodied and disembodied cases, technological change influences input productivity by shifting the isoquants toward the origin and altering the marginal rates of substitution.

In the case of disembodied technology, technological change is defined as Hicks-neutral, or

II-7 $$Q = A_t F(K, L)$$

if the marginal rate of substitution between inputs is unchanged. Productivity of inputs would increase proportionately under the Hicks-neutral as-

sumption, if the production function is linearly homogeneous. If technology is biased toward favoring capital, technological change is Solow-neutral, or

II-8 $$Q = F(A_t K, L)$$

and the productivity of capital tends to be enhanced relative to other inputs. If technology augments the effectiveness of labor, technological change is Harrod-neutral, or

II-9 $$Q = F(K, A_t L).$$

In the case of embodied technology, an input is simply measured as a function of technological change along with other determining variables. "Embodied" implies that, because of technological changes, an input is not homogeneous over time. Old and new inputs do not participate equally in technological change and, therefore, newer inputs are inherently different from older inputs.[12] For example, capital stock would be a function of investment, depreciation, and technological change over time. Technological change would tend to enhance the productivity of newer capital relative to older capital.

The relationship of the above properties of production functions—technological change, technological bias, elasticity of substitution, relative prices, and scale economies—to productivity changes can be illustrated graphically. In chart 2 below, a labor-saving technological change is assumed to be the only cause for a shift in the production function. Isoquants for a given level of output, Q, are mapped for two periods, before (t) and after $(t+1)$ when the technological change took place. The total change in productivity representing optimal combinations of inputs appears as a shift from point A to B, as adjustments are made to the new production process and to the resulting changes in relative factor prices. The rate of productivity change is the rate at which the isoquant shifts towards the origin. (Note: the axes represent unit capital and unit labor requirements, which is average physical product—or the reciprocal of input productivity, instead of the more conventional levels of capital and labor. As a result, movements within the graph can be related to changes in relative factor productivity).[13]

To isolate the purely technological influences of productivity between two time periods, factor substitution due to relative price changes must first be eliminated. To do this, relative prices in the initial period, t, are retained after the technological change has occurred, $t+1$. In other words, the slope of the isocost line (i.e., contours of equal cost) in period t is preserved. (Alternatively, the analysis could have been done using constant prices at period $t+1$, with the effects being essentially the same but the magnitudes of change being different—the "indexing problem" in measuring productiv-

CHART 2
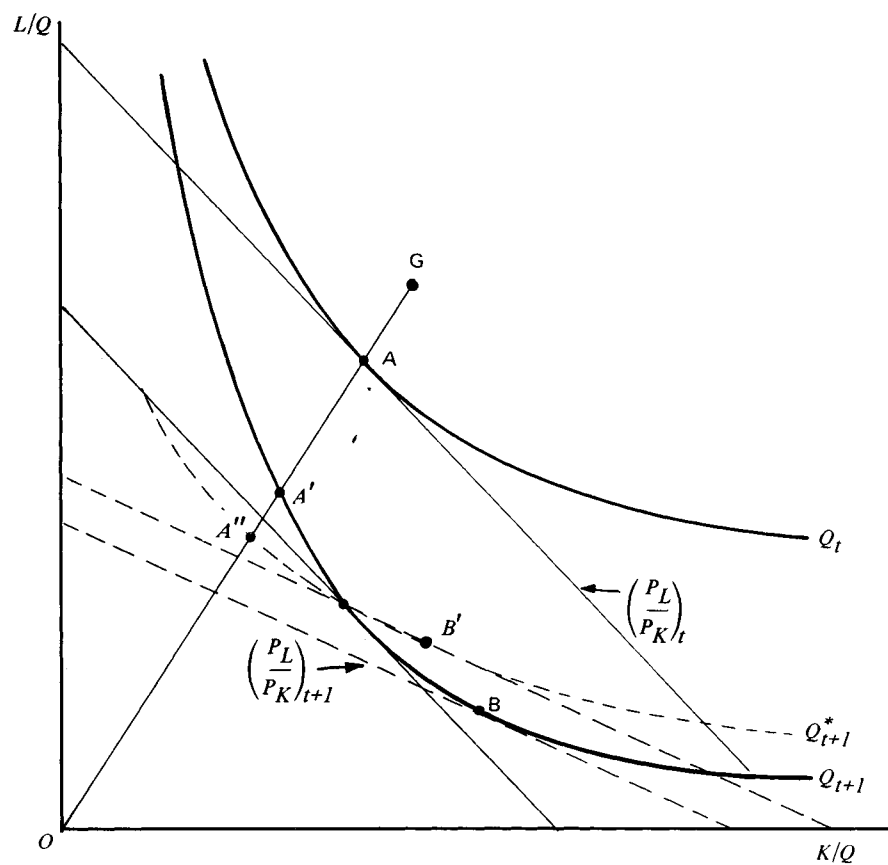
SOURCE: Adapted from Salter, 1969, p. 30.

ity changes). The relevant shift in total factor productivity becomes the move from point A to A'.

But even controlling for relative price changes, the move from A to A' captures some factor substitution, because in this example the nature of the technological change is labor-biased, and relative factor productivity changes. If the technological change had been Hicks-neutral (i.e., constant capital-to-output ratio), the shift in the isoquant would have been from Q_t to Q^*_{t+1} and the relevant shift in optimal combinations of inputs would have been from A to A' '. The extent of the labor bias in TFP growth is represented then by the move from A' ' to A'. The capital-to-labor ratio would, however, remain constant as would the productivity of labor relative to capital.

The final shift in optimal combinations of inputs would be the move from A' to B, which results from the change in factor prices. The substitution of inputs that occurs will also affect relative factor productivity—labor productivity would increase relative to capital. Here, particularly, the elasticity of substitution can be seen to play the dominant role in the extent of the relative productivity change. If the shift of the isoquant had been Hicks-neutral (i.e., less labor biased), the substitution can be described by the movement from A' ' to B' and the increase in labor productivity relative to capital productivity would have been less.

Of course, the firm need not be producing on its cost curve at all, due to mismanagement or some unavoidable reason, such as quasi-fixed factors, and movement toward the cost curve would certainly improve efficiency. For example, suppose a firm is operating at point G on chart 2, because of a fixed stock of capital. The "technical efficiency" of the firm would be measured by the ratio OA/OG. Under the assumption of constant returns, the ratio would be the percentage of potential output that is actually attained.[14]

The scale effects can also be illustrated in chart 2 by making one modification: output is allowed to change. In this case, the input savings due to an increase in the scale of operations is represented by a shift of the isoquants from Q_t to Q_{t+1}, where $Q_{t+1} > Q_t$. Now the shift from A to A' ' represents the input savings, assuming the production function is homothetic (i.e., the benefits of returns to scale are evenly distributed between inputs). The shift from A' ' to A' represents the nonhomotheticity effect of the production function (i.e., its labor-saving bias). If a change in relative prices occurs, a further shift will occur, such as from A' to B, accompanied by adjustments in relative productivity. (Graphic analysis does not allow simultaneously separating technological changes and scale effects).

Capturing all the adjustments that can take place as a result of changes in technology is impossible, because of the complexities of real production functions. Nevertheless, the discussion above provides the foundation to compare the strengths and weakness of alternative approaches to productivity analysis. While the discussion has been a basic review of well-known properties of production functions, the assumptions are often glossed over in most papers. As a result, differences in the assumptions of alternative approaches are obscured and comparisons made difficult. The next three sections focus on the underlying assumptions of three major approaches to TFP as a measure of the productivity performance of a firm.

III) The Network Approach

Since managers are not restricted to the single objective of maximizing physical output relative to any or all physical inputs, they need a framework of analysis that provides links to all aspects of decision making and performance evaluation in the production process. In order to integrate productivity analysis into their decision making, managers must be able to identify what physical and economic relationships are truly constant and where better results (typically in terms of profits) can be obtained.[15] As such, a measure indicating that a change in partial productivity has occurred but offering little insight into the implications of such a change on the overall performance of the firm or what actions must be taken to derive the full benefit from that partial productivity gain is likely to be viewed by businessmen as having little value. The issue is not that managers have no interest in productivity analysis, but rather that their meticulous knowledge of their own production processes make them skeptical of the economist's penchant for simplifying assumptions.

For strategic purposes of comparing a firm's performance with industry standards or among individual divisions of the firm, managers have typically resorted to financial analysis ratios that are more closely related to the accounting concept of profit margins than the economic concept of productivity.[16] By far, the most economically oriented of the structure-of-ratios approaches is the network of productivity, cost, and profitability relationships, which is a blend of financial and physical aspects of resource flows. (See chart 3.) The network approach to evaluating the impact of productivity changes, particularly those directly related to the adoption of new technologies, avoids the manager's concerns with empirically derived production functions, but is still deeply rooted in the economic theory of production.

Unlike economists who may be interested in understanding productivity growth as an end in itself, managers are foremost interested in the profitability of investment, which may not be enhanced by a productivity gain in

CHART 3

Productivity Ratios

$$\frac{Output}{Man\ Hrs.} \qquad \frac{Man\ Hrs.}{Materials\ Vol.} \qquad \frac{Output}{Materials\ Vol.}$$

$$\frac{Man\ Hrs.}{Fixed\ Invest.} \qquad \frac{Materials\ Vols.}{Fixed\ Invest.}$$

$$\frac{Capacity}{Fixed\ Invest.}$$

$$Return\ on\ Capital$$

$$Wage\ Rate \qquad\qquad\qquad\qquad Price\ of\ Materials$$

Cost Ratios

$$\frac{Fixed\ Cost}{Output}$$

$$\frac{Labor\ Cost}{Output} \qquad\qquad\qquad \frac{Material\ Cost}{Output}$$

$$\frac{Fixed\ Cost}{Total\ Cost}$$

$$\frac{Labor\ Cost}{Total\ Cost} \qquad\qquad \frac{Material\ Cost}{Total\ Cost}$$

Managerial Control Ratios

$$\frac{Total\ Cost}{Output}$$

$$\frac{Product\ Vol}{Output} \qquad\qquad\qquad \frac{Output}{Capacity}$$

$$\frac{Profit}{Total\ Invest.}$$

$$\frac{Fixed\ Invest.}{Total\ Invest.} \qquad\qquad\qquad \frac{Capacity}{Fixed\ Invest.}$$

SOURCE: Adapted from Bela Gold, 1979, p. 50.

any one input (i.e., if the adoption of a new technology results in higher material costs because higher input standards are required.) Even plant or division managers, who may not be directly involved in investment decisions, are concerned with cost minimization of those portions of the firm directly under their control, which may have a broader scope than productivity performance. In order to take account of managers' objectives, the network approach simultaneously monitors three distinct sets of ratios: 1) six productivity ratios implicit in the production function, 2) seven cost ratios implicit in the cost function, and 3) six managerial control ratios implicit in the profit function. Although interdependent, each set monitors different aspects of the economic and financial structure of the firm and provides different signals to decision makers.

The set of network relationships for analyzing productivity explicitly are centered around the physical input-output relationships for the firm as a whole or each of its separable units (i.e., plants, divisions within plants, or possibly product lines within a division). To the extent a single product is involved, output is measured in physical units (e.g., tons of finished steel) and each input (e.g., labor) in homogeneous units (e.g., man-hours). Thus, one group of three ratios is simply the partial productivity indexes for the major inputs—output per man-hour, output per materials volume, and capacity per net fixed investment. While the first two ratios are reasonably straightforward, representing labor and materials productivity, the third ratio needs some explanation.

Capacity per fixed investment represents the productivity of investment, which is substituted for capital productivity for two basic reasons. First, aggregated investment is used instead of capital stock, because physical capital is extremely difficult to measure and the data is readily available to managers. The measure of investments, $\bar{I}$, would simply be the sum of past capital expenditures, or $\bar{I}_t = \sum_{i=0}^{\infty} I_{t-i}$. Second, managers are more concerned about the return on their investment than on the size of their capital stock, per se. Thus, the focus on investment productivity avoids a serious criticism that most managers have of production function-based productivity analyses (i.e., the measurement of physical capital stock) and gets directly to the key concern of managers (i.e., profitability). The advantage is limited, however, by the need to include depreciation in measuring investment.

The reason capacity, or potential output, is used in the analysis of investment productivity instead of actual output is to "...differentiate between what the capital goods of the firm embodying such investment can produce and the extent to which they are under-utilized because of market forces."[17] This adjustment for potential output is more appropriate than actual output in capturing the economic concept of capital services derivable from a given stock of capital. To the manager, the relevant concept

of capital productivity is based on potential output, which is likely to provide a lower measure of capital productivity than one based on actual output. Using actual output assumes that capital is always fully utilized, when in fact actual output is typically less than potential output. Indeed, investment decisions are commonly made with the assumption that a portion of the capital stock's time will be idle—either for repairs or from inadequate demand for the product.

The remaining group of physical ratios are input-input ratios: materials volume to man-hours, materials volume to actively utilized fixed investment (or, fixed investment times the utilization rate), and man-hours to actively utilized fixed investment. Again, to be analytically correct, the relevant measure of the capital input is not capital stock, but actively utilized investment (or, fixed investment times the output-to-capacity ratio). In equilibrium, these input ratios should represent the most efficient combinations of inputs, in accordance with traditional productivity theory.

Under assumptions of constant returns to scale and producer equilibrium, for example, a Hicks-neutral change in technology will increase the productivity of the three inputs proportionally and leave the input ratios unchanged. But since an empirical production function is not required for the analysis, these assumptions (which are made simply for the convenience of the econometrician) are not necessary. In the real-world environment of managers, changes in one of these ratios (e.g., the substitution of capital for labor) will induce a series of adjustments over time in all three input-input ratios until a new optimal combination is reached, which will also alter the input-output ratios.

By tracking the movement of all six ratios over time, therefore, the manager is able to analyze the effects of changes in technology on plant productivity. In the case of labor productivity, for example, the ratios are combined in a single equation, such that:

III-1
$$\frac{\text{Physical Output}}{\text{Man-hours}} = \frac{\text{Fixed Investment} \times \dfrac{\text{Output}}{\text{Capacity}}}{\text{Materials Volume}} \times$$

$$\frac{\text{Material Volume}}{\text{Man-Hours}} \times \frac{\text{Capacity}}{\text{Fixed Investment}}$$

This equation can be loosely interpreted as decomposing labor productivity into a capital-to-material ratio, a materials-to-labor ratio, and capital productivity. The advantage of this formulation is that changes in labor productivity, especially in a capital-dominated industry, can be traced to changes in the level of investment in plant and equipment per man-hour and to changes in the productive capacity obtained from that investment.[18]

As long as the source of the technological change (e.g., displacement of labor by new machines) is known, therefore, the changes in "apparent" productivity of labor (associated with perhaps an increase in the capital to labor ratio or with an increase in capital productivity) can be distinguished from gains in labor productivity caused by other sources, such as intensified labor effort.

The productivity network is constructed around physical measures of inputs and outputs and must be linked through input prices to a cost function, in order to measure the economic effects of changes in productivity on costs and ultimately profitability. In chart 3, composite prices of labor, capital, and materials are represented by wage rates, return on investment, and material prices respectively. The product of the physical input and its price equals total input costs. The cost function is built up around seven inter-related ratios. A group of four ratios measure unit costs: total costs per output, labor costs per output, fixed costs per output, and materials costs per output. An analysis of the effect of a change in labor productivity on unit labor costs, for example, may have to consider possible changes in wage rates. Again, assuming competitive markets, managers will seek to equate the marginal product of labor with the price of labor, which is set by the market and presumably insensitive to factor-demand shifts induced by a technological change in any firm's production process. The assumption may break down for firms in which collective bargaining centers around the issue of labor productivity or in which an increase in labor productivity requires higher quality labor.

Another group of cost ratios measure factor shares of total costs: labor costs per total costs, fixed costs per total costs, and materials costs per total costs. Factor shares are essentially the products of the ratios of physical inputs to outputs and the ratios of input and output prices. By analyzing the three share ratios together, a change in one input price can be traced through its effects on relative prices and back to substitution effects and productivity changes in the network ratios. Aside from input price changes, the effect of a change in input productivity must also consider what proportion that input's costs are of total costs, especially if an opposite change in other unit input costs with higher proportions of total cost is induced by the initial change. These relationships can be expressed algebraically as:

III-2 $\quad \Delta \dfrac{\text{Total Cost}}{\text{Output}} = \Delta \dfrac{\text{Wage Cost}}{\text{Output}} \left(\dfrac{\text{Wage Cost}}{\text{Total Cost}} \right) +$

$\Delta \dfrac{\text{Material Cost}}{\text{Output}} \left(\dfrac{\text{Material Cost}}{\text{Total Cost}} \right) + \Delta \dfrac{\text{Other Cost}}{\text{Output}} \left(\dfrac{\text{Other Costs}}{\text{Total Cost}} \right)$

For example, a 10 percent reduction in labor costs would offset a 2 percent increase in material costs for a firm in which labor accounts for only 10 percent and material costs account for 50 percent of total costs.

The role of the cost ratios is "...to emphasize the need to consider possible *interactions* between adjustments in factor productivities and in the factor prices, instead of encouraging continued reliance on the common but unrealistic assumption that factor prices remain unchanged."[19] It should be noted that, although no direct effort is made to weight the contributions of each input to productivity changes as occurs in neoclassical TFP measures, factor cost proportions are explicitly included in the analysis (i.e., the percentage increase in labor productivity can be compared to accompanying changes in wage rates). In fact, under the assumptions of constant returns to scale and producer equilibrium, the input share would be equal to the elasticity of cost with respect to input price, which is used explicitly in the neoclassical approach to measure the share-weighted contribution of an input to total costs. Moreover, not only are input prices likely to change over time in response to a change in input demand, but the resulting relative price change would again alter the optimal combination of physical inputs, which induces another round of input adjustments, and so on until a new equilibrium is achieved.

To complete the network approach to TFP analysis, a profit function is introduced to provide a group of ratios through which managers "control" the determinants of their firms' profitability. The relevant set of ratios for the profit function are: total cost per output (average unit costs), product value per output (product prices), fixed investment per total investment (internal allocation of capital between fixed and working components), output per capacity (capacity-utilization rate), profit per total investment (profit rate), and—to come full circle back to the production function—capacity per fixed investment (productivity of fixed investment). These ratios place the role of productivity gains in the broader context of economic factors that influence the firm's profitability. A cost-reducing investment may actually prove undesirable if it leads, for example, to reductions in product prices by market forces or to increases in idle capital that more than offset profit gains from the initial cost reduction. Algebraically, the profitability ratios are related in the following way:[20]

III-3 $$\frac{\text{Profit}}{\text{Total Investment}} = \frac{\text{Profit}}{\text{Output}} \, x \, \frac{\text{Output}}{\text{Total Investment}}$$

$$\frac{\text{Profit}}{\text{Output}} = \frac{\text{Product Value}}{\text{Output}} - \frac{\text{Total Cost}}{\text{Output}}$$
$$\text{(average price)} \quad \text{(average unit cost)}$$

$$\frac{\text{Output}}{\text{Total Investment}} = \frac{\text{Output}}{\text{Capacity}} \; x \; \frac{\text{Capacity}}{\text{Fixed Investment}} \; x \; \frac{\text{Fixed Investment}}{\text{Total Investment}}$$

| (utilization rate) | (productivity of internal investment) | (internal allocation of capital) |

The interactions of the inputs are the basis for three serious shortcomings that managers (and economists) find in partial productivity measures: 1) partial productivity obviously does not measure efficiency of the production process as a whole or even the productive contribution of the single input, 2) increases in an input's productivity may or may not be desirable, and 3) even if increases in an input's productivity are accompanied by only proportionate increases in unit input costs, total production costs are more likely to increase than to remain unchanged if another input is more dominant.

Consider, for example, the effect of a technological improvement that increased the capacity of existing capital stock through relatively minor outlays in ancillary equipment (e.g., computers). Labor productivity would obviously increase, but the means by which that productivity gain is accomplished can be seen through the link between labor and capital productivity in the network relationships. Changes in labor productivity result from some combination of a change in adjusted capital-labor ratio and a change in capital productivity itself, which should be observed in the values of the ratios. The increase in productive capacity also induces a change in capital charges (price of capital), but an increased utilization (assuming the improvement was dictated by excess demand) would reduce capital charges per unit of output (average capital costs). Moreover, wage costs per unit of output may be reduced, which may or may not be passed along in the form of higher wage rates. If wage rates do rise, labor may share in the gains from the technological improvement even though all of the productivity gains in this case were derived from capital. (To be sure, the division of labor and capital productivity is never as clear as is suggested here. If capital is more productive, then virtually by default labor is also more productive and is likely to benefit from the technological gain in the form of higher wages whether earned or not).

If capacity is expanded by replacing old, depreciated capital with new, undepreciated capital (i.e., modernization of a plant), the profitability of the plant may actually decline. Consider, for example, that the increase in capital costs was not offset by large enough increases in capital productivity. Labor productivity would still have increased and induced an increase in wage rates. The net result would be an increase in total average costs, or a reduction in profit margins. Moreover, the new technology could re-

quire an improvement in the quality of the materials input, which would further increase average cost.[21]

By analyzing patterns of change in the ratios over time, managers clearly have a solid method for analyzing the impact of productivity gains on the profitability of their firm, assuming the firm's production process is stable so that changes in ratios can be attributed to the introduction of a new technology. The new technology would create a disturbance in the pattern of the ratios, which could then be analyzed. If the production process is unstable, it may be unclear what a change in the ratios signifies. The approach also burdens the manager with a substantial amount of data that must be monitored and a large number of ratios that must jointly be analyzed. The question is whether there is better way to organize the information contained in the ratios into a more comprehensible measure of a firm's productivity performance.

IV) The Neoclassical Approach

The basic challenge to the neoclassical approach is to match the scope of the network ratios, but at the same time to combine the ratios into a single unified measure that isolates the real gains in efficiency of production, or growth in TFP. Given a willingness to accept the assumption of an estimable production function, the separation of movement along the production function (factor substitution) from the actual shift in the production function (change in TFP) becomes a relatively simple task. As illustrated above, the shift is measured by a residual—the difference between output and input growth or any source of output growth other than changes in inputs. The trick is to remove as many sources of bias in the empirical work, so that the residual can reasonably be assumed to be proportional to real gains in efficiency due to technological change (A), or that portion of TFP that is directly the result of the advance of technology.

Given the complexities of production functions described in section II, a more versatile model than a simple Cobb-Douglas production function is obviously required to demonstrate the full analytical power of the neoclassical approach. However, to show the development of the neoclassical approach as a microeconomic tool, it is appropriate to start with the simplest and most intuitively appealing models that have been used for analysis at the macroeconomic level and then to build in complexity by dropping assumptions. In order to emphasize the regional application of the models, all the following equations are presumed to apply to a regional firm (i.e., subscripts to designate region are omitted) unless otherwise specified. The same analysis can be applied to a regional industry essentially by aggregating the production functions of regional firms in the same industry.

In the previous section, TFP was presented as a ratio of growth in aggregated real output to aggregated real input. The first issue to confront is how to handle multi-product firms. Under the appropriate assumptions of competitive equilibrium, the sum of the output values (i.e., total revenues from a set of final products of a firm) should equal the sum of the input values (i.e., total costs of all factors of production), or:

IV-1
$$\sum_j P_j Q_j = \sum_i P_i X_i$$

where

P_j = price of the jth output
P_i = price of the ith input
Q_j = quantity of the jth output
X_i = quantity of the ith input

Differentiating with respect to time and dividing both sides by total value will result in the following relationship:

IV-2
$$\sum_j u_j [\dot{P}_j + \dot{Q}_j] = \sum_i v_i [\dot{P}_i + \dot{X}_i]$$

where the corresponding weights are simply factor shares, or:

$$u_j = \frac{P_j Q_j}{\sum_j P_j Q_j} \, , \, v_i = \frac{P_i X_i}{\sum_i P_i X_i} \, ; \, \sum_j u_j = \sum_i v_i = 1 \, ; \, u_j \, , v_i \geq 0$$

If we denote $\dot{Q} = \sum_j u_j \dot{Q}_j$, $\dot{X} = \sum_i v_i \dot{X}_i$, $\dot{P}_q = \sum_j u_j \dot{P}_j$, and $\dot{P}_x = \sum_i v_i \dot{P}_i$, the above equation is simplified to the form:

IV-3
$$\dot{P}_q + \dot{Q} = \dot{P}_x + \dot{X}$$

For the moment, let us assume that technological change (A) is the only source of productivity growth. By substitution then, the rate of growth in TFP can now be expressed either in its primary form (i.e., the production function):

IV-4
$$\dot{A} = \dot{Q} - \dot{X}$$

or in its dual form (i.e., the cost function):

IV-5
$$-\dot{B} = \dot{P}_x - \dot{P}_q$$

In this simple formulation, A is the outward shift in the production function due to techological change, B is the inward shift in the cost function due to technological change, and $TFP = A = -B$.

The main point is that even if all outputs and inputs are correctly measured and the direct aggregation of their values must be equal, the growth rates of physical outputs can still exceed the growth rate of physical inputs. The difference of these growth rates will be TFP growth and will be reflected in the differential growth rates of inputs and outputs and their prices. In other words, output prices need not grow as fast as production costs, if costs are offset by productivity gains.[22]

The effect of quality changes in the output or in any of the inputs can also distort the desired measure of productivity. Other things being equal, for example, an improvement in quality will be reflected in an increase in price. But, prices are also used to derive physical quantities (i.e., real value) from nominal money values of inputs and outputs. Does a change in price reflect a change in quality and, therefore, an increase in physical quantities? Or, does it reflect inflation or even a change in relative price, but not a change in quantity? Fortunately, the problem can be dealt with directly in the above analytical framework. The output measure, as in the case of inputs, is typically an aggregate index of products created by the industry or firm (in the case of a multi-product firm). The preferred method of aggregating

is to use a share-weighting, or Divisia, index: $\dot{Q} = \sum_j \dfrac{P_j Q_j}{P_q Q}(\dot{Q_j})$.

The difference between the actual total (i.e., direct aggregation) and the share-weighted total (i.e., Divisia aggregation) measures quality changes of a particular input or output. The rate of change in the direct aggregation of the index should contain both the quality improvements and the quality-adjusted gains in output. The quality-adjusted gains are calculated by first disaggregating the output index into product lines and computing individual growth rates for each line. Then, the separate growth rates can be combined into a share-weighted growth index, based on each product line's contribution to total revenue, in the case of output, and total cost, in the case of inputs. Any difference between the aggregated growth rate and the share-weighted growth rate—the residual—can be attributed to quality enhancement of the input. A parallel approach on the cost function will distinguish between quality-related adjustments in input prices and technology-related adjustments in output prices.

An approach developed by Gollop and Jorgenson provides the foundation on which more complex models of production functions at the microeconomic level can be discussed.[23] The method applied by Gollop and

Jorgenson is based on an expanded production function for an industry, which in this discussion is applied to a regional firm, and still assumes constant returns to scale and Hicks-neutral technological change:

IV-6 $$Q = F(K,L,M,A)$$

where A is expressed explicitly as a function of time and Q is output expressed as value of total shipments rather than value added. Compared to the form presented earlier, the only additional variable is materials, M, and the production is presented in logarithmic form. In order to solve for TFP growth, output is differentiated with respect to time (i.e., the proxy for technology). Since $\frac{d\ln Q}{dt}$ is equivalent to $\dot{Q}$, the differentiated equation can be expressed as:

IV-7 $$\dot{Q} = \frac{\partial \ln Q}{\partial \ln K} \dot{K} + \frac{\partial \ln Q}{\partial \ln L} \dot{L} + \frac{\partial \ln Q}{\partial \ln M} \dot{M} + \frac{\partial \ln Q}{\partial t}$$

where $\frac{\partial \ln Q}{\partial t} = \dot{A}$, or the desired measure of productivity growth.

Assuming the necessary conditions for producer equilibrium (i.e., perfectly competitive factor markets, so that marginal product equals the input price), the partial derivatives of output with respect to inputs in log form (e.g., $\frac{\partial \ln Q}{\partial \ln M}$, which is also an elasticity) must equal their respective shares of the total value of output, or:

$$\frac{wL}{P_q Q} = \frac{\partial \ln Q}{\partial \ln L} , \frac{rK}{P_q Q} = \frac{\partial \ln Q}{\partial \ln K} , \text{ and } \frac{mM}{P_q Q} = \frac{\partial \ln Q}{\partial \ln M}$$

or,

$$v_L = \frac{wL}{P_q Q} , v_K = \frac{rK}{P_q Q} , \text{ and } v_M = \frac{mM}{P_q Q}$$

where w = price of labor, r = price of capital, m = price of materials, and $\sum_i v_i = 1$ (with $i = L, K,$ and M), because of the constant returns to scale assumption. (P_q is traditionally considered to be a numeraire, which is assumed to be 1 without loss of generality). The differential equation can then be simplified to an expression of a growth rate of output less the weighted average growth rates of inputs:

IV-8 $$\dot{Q} = [v_K \dot{K} + v_L \dot{L} + v_M \dot{M}] + \dot{A}$$

The neoclassical approach at this point is essentially identical to the productivity ratios in the network approach. All three partial productivity measures are implicit in the formula and can be presented explicitly by

subtracting the growth rate of the chosen input (i.e., labor) from both sides of the TFP equation:

IV-9 $$\dot{Q} - \dot{L} = v_K(\dot{K} - \dot{L}) + v_M(\dot{M} - \dot{L}) + \dot{A}$$

where factor intensities are represented by $v_K(\dot{K} - \dot{L})$ and $v_M(\dot{M} - \dot{L})$ and remembering that $v_K + v_L + v_M = 1$. This rearrangement of the TFP equation shows the sources of labor productivity growth depending on factor intensities and changes in TFP. The partial productivity of an input already adjusts for input substitution, which under the network approach required analyzing three additional sets of ratios—the input-input ratios.

Moreover, the number of inputs can easily be expanded to a larger set of inputs, which would only further encumber the network approach. An obvious choice for a fourth input is energy, given its topical interest, but recent studies have also differentiated between domestically produced intermediate inputs and imported inputs. Other studies have included production and non-production workers, instead of aggregating into a single labor input. Capital could be divided into separate parts, such as machinery, structures, and land.[24] Indeed, the only limitations to further disaggregating the four basic inputs are data on the inputs and time to process them. (As we shall see in the next section, an input-output table potentially overcomes both limitations.)

The inclusion of multi-product firms also raised the problem of a changing composition of total output of the firm over time. The firm may be affected by a changing composition of aggregate demand (i.e., sectoral shift) or simply by a changing market share. If so, some portion of the firm's TFP growth should be attributed to the reallocation of its resources from low productivity product lines to high productivity product lines and only the remainder to technological improvements. In some sense, multiple products introduce an aggregation problem similar to quality changes discussed above. That is, each product line will have its own technological improvements and the contribution each output makes to total productivity growth will depend on its share of total output of the firm. At the same time, there will be a reallocation of factors among products.

The basic equation to analyze the compositional-bias effects is derived by aggregating the productivity measures for each product, using the following equation:

IV-10 $$\dot{A} = \sum_j \frac{P_jQ_j}{P_vV}\dot{A}_j + \sum_j \frac{P_v - P_{vj}}{P_vV}\frac{dV_j}{dt} + \sum_j \frac{r_j - r}{P_vV}\frac{dK_j}{dt} +$$

$$\sum_j \frac{w_j - w}{P_v V} \frac{dL}{dt}$$

where $\dot{A}$ = aggregate rate of productivity growth for the firm
$\dot{A}_j$ = rate of productivity growth for product-line j
V = total value added of the firm's output
V_j = value added of product-line j
Q_j = output of product-line j
P_v = price of total value added
P_{vj} = price of value added of product-line j
P_j = price of product-line j's output (i.e., specific product price)
r = total price of capital
r_j = price of capital associated with product-line j
w = total price of labor
w_j = price of labor associated with product-line j

The equation decomposes $\dot{A}$ into the following four components: a

weighted sum of product-line productivity growth rates ($\sum_j \frac{P_j Q_j}{P_v V} \dot{A}_j$),

the effect of reallocation of capital and labor ($\sum_j \frac{r_j - r}{P_v V} \frac{dK_i}{dt}$ and
$\sum_j \frac{w_j - w}{P_v V} \frac{dL}{dt}$, respectively), and the effect of shifts in total value added

of the firm's products ($\sum_j \frac{P_v - P_{vj}}{P_v V} \frac{dv_j}{dt}$) .[25] For example, capital could be

moved from one product line where it earns a low rate of return ($r_j < r$) to another product line where it earns a relatively high rate of return ($r_j > r$). Although the aggregated capital input need not change, the shift to a more productive product would increase total output. Similarly, holding the allocation of capital and labor constant, the firm's total productivity could be increased by a shift in the composition of value added from product-lines with high marginal cost ($P_{vj} > P_v$) to product-lines with low marginal cost ($P_{vj} < P_v$).

If a set of fixed weights, selected for some base period, are used to compute what each product-line's contribution would have been if its share of total output remained constant, one can identify the pure contribution of changing product-line productivity within each product-line and also the effects of shifts in demand activity among products. Adding in the base

period weights to hold the distribution of products constant gives the following equation:

$$\text{IV-11} \quad \dot{A} = \sum_j \left(\frac{P_jQ_j}{P_vV}\right)_B \dot{A}_j + \sum_j \left[\frac{P_j - Q_j}{P_vV} - \left(\frac{P_jQ_j}{P_vV}\right)_B \right]\dot{A}_j +$$

$$\sum_j \frac{P_v - P_{vj}}{P_vV}\frac{dV_j}{dt} + \sum_j \frac{r_j - r}{P_vV}\frac{dK_j}{dt} + \sum_j \frac{W_j - w}{P_vV}\frac{dL_j}{dt}$$

where $\sum_j \left(\frac{P_jQ_j}{P_vV}\right)_B \dot{A}_j$ identifies the pure contribution of changing productivity within product lines and $\sum_j \left[\frac{P_iQ_j}{P_vV} - \left(\frac{P_iQ_j}{P_vV}\right)_B \right]\dot{A}_j$ identifies the effect of shifts in activity among product lines.

While working with the primary form of the production function provides indepth analysis of productivity growth, the dual form has been useful in identifying other sources that bias the interpretation of the rate of technological change contained in the productivity measure. Among the most important developments in productivity analysis using cost functions has been the measurement of scale economies and factor-augmenting technology bias. Using Sheppard's lemma, the production function can be converted to a cost function from which a productivity measure can be derived. Starting with the simplest input-output relationship:

$$\text{IV-12} \qquad P_qQ = rK + wL + mM, \text{ or } P_q = r\frac{K}{Q} + w\frac{L}{Q} + \frac{m}{Q}$$

where total revenue equals total cost and output price (P_q) equals average cost. The average cost function of the firm can be expressed in terms of input prices as:

$$\text{IV-13} \qquad\qquad P_q = G(r,w,m,T)$$

Again, taking the total derivative to isolate the productivity measure:

$$\text{IV-14} \qquad\qquad \dot{P}_q = v_K\dot{r} + v_L\dot{w} + v_M\dot{m} + \dot{B}$$

and $T\dot{F}P$ is measured by $\dot{B}$ (where $\dot{B} = \dfrac{\partial \ln P_q}{\partial t}$) is perhaps more appro-

priately called the rate of cost reduction. Now, all the basic cost ratios in the network approach are revealed and changes in relative prices controlled for in the analysis of productivity growth.

In the form of a total cost function, returns to scale can be measured or explicitly controlled by making cost a direct function of output. The cost function takes the form:

IV-15 $$C = G(r,w,m,Q,T)$$

where now average cost is a function of output as well. Differentiating in log form with respect to time and applying Sheppard's lemma (e.g., $\frac{\partial C}{\partial r} = K$) decomposes the cost function into its sources of growth, or

IV-16 $$\dot{C} = v_K \dot{r} + v_L \dot{w} + v_M \dot{m} + \varepsilon \dot{Q} + \dot{B}$$

The derivative of cost with respect to output, ε (where $\varepsilon = \frac{\partial \ln C}{\partial \ln Q}$). can be interpreted as a cost elasticity representing an index of scale economies and the assumption of constant returns to scale can be dropped. The value, ε, measures the change in cost that is independent of technological change and changes in input prices. If the resulting value is equal to one, then the firm exhibits constant returns to scale and and does not bias the productivity growth measure if it is excluded. If the value is greater (less) than one, the contribution of economies (diseconomies) of scale to the residual is eliminated.[26]

If the firm is viewed as having a multi-product production and cost function, there is the potential for scale economies to be increasing for some product lines and decreasing for others. In order to analyze the effects of multi-products on productivity, assume a cost function of the form:

IV-17 $$C = G(r,w,m,Q_1,Q_2,....Q_n,T)$$

where the producer is minimizing the cost of producing m outputs with the three basic inputs: capital, labor, and materials. Again, totally differentiating the cost function in log form with respect to time:

IV-18 $$\frac{d\ln C}{dt} = \frac{\partial \ln C}{\partial \ln w} \frac{d \ln w}{dt} + \frac{\partial \ln C}{\partial \ln r} \frac{d \ln r}{dt} + \frac{\partial \ln C}{\partial \ln m} \frac{d \ln m}{dt} +$$

$$\frac{\partial \ln C}{\partial \ln Q_1} \frac{d \ln Q_1}{dt} + \frac{\partial \ln C}{\partial \ln Q_2} \frac{d \ln Q_2}{dt} + ... \frac{\partial \ln C}{\partial \ln Qn} \frac{d \ln Qn}{dt} + \frac{\partial \ln C}{\partial t}$$

Now, setting $\frac{\partial \ln C}{\partial \ln w} = L, \frac{\partial \ln C}{\partial \ln r} = K,$ and $\frac{\partial \ln C}{\partial \ln m} = M$ (from Sheppard's lemma) will result in:

IV-19 $\dot{C} = v_K \dot{r} + v_L \dot{w} + v_M \dot{m} + \dfrac{\partial \ln C}{\partial \ln W_1} \dot{Q}_1 + \dfrac{\partial \ln C}{\partial \ln Q_2} \dot{Q}_2 \cdots +$

$$\dfrac{\partial \ln C}{\partial \ln Q_N} \dot{Q}_n + \dfrac{\partial \ln C}{\partial t}$$

If $\dfrac{\partial \ln C}{\partial \ln t}$ is defined as proportional shift in the cost function, $\dot{B}$, and ε_j
(where $\varepsilon_j = \dfrac{\partial \ln C}{\partial \ln Q_j}$ is the elasticity of cost with respect to each product
line, then the the equation can be rearranged, as follows:

IV-20 $\qquad \dot{B} = \dot{C} - [v_K \dot{r} + v_L \dot{w} + v_M \dot{m}] - \sum_j \varepsilon_j \dot{Q}_j$

The equation thus divides the shift in the cost function into 1) a change in
costs, 2) change in aggregate inputs, and 3) aggregate scale effects. By total
differentiation of C, one can derive $\sum_i v_i \dot{P}_i = \dot{C} - \sum_i v_i \dot{X}_i$ Substituting this
equation into IV-20 and denoting $\sum_i v_i \dot{X}_i = \dot{X}$. one can then derive
$\dot{B} = \dot{X} - \sum_j \varepsilon_j \dot{Q}_j$. Adding $\dot{Q}$ to both sides and recalling that $T\dot{F}P = \dot{Q} - \dot{X}$,
one can relate $\dot{B}$ to the measure of TFP:

IV-21 $\qquad T\dot{F}P = -\dot{B} + (1 - \sum_j \varepsilon_j)\dot{Q}_j$

Under conditions of constant returns to scale, $T\dot{F}P$ will equal $-\dot{B}$. Ear-
lier, $\dot{A}$ was equated with $-\dot{B}$. Since the measure of $\dot{B}$ is affected by scale
economies, one would naturally expect $\dot{A}$ also to be affected. Indeed, the
equation $T\dot{F}P = \dot{A} = -\dot{B}$ holds only under conditions of constant returns
to scale ($\varepsilon = 1$). More generally, the relation between the two measures is
$-\dot{B} = \varepsilon \dot{A}$, (where again ε is the elasticity of cost to output, or the index for
scale economies). In other words, shifts in the cost function are not equal
to shifts in the production function except under the assumption of constant
returns to scale.[27] Thus, another way of viewing the sources of technolog-
ical change for the firm, $\dot{A}$, is to distinguish between a static component
(scale economies), which is a shift in the production function due essentially
to proportional changes in inputs, and a dynamic component (cost-reducing
new technology), which is the shift in the production function. It is this
dynamic component that is the Holy Grail of productivity analysis. If
$\dot{B} > 1$, the term can be referred to as technological progress; if $\dot{B} < 1$, then
the appropriate term would be technological retardation.

The cost function can also facilitate identifying which inputs are primarily
augmented by technological change, but introduces the need for empirical

estimation. In the case of a Cobb-Douglas production function, for example, the factor coefficients were held constant on the (required) assumption of Hicks-neutral technological change. If technology is factor specific, however, changes in that factor will be a function of technological change, independent of the other factors. As such, fixed coefficients no longer apply and the cost function can be rewritten:

IV-22
$$C = H[\frac{w}{A_L}, \frac{r}{A_K}, \frac{m}{A_M}, Q]$$

where each A_i represents productivity change in the ith factor in time, independent of other input factors.[28] Each augmentation function, A_i, can be expressed as a translog function of time:

IV-23
$$A_i = \exp[\eta_i T + \frac{1}{2}\phi_i T^2] \quad (i = L, K, M)$$

which is now a second-order approximation. When combined in a translog cost function, the coefficients can be estimated with time series data. The translog approximation of the cost function takes the form:

IV-24
$$C = \exp[Z - \sum_i \beta_i \eta_T T - \frac{1}{2}\sum_i\sum_j \gamma_{ij}\eta_i \ln P_j T - \sum_i \gamma_{iQ}\eta_i \ln QT$$

$$- \frac{1}{2}\sum_i \beta_i\phi_i T^2 + \frac{1}{2}\sum_j(\sum_i \gamma_{ij}\eta_i)\eta_j T^2]$$

where $Z = \alpha_0 + \sum_i \beta_i \ln P_i + \beta_Q \ln Q + \frac{1}{2}\sum_i\sum_j \gamma_{ij} \ln P_i \ln P_j$

$$+ \sum_i \gamma_{iQ} \ln P_i \ln Q + \frac{1}{2}\gamma_{QQ}(\ln Q)^2$$

and $-\frac{1}{2}\sum_i \beta_i\phi_i + \frac{1}{2}\sum_j(\sum_i \gamma_{ij}\eta_i)\eta_j$ is the component that allows the augmentation factor to be flexible.

For measurement purposes, A_i is restricted to a first order form, such that $A_i = \exp[\eta_i T]$ parameters containing T will change. For example, the parameter for T^2 becomes $\frac{1}{2}\sum_j(\sum_i \gamma_{ij}\eta_i)\eta_j$. The rate of cost reduction due to technological change (i.e., the partial derivative of cost with respect to time) is:

IV-25
$$-\dot{A} = \sum_i \beta_i \eta_i - \sum_i (\sum_j \gamma_{ij}\eta_j) \ln P_i - (\sum_i \gamma_{iQ}\eta_i) \ln Q$$
$$+ \sum_i (\sum_j \gamma_{ij}\eta_i)\eta_j T$$

Since η_i enters each component of technological change, $\dot{A}_i$ (in equation IV-25) can be decomposed into the contributions of each input by taking the derivative with respect to each factor's price. For example, labor's contribution is

IV-26
$$-\dot{A}_L = -\eta_L[\beta_L + \sum_i \gamma_{iL} \ln P_i + \gamma_{LQ}\ln Q - (\sum_i \gamma_{iL}\eta_i)T].$$

$$(i = L, K, M)$$

The sum of these expressions is equal to $-\dot{A}$. If η_i equals zero, the corresponding augmentation coefficient will equal unity and the conclusion would be that the input did not contribute to technological change. However, if η_i is greater than zero, then that input is a positive source of the firm's technological growth (and vice versa). If $\eta_K = \eta_L = \eta_M$, each input contributed equally to TFP growth (i.e., Hicks-neutral growth). And, finally, if $\eta_K > \eta_L$, then the firm experienced capital-augmented growth.

Two final assumptions warrant a brief mention. The first is the assumption that price equals marginal cost, or that factor markets are in competitive equilibrium. The literature abounds with examples of imperfect competition and, in the case of regulated industries, the assumption is also unlikely to apply. Denny, Fuss, and Waverman have developed an approach to analyzing productivity growth under conditions of non-marginal cost pricing and much research is underway in the use of shadow pricing among firms where direct prices do not exist.[29] A second, related assumption is that firms operate where long-run and short-run costs are minimized, which is linked to the issue of idle capacity. Two basic approaches have been developed to deal with the problem. Morrison and Brendt have developed a dynamic equilibrium model that measures the difference between short-run and long-run costs, which is attributed to excess capacity.[30] An alternative approach incorporates the stock of fixed capital directly into the model and estimates a restricted cost function.[31] Both approaches are in the early stages of refinement and are somewhat controversial.

The decomposition of TFP into scale and technological change illustrates the complexity of disentangling the effects of technological change on productivity growth even if other sources of growth have been eliminated. The term remains, in effect, a residual and is still subject to distortions that can

not be quantified. To the extent that A is never a pure measure of technological change, managers must be careful in analysis changes over time. Clearly, more research is need to improve the measurement of the sources of productivity change, especially in the handling of capacity and shadow pricing. Nevertheless, research in the last few years has dramatically improved our ability to analyze productivity growth, so that managers can place greater confidence in the critical assumption that the residual is at least proportional to the growth in productivity due to technological change.

V) The Input-Output Approach

The primary strength of input-output (I-O) models is widely known to be their ability to analyze resource-allocation problems. Yet, to the extent that productivity analysis evaluates the effect of changes in various input-output and input-input relationships on the production process, an I-O approach would seem to be ideally suited for productivity analysis. Indeed, the potential for deriving an I-O-based productivity measure has existed for some time. However, only recently have developments linking an I-O framework to the neoclassical approach sparked new interest in the application of I-O models to productivity analysis. This new interest in I-O analysis as an extension of the neoclassical approach is centered around the ability of I-O models to make three contributions to productivity analysis: 1) extend the number of intermediate inputs that can be explicitly captured in the analysis, 2) adjust for compositional shifts among inputs and final demand at a very detailed level, and 3) measure both direct and indirect components of productivity changes.

Starting with Leontieff's original approach, a general production function for a firm, based on the firm's own I-O matrix for the products it produces, is given such that:

V-1
$$Q_j = F(X_{1j}, X_{2j}, ... X_{Nj})$$

where Q_j = (column) vector of physical outputs from each of the firm's product lines
   X_{kj} = the amount of kth input absorbed by product-line j.

(This production function includes only material inputs, but can be expanded to include labor and capital in the form of value added, as will be illustrated below). From the production function, the input-output ratios—or, coefficients of production—are derived simply as:

$$a_{1j} = \frac{X_{1j}}{Q_j}, \ a_{2j} = \frac{X_{2j}}{Q_j}, \ ... \ a_{Nj} = \frac{X_{Nj}}{Q_j}$$

with a = matrix of the direct coefficients of
production for all inputs and outputs of the firm.

As formulated above, the a_{ij}s are measured in physical terms, i.e., tons of coal per ton of steel shipped. But, for most empirical work, value coefficients are required and the physical coefficients will be redefined later in the text (eq. IV-6) as:

$$a_{ij} = \frac{P_i X_{ij}}{P_j Q_j}$$

where P_i and P_j are prices of inputs and outputs respectively. Now, $P_i X_{ij}$ can be interpreted as the transaction flow of input i to the output of product-line j and a_{ij} is the production coefficient in value terms.

The underlying assumptions for the I-O model are similar to those found in neoclassical production functions. For example, the production function is assumed to be linearly homogeneous (i.e., constant returns to scale) and is unique for each output (i.e., no joint products). Both assumptions are subject to the same criticisms as appeared in the neoclassical approach. The linearly homogeneous assumption can be dropped either by introducing capital and labor constraints into the static model or by using a dynamic model with investment as an endogenous vector that determines product-line capacity levels.[32] The homogeneous output assumption is essentially the aggregation problem, which becomes less important as product-lines are defined at increasing levels of disaggregation.

Another key assumption that is required by the very nature of an I-O model is fixed-input coefficients (i.e., no input substitution). This restriction is typically defended out of practical necessity rather than theoretical convenience.[33] The restriction of fixed coefficients, however, only applies to analyses using the same matrix over time for forecasting purposes. If new matrices are used for each time period, the coefficients are obviously no longer fixed. With input substitution and compositional shifts at least conceptually adjusted for, all that remains is to capture movements in the production function through changes in the I-O coefficients over time.[34]

Leontieff's original approach to deriving a productivity index from an I-O model was to calculate what he called an index of structural change. This index was intended to indicate the change in output that the firm is capable of producing, given the same amount of inputs. The index can be formulated explicitly by taking the average change in the direct production coefficients as follows:

V-2 $$\bar{a}_{ij} = (a_{ij}^1 - a_{ij}^0)/[(a_{ij}^1 + a_{ij}^0)/2]$$

where $0,1$ = base and end years, respectively. If $\bar{a}_{ij}$ is based on a change in a material input coefficient over time, the index can be interpreted as a measure of structural change in the sense that there has been a compositional change in the share of the input consumed in the production of the output in product-line j. Since $\bar{a}_{ij}$ is based on a change in the input coefficient, the index can be loosely described as a measure of the change in input productivity. In fact, $\bar{a}_{ij}$ is nothing more than the reciprocal of Q_{ij}/X_{ij}, which is the typical productivity ratio. If the coefficient for production of the input declined between two periods, the productivity of that input increased at the inverse rate. (Although Leontieff did not originally include primary inputs in his matrix, labor or capital productivity would be analyzed in the same way as any other input once they are added to the matrix.)

Since $\bar{a}_{ij}$ is an average-growth measure of the direct coefficients of an individual input in the production of output j, the total change over all the inputs of the firm should be measured by a weighted-average value of the sum of the changes in all inputs used in the production of output j over the two periods. The appropriate weights would be $(X_{ij}^1 + X_{ij}^0)/2$, or the average share of each material input in the total amount of materials used over the two periods. Therefore, for product-line j the direct productivity index of all materials is:

V-3
$$I_j = \frac{\sum_i [(P_i X_{ij}^1 + P_i X_{ij}^0)/2]\, \bar{a}_{ij}}{\sum_i (P_i X_{ij}^1 + P_i X_{ij}^0)/2}$$

In other words, the change in each cell of a column is combined by a weighting system based on size of the transaction flows of each cell $(P_i X_{ij})$ and the total of the firm. While the value of each cell may be obvious, the aggregate measure can be problematic because the change in any given co-efficient between two different matrices can be positive or negative, and its contribution to total productivity change will depend on its assigned weight.[35]

Surprisingly, little attention was given to Leontieff's index in the productivity literature until recently. One possible explanation is that the index has been aimed at measuring only material productivity (and thus its relation to structural change) as opposed to the productivity of primary factors—labor and capital—that are the paramount interest of neoclassical economists. By introducing a measure of indirect effects of productivity changes, however, the I-O approach is a natural extension to the neoclassical measures of productivity change. For example, labor productivity can increase in the food industry, but does that mean less total labor is used per unit of food-industry output—i.e., direct labor used by the food

industry plus the indirect labor that went into the production of the capital used in the production process? If labor productivity in the food industry itself rose, but labor productivity in the food-processing machinery industry declined, it may be that total labor per unit of food processed, relative to the base period, also declined. In other words, by measuring output in terms of final demand rather than total output (direct and indirect), the productivity measure cannot reflect the overall effects of productivity growth.[36]

In order to show formally how the neoclassical measures may be misrepresenting productivity effects, consider the following demand-supply relationship:

V-4
$$P_i Q_i = \sum_j P_i X_{ij} + \sum_m P_i Y_{im}$$

where Y_{im} is demand for input i by final demand category m. In the case of a firm, $P_i Q_i$ would be shipments of output i, $\sum_j P_i X_{ij}$ would be total intermediate purchases of output i, and $\sum_m P_i Y_{im}$ would be final consumer purchases of output i. The aggregate value added identity would then be:

V-5
$$\sum_k \sum_j P_k V_{kj} = \sum_m \sum_i P_i Y_{im}$$

where V_{kj} = the amount of input of the kth factor of value added into product-line j. Then, the total input productivity (TFP) change can be derived as a change in output minus the weighted change in all inputs (material and primary inputs):

V-6
$$e_j = \dot{Q}_j - \sum_j a_{ij} \dot{X}_{ij} - \sum_j f_{kj} \dot{V}_{kj}$$

where $f_{kj} = P_k V_{kj}/P_j Q_j$ and $a_{ij} = P_i X_{ij}/P_j Q_j$. Combining terms, the aggregate index of total productivity change for the firm can be expressed as:

V-7
$$e = \sum_j w_j e_j$$

where $w_j = P_j Q_j/\sum_j P_j Q_j$.

To measure the primary inputs productivity without material productivity, which would be analogous to the productivity measure presented in the previous section, the change in value added is derived as:

$$V-8 \qquad Z_j^* = \frac{P_j^0 Q_j^1 - \sum_i P_i^0 X_{ij}^1}{P_j^0 Q_j^0 - \sum_i P_i^0 X_{ij}^0}$$

The numerator is the value added of period 1 in constant dollars and the denominator is the value added in the base period. The primary input productivity change in the jth sector is:

$$V-9 \qquad e_j^* = Z_j^* - \sum_j f_{kj}^* \dot{V}_{kj}$$

where $f_{kj}^* = P_k V_{kj}/(P_j Q_j - \sum_j P_i X_{ij})$. Equation V-9 calculates productivity of primary factors only (i.e., capital and labor), which is similar to the productivity calculation based on the neoclassical approach (see IV-4). The total productivity change is:

$$V-10 \qquad e_j = e_j^* (1 - \delta_j)$$

where $1 - \delta_j = (P_j Q_j - \sum_j P_i X_{ij})/P_j Q_j$ and $0 < 1 - \delta_j < 1$. It can be seen that primary input productivity change is always greater than the TFP change. Obviously, δ_j can be different for each product line as well as for each period. Therefore, productivity based on aggregate value added might be misrepresented for both time and cross-product comparisons.[37]

Most recently, Wolff has been able to enrich the I-O productivity analysis by linking the I-O model directly to the TFP approach.[38] An adaptation of the Wolff approach is presented briefly below. The basic TFP formula for primary inputs would be:

$$V-11 \qquad \rho = (P\frac{dy}{dt} - w\frac{dL}{dt} - r\frac{dK}{dt})/y$$

where $y = PY$ (i.e., P and Y are vectors of prices and final demand respectively). To show the identity with neoclassical equations used in the previous section, the TFP growth equation (V-11) can be rewritten as:

$$V-12 \qquad \rho = \frac{P\hat{Y}}{y}\dot{Y} - \frac{w\hat{L}}{y}\dot{L} - \frac{r\hat{K}}{y}\dot{K}$$

(since $\dot{Y} = \frac{\partial \ln Y}{\partial t}$), where the symbol, ^, above a vector represents a diagonal matrix with the diagonal composed of the vector elements. The coefficient $P_j Y_j/y$ is a share of the value of the output of product-line j. Coefficients wL_j/y and rK_j/y represent aggregate shares of labor and capital

inputs into product-line j, respectively. From the aggregate value added identity (V-5), we know that

$$\text{V-13} \qquad \sum_j \frac{wL_j}{y} + \sum_j \frac{rK_j}{y} = 1$$

Therefore, the I-O based TFP equation (V-12) is similar to the neoclassical-based TFP equation with constant returns to scale (IV-8).

With the above analogy in mind, the dual (cost) form of the TFP model can be derived by making the following simplification of notation. First, express a proportional change, dz/dt, as dz, and rewrite the I-O based TFP equations (V-11) as:

$$\text{V-14} \qquad \rho = (PdY - wdL - rdK)/Y$$

The basic Leontieff equation presented above (V-4) can be rewritten in the dual form of the production function as:

$$\text{V-15} \qquad P_j = \sum_i a_{ij}P_i + wl_j + rk_j$$

where $l_j = \dfrac{L_j}{Q_j}$ and $k_j = \dfrac{K_i}{Q_j}$. Based on equation V-12, Wolff modified the I-O based TFP equation V-15 into:

$$\text{V-16} \qquad \rho = -[Pda + wdl + rdk]Q/y$$

where "a" is the matrix of direct coefficients and l and k are vectors composed of elements l_j and k_j. The equation V-16 simply states that a decrease in any cell of matrix "a" or primary factor per unit of output will increase TFP. For an individual firm, the TFP growth measure for product-line j is:

$$\text{V-17} \qquad \pi_j = -(Pda_j + wdl_j + rdk_j)/P_j$$

where a_j is a jth column of matrix "a", and π_j is the equivalent of Leontieff's measure of technological change presented at the beginning of the section (V-3). In the Leontieff measure, the change in each direct coefficient is weighted by the average value of transaction flows in the numerator and the total difference is divided by the two-period average of the intermediate inputs. In Wolff's measuring scheme, weights are assigned according to prices of each intermediate input (the numerator) and total change is divided by the price of the industry's output. In some sense, therefore, Wolff's measure can be viewed as a dual measure to Leontieff's. The important addition made by Wolff is the inclusion of primary factor changes in Leontieff's measure of structural changes.

By combining the equations V-16 and V-17, Wolff has derived a definition of aggregate TFP in I-O form:

V-18
$$\rho = \pi \frac{\hat{P}Q}{y}$$

However, his major contribution is identifying the sources of TFP growth in an I-O format, which allows the inclusion of change in material inputs and the direct and indirect effects of each primary input—both of which are lacking in earlier neoclassical approaches. With that intent in mind, the TFP measure presented in V-17 can be expressed in terms of the total primary inputs requirement as:

V-19
$$\rho = - [wd\lambda + rd\gamma]Y/y$$

where λ and γ are total (direct and indirect) requirements for labor and capital per unit of total output. The corresponding measure of product-line technological progress can be expressed as:

V-20
$$\pi_j^* = - (wd\lambda_j + rd\gamma_j)/P_j$$

where π_j^* indicates the change in technology that is attributed to the labor and capital inputs involved in each of the material inputs used to produce a unit of output in product-line j of the firm. The difference between the TFP measure π_j and the technological progress measure π_j^* is established by the equation:

V-21
$$\pi_j^* = \pi(\hat{P}q\hat{P}^{-1})$$

where q is a Leontieff inverse: $q = (I - a)^{-1}$. Each π_i^* is a weighted average of each π_j. Weights are distributed according to the direct and indirect impact of each final demand m on product-line j (i.e., q_{ij}), which in turn is weighted by the price of good j (i.e., P_j^{-1}). Finally, these weights are multiplied by the price of good i, or:

V-22
$$\pi_j^* = P_i \sum_j \pi_i P_j^{-1} q_{ij}$$

Thus, π_j^* is TFP growth for product-line j.

The TFP growth for the firm can be expressed as technological progress, which is weighted by shares of each good in final demand, or:

V-23
$$\rho = \pi \cdot \frac{\hat{P} Y}{y}$$

By substituting V-21 into V-23, we derive

V-24
$$\rho = \pi S \beta$$

where $S = \hat{P}q\hat{P}^{-1}$ and $\beta = \frac{P_j Y_j}{y}$. Equation V-24 can now be used as the major equation for the derivation of the sources of change in TFP growth in terms of rate of the change in TFP, which is:

V-25
$$\Delta\rho \cong \pi S(\Delta\beta) + \pi(\Delta S)\beta + (\Delta\pi)S\beta$$

(For convenience, the second-order terms in the derivation of this equation are ignored). The first term indicates the effect of the change in the composition of final demand on the change in ρ. The second term indicates the effect of the cross-product-line transactions on the change in ρ. The last term reflects the effect of TFP growth of each product-line on the change in ρ.

As stated earlier, the neoclassical approach does not have a cross-product effect, ΔS. If $S\beta$ is aggregated into one term, $\alpha = S\beta$, however, the neoclassical approach can be represented as:

V-26
$$\Delta\rho \cong \pi(\Delta\alpha) + (\Delta\pi)\alpha$$

which is the analog of the TFP measure developed by Gollop. The last term on the right-hand side is identical to the last term of the equation V-25. Therefore, the first term is a combination of two effects: the effect of the change in the composition of final demand and the effect of the change in cross-product-line transactions. In other words, Wolff has demonstrated that the I-O model allows one to differentiate the effect of activities related to the change in the final demand composition from those related to the cross-product-line transactions. Moreover, in the neoclassical approach $\Delta\pi$ would include only the change of the primary factors, while the I-O approach allows one to examine all the intermediate inputs as well.

In traditional I-O studies, capital is considered as a primary input factor. By considering capital as a primary factor, however, one losses the indirect effect devoted to production of capital, which can be fairly significant (according to Wolff's empirical results). To avoid this shortcoming, Wolff augmented the I-O matrix by including the row and column that reflects the consumption and production of capital. Wolff also included noncompetitive imports inside the I-O table. More often, studies have assumed imports to be a primary factor. If so, however, the calculated total requirements treat imports as though they are free goods. For example, in the computation of total labor input, some part of productivity of foreign

workers would be ascribed to domestic labor. On the other hand, exports are included in final demand. Of course, one could assume that the economy produces just enough exports to compensate for its imports.[39] Wolff operated under a somewhat similar assumption. He assumed that non-competitive imports are equal to the "exogenous exports." The estimated endogenous export was netted out from the total exports, leaving "exogenous exports" as part of final demand. Endogenous exports were included inside the I-O table.

Wolff's paper was a considerable advance toward demonstrating the similarities between neoclassical and I-O productivity measurement. One criticism of Wolff's productivity measure is the output measure, which is a final demand (represented by equation V-6). However, it has already been shown (by Watanabe) that such measures of TFP are an underestimation. The full measure of output should include total sector's output (which is represented by equation V-11). This is an important point in TFP measurement within the I-O model. Nevertheless, Wolff should not be criticized too harshly for this shortcoming, because his purpose was to compare the I-O approach with the neoclassical approach. Since the neoclassical approach uses final demand as an output, Wolff was required to do the same.

There have been a number of attempts to satisfy the managers' analytical needs by conceptualizing an I-O approach at the firm level.[40] But for the most part, there have been very few actual applications of the I-O approach at the firm level. The reason for the lack of interest is not clear. One explanation that has been offered is that..."perhaps the key stumbling block to input-output use in corporate planning is that the assumptions and definitions concerning technological change, structural change, pricing and demand sensitivity, on which the input-output coefficient changes are based, are of very great interest to the planner and of perhaps least interest to the input-output analyst; in fact, most input-output analysts tend to sweep these assumptions under the rug, because they interfere with the mathematical elegancies of the input-output model or the simplicity of the analysis task."[41]

The latest developments in the application of I-O models for productivity measurement at the macroeconomic level, however, should be directly applicable at the micro level of the firm, once the techniques are refined for application with I-O tables of actual firms. The I-O model of the firm describes inputs of different materials and primary resources necessary to produce the multiproducts of a firm. It can, therefore, be a useful tool for managers to analyze productivity of each resource that enters the firm's particular production process. Utilizing Wolff's approach to productivity analysis, resources can be analyzed from the point of view of each output.

This could prove to be a major breakthrough for the I-O analysis of the firm.

VI) Conclusion: Comparative Strengths and Weaknesses

The production function was originally developed by economists as a purely economic concept to simplify the analysis of the behavior of markets. Production functions were never intended to be used to explain the behavior of real firms or industries, which is their primary value—if they are to have any at all—to managers. Managers are concerned about actual physical relationships among inputs in their own particular and complex production processes. From the managers' perspective, these relationships are predetermined for each of their products on the basis of engineering and production decisions relating to such factors as product design, material input specification, mix of labor skills, equipment characteristics, production schedules, quality standards, and maintenance requirements— none of which entered the above discussion of production functions and, indeed, have never been modelled successfully by empirical production functions. Nevertheless, TFP analysis is increasingly being used at all levels of the economy. Integrating productivity analysis into the decision-making processes of managers under these limitations is successful only with a comprehensive understanding of the relative strengths and weakness of the alternative approaches available to them.

To begin with, managers are likely to be most comfortable with the network approach, because of their training and their detailed understanding of the engineering design of their own firm's plant and technology. Access to confidential company data, plus a first-hand knowledge of each stage of the production process allows what might be termed "micro-micro" analysis of the firm's productivity performance and its relationship to the firm's overall profitability.

From a conceptual perspective, the basic strengths of the network approach are: 1) its ability to allow the managers to identify the sources of a technological impact and trace their impact through the economic structure of the firm; 2) because of its independence from an empirical production function, its ability to analyze productivity without the restrictive assumptions of the economists' production functions that managers tend to be critical of; and 3) its ability to utilize analytical tools, such as economic and financial ratios, with which managers are typically familiar. But, these strengths are not costless.

The primary weakness of the network approach, relative to neoclassical or I-O approaches, is its dependence on a substantial number of economic ratios without knowing what the equilibrium values would be, except by

using engineering knowledge or the industry performance as a norm. While the firm may be reassured to know that labor or investment productivity actually did increase as a result of a particular innovation, it may also be difficult to make relative comparisons of different innovations introduced at different plants or to compare different innovations introduced at the same time in a single plant. Nor does tracking the ratios simplify the analysis of scale effects, regulatory effects, or total productivity. Moreover, the requirement of detailed knowledge of the production process to extract the full benefit from the approach limits its usefulness to less experienced managers or outside analysts. In some sense, the network approach is like a Rubic's cube, without any guide as to how to condense all the information contained in the complete network of ratios.

In stark contrast, the neoclassical approach condenses much of the same information used in the network approach into a single, quantifiable measure of productivity performance that can easily be compared to productivity performance at any level of aggregation. Certainly, some knowledge of the firm, region, or industry is necessary to derive the full benefit from the approach. The analysis is standardized, however, so that contributions to productivity growth can be easily evaluated. Moreover, interplant and interindustry comparisons are greatly simplified. That simplification is achieved by a willingness to use an empirical production function as a tool to obtain quantifiable results, which is not a step lightly taken without a full understanding of its implications on productivity analysis. Obviously, the manager should not expect to be able to remove all the sources of error or bias in the TFP measure. If the major sources are addressed, however, the growth rate of the residual may reasonably be assumed to be proportional to the actual productivity growth of the firm.

Conceptually, the neoclassical approach is only as good as the underlying assumptions. In recent years, most of the more restrictive assumptions of the original Cobb-Douglas production function have been loosened or removed— some, such as the handling of idle capacity, are still open to debate. Nor have all the limitations in the neoclassical approach, relative to the network approach, been removed. Perhaps the most important restriction of the TFP approach relative to the network approach from the standpoint of managers is the link to profit performance. In its simplest form, the neoclassical approach assumes that every product line receives a normal profit, which in the long run is zero (i.e., total revenue equals total cost). Thus, all the managers need be concerned about is efficient allocation of inputs. Without a measure of the direct impact of technological changes on the firm's profitability, managers lose a valuable tool in formulating policies designed to improve profit performance through innovations.

Finally, in Leontieff's original formulation, the I-O approach had essentially only one advantage over the other approaches in its ability to analyze a complete set of intermediate (or, materials) input productivities, including direct and indirect effects of each input. Neither the network or neoclassic approach captured indirect effects and, until recently, the neoclassical approach included only primary inputs. With the modification of the I-O approach to include primary inputs, the I-O approach can be extended to incorporate many of the elements of the neoclassical approach, but still retain the unique perspective provided by its matrix. Moreover, even though the I-O approach lacks the financial ratios found in the network approach, it does extend the physical ratios of the firm in comparable detail across the whole set of production processes. In a sense, these ratios create a seamless web of interlocking relationships between the firm's performance and the regional or national economy that is unattainable in the networking approach. Among the strengths of the enhanced I-O approach then are that: 1) it allows greater differentiation between sources of productivity growth (particularly, the I-O approach can differentiate between changes in the composition of final demand and the change in the transactions between industries); 2) it goes beyond both neoclassical and network approaches to capture indirect effects of changing production processes on productivity growth; 3) it considers capital as a means of production, which is reproducible within the firm or industry, while the neoclassical approach treats capital essentially as exogenous (in other words, investment is endogenous and the productivity growth of the firm can be analyzed in a truly dynamic model), and 4) its ability to differentiate inputs of each industry into a single output makes the I-O approach ideally suited for analyzing the effects of sectoral shift on a plant's productivity.

The article of faith in the I-O approach is the validity of the matrix coefficients--whether at the national or firm level. Without confidence that the coefficients are reasonably accurate, the productivity measure is meaningless. Moreover, filling the matrix and updating it over time is a difficult and timely task that can limit its usefulness to managers. Finally, the labor and capital inputs are given only one price over all production processes, while both the neoclassical and network approaches allow differential wage rates and cost of capital. For these reasons, the neoclassical approach has several advantages over the I-O approach to productivity analysis: 1) the neoclassical approach utilizes a more flexible production function, allowing estimation of many important economic factors such as returns to scale and factor-augmenting technological bias; 2) it differentiates labor and capital into components (16 labor categories, in one case) and estimates the individual effect of each on productivity, compared to only one row of labor and capital coefficients in the I-O approach; and 3) with its flexibility, it potentially can provide a dynamic approach to the effects of capital adjustment and capital accumulation on productivity growth. (Although a

number of dynamic models have been constructed for I-O models, none have yet been applied successfully to productivity analysis).

After years of research and controversy, productivity analysis is finally coming of age. Interest in detailed analysis at all levels of the economy has never been greater, nor the tools more sophisticated and flexible. In response to the challenge to go beyond the popular but simplistic techniques based on labor productivity, more realistic models of complex production processes and the interrelationship of factors of production with productivity have been developed with the explicit needs of microeconomic analysis in mind. While much research remains, managers have at least three major alternatives to partial productivity analysis, which are all soundly based in economic theory. None is clearly superior, but each has relative strengths and weaknesses that make them more or less valuable to the manager, depending on the types of questions that need to be answered at any given time. The objective of productivity research should be directed towards bridging these three approaches. As this research continues, these tools can only improve the economic content of managers' decision making.

---

[1] Labor productivity derives its popularity both from its ease of measurement and from questions of relative welfare gains to society over time associated with labor productivity growth. See Fabricant, 1984, p. 4-8.

[2] Abramovitz and others were working on TFP measures in the 1950s and earlier, but Solow in his 1957 article is generally credited for the theoretical breakthroughs that led to more sophisticated measures of TFP. See Nelson, 1981.

[3] See Bela Gold, et al., 1984, p. 79.

[4] See Denny and Fuss, 1977, for a discussion of the separability issue.

[5] Because the second-order cross-partial derivatives are unknown, the rate of change is actually indeterminant, but is typically assumed to be negative in order to allow the isoquant to be strictly convex to the origin. See Hadar, 1971, p. 27.

[6] See Kendrick, 1973, p. 32.

[7] Implicit assumptions are constant returns to scale and producer equilibrium. For details, see Robert M. Solow, 1957, p. 313.

[8] The formal proof that $T\dot{F}P$ is the distance O^*M/q_t is as follows:

$$\overline{LO^*} = \overline{NP}, \text{ and}$$

$$\overline{NP} = (\partial q/\partial k)\Delta k, \text{ where } \Delta k = k_{t+1} - k_t$$

$$\overline{O^*M} = q_{t+1} - q_t - (\partial q/\partial k)k = \Delta q - (\partial q/\partial k)\Delta k.$$

Since $T\dot{F}P = \overline{O^*M}/q_t$,

$$T\dot{F}P = \Delta q/q_t - (\partial q/\partial k)(k_t/q_t)(\Delta k/k_t) = \dot{q} - v_K(\dot{k}) .$$

Q.E.D.

[9] For a standard discussion of the properties of linearly homogeneous functions, see Chiang, 1967, p. 371-376.

[10] For a more detailed discussion, see Sato, 1981.

[11] See Nadiri, 1970, p. 1142.

[12] See Nadiri, ibid, p. 1143.

[13] See Salter, 1969

[14] For a description of the Farrell index, see Baird, 1977.

[15] See Ames, 1969.

[16] For a detailed critique of these approaches, see Samuel Eilon and Judith Soesan, 1976, p. 7-13.

[17] See Gold, 1974, p. 79.

[18] A rearrangement of the variables would produce the relevant equations for investment and materials productivity. See Gold, 1955, pp. 88-91, for an example of how investment productivity can be related to labor productivity in the following way:

$$\frac{\text{Productive capacity}}{\text{Fixed Investment}} = \frac{\text{Man-Hours}}{\text{Material Volume}} \, x$$

$$\frac{\text{Materials Volume}}{\text{Fixed Investment} \, x \, \dfrac{\text{Output}}{\text{Capacity}}} \, x \, \frac{\text{Physical Output}}{\text{Man-Hours}}$$

[19] See Gold, 1974, p. 66.

[20] See Gold, 1979.

[21] For an example of an actual application of the networking approach at the firm level see, Eilon, et al., 1976.

[22] See Jorgenson and Griliches, 1967.

[23] For a more detailed discussion, see Gollop and Jorgenson, 1980.

[24] Denison, 1972, and Norsworthy, et al., 1979, have stressed the diversity of factors that might be captured in the residual, using their accounting approach to TFP analysis. The emphasis in this discussion, however, is in analyzing the shift in the production function, which was the focus of Solow's original article.

[25] For the exact derivation of the equation, see Gollop, 1985, p. 181-184. The price of value added is a composite price of factors contributing to value added, such as wages, cost of capital, taxes, and imports. Since many of these prices are difficult to measure, the value-added price is derived by a double-deflation process (i.e., first, physical value added is derived by taking the difference between physical total output and physical material inputs, and then nominal value added is divided by physical value added to compute a price).

[26] See Gollop and Roberts, 1984. Sato, 1981, has an alternative approach using Lie-Group Theory.

[27] For a formal proof, see Denny, Fuss, and Waverman, 1981.

[28] For a detailed discussion, see Gollop and Roberts, 1984.

[29] For example, see Atkinson and Haverman, 1984. For another discussion of the issue, see Denny, Fuss, and Waverman, 1981.

[30] See Morrison and Brendt, 1981.

[31] See Shankerman and Nadiri, 1984.

[32] See A. Erdilek, 1977, p. 45.

[33] See Leontieff, 1955, p.18.

[34] See Leontieff, 1941, p. 46. Elsewhere, it has been argued that there are three cases in which technological change can occur: 1)a change in one or more coefficients, 2)a change in the given lower limits of the primary inputs, and 3)a change in the function of the primary inputs that is to be minimized. See Herbert Simons, 1951, p. 260.

[35] See Domar, 1961.

[36] The calculation of this type of labor productivity was presented in Beeson and Schlutter, 1981. The article, as well, represents an example of how labor productivity along with material productivity can be measured with the I-O model.

[37] For a more detailed discussion at the industry level, see Watanabe, 1971.

[38] See Wolff, 1985. Wolff's research follows the works of Rymes, 1971, and Peterson, 1979.

[39] For a discussion of the limitations of this assumption, see Beeson and Schlutter, 1981.

[40] See A. Erdilek, 1977.

[41] See Gols, 1985, p. 27.

[42] For a discussion of the shortcomings of empirical production functions in measuring productivity, see Baird, 1977.

[43] From a measurement perspective, there are also advantages related to the ability of the network approach to adjust for idle capacity and thus avoid the need to estimate capital stock altogether.

Bibliography

Ames, E., 1961, "Research, Invention, Development and Innovation," *American Economic Review*, June, pp. 370-381.

Atkinson, S. E., and R. Halvorsen, 1984, "Parametric Efficiency Tests, Economies of Scale, and Input Demand in U. S. Electrical Power Generation," *International Economic Review*, Vol. 25, no. 3, pp. 647-659.

Baird, R. N., 1977, "Production Functions, Productivity, and Technological Change," in Bela Gold (ed.), *Research, Technological Change, and Economic Analysis*, Lexington: D.C. Heath Co., pp. 11-41.

Beeson, P. and G. Schlutter, 1981, "Components of Labor Productivity Growth in the Food System," *Journal of Political Economy*, Vol. 92, August, pp. 378-384.

Carter, Anne P., 1970, "A Linear Programming System Analyzing Embodied Technological Change," in A. P. Carter and A. Brody (ed.), *Contributions to Input-Output Analysis*, New York: American Elsevier Publishing Co., pp. 77-98.

Chiang, A. C., 1967, *Fundamental Methods of Mathematical Economics*, New York: McGraw Hill Co.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau, 1973, "Transcendental Production Frontiers," *The Review of Economics and Statistics*, Vol. 55, no. 1, pp. 28-45.

Dennison, E. F., 1972, "Some Major Issues in Productivity Analysis: An Interpretation of Estimates by Jorgenson and Griliches," *Survey of Current Business*, Vol. 52, no. 5, May, pp. 37-63.

Denny, Michael, and Melvyn Fuss, 1977, "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates," *The American Economic Review*, Vol. 67, no. 3, June, pp. 404-418.

Denny, M., and M. Fuss, and L. Waverman, 1981, "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications," in Thomas C. Cowing and Rodney E. Stevenson (ed.), *Productivity Measurement in Regulated Industries*, New York: Academic Press, pp. 179-217.

Domar, E. D., 1961, "On the Measurement of Technical Change," *Economic Journal*, December, pp. 709-729.

Eilon, Samuel, and Judith Soesan, 1976, "Definitions and Prevailing Approaches" in S. Eilon, et al, (ed.), *Applied Productivity Analysis for Industry*, Elmsford: Pergamon Press, Inc.

Eilon, Samuel, Bela Gold, and Judith Soesan, 1976, *Applied Productivity Analysis for Industry*, Elmsford: Pergamon Press, Inc.

Erdilek, Asim, 1977, "Productivity, Technological Change, and Input-Output Analysis," in Bela Gold (i.e.,), *Research, Technological Change, and Economic Analysis*, Lexington: D.C. Heath and Company, pp. 43-85.

Fabricant, Solomon, 1984, "Measurement: an Overview," in *Measuring Productivity*, New York: UNIPUB.

Gold, Bela, 1955, *Foundations of Productivity Analysis*, Pittsburgh: University of Pittsburgh Press.

Gold, Bela, 1979, *Productivity, Technology, and Capital*, Lexington: Lexington Books.

Gold, Bela, William S. Peirce, Gerhard Rosegger, and Mark Perlman, 1984, *Technological Progress and Industrial Leadership*, Lexington: D.C. Heath and Company.

Gollop, Frank M., 1985, "Analysis of the Productivity Slowdown: Evidence for a Sector-Biased or Sector-Neutral Industrial Strategy," in W. J. Baumol and K. Melennan (ed.), *Productivity Growth and U.S. Competitiveness*, New York: Oxford University Press.

Gollop, F. M., and M. J. Roberts, 1981, "The Sources of Economic Growth in the U.S. Electric Power Industry," in Thomas G. Cowing and Rodney E. Stevenson (ed.), *Productivity Measurement in Regulated Industries*, New York: Academic Press, pp. 107-143.

Gollop, F. M., and M. J. Roberts, 1981, "Imported Intermediate Input: Its Impact on Sectoral Productivity in U.S. Manufacturing," in Ali Dogramaci and Nabil R. Adam (ed.), *Aggregate and Industry Level Productivity Analyses*, Boston: Martinus Nijhoff Publishing, pp. 149-185.

Gollop, Frank M., and Dale W. Jorgenson, 1980, "U.S. Productivity Growth by Industry, 1947-1973," in John W. Kendrick and Beatrice N. Vaccara (ed.), *New Developments in Productivity Measurement and Analysis*, vol. 44, Chicago: University of Chicago Press, pp. 17-136.

Gols, A.G., 1985, "The Use of Input-Output in Industrial Planning," *Business Economics*, May, pp. 19-27.

Hadar, Josef, 1971, *Mathematical Theory of Economic Behavior*, Reading: Addison-Wesley.

Jorgenson, D. W., and Z. Griliches, 1972, "The Explanation of Productivity Change," *Survey of Current Business*, Vol. 52, no. 5, May, pp. 3-36.

Kendrick, John W., 1973, "Postwar Productivity Trends in the United States, 1948-1969," *NBER General Series* # 98.

Kennedy, Charles, and A. P. Thirlwall, 1972, "Surveys in Applied Economics: Technical Progress," *The Economic Journal*, March, pp. 11-72.

Leontief, W.W., 1941, *The Structure of the American Economy, 1919-1929*, Cambridge: Harvard University Press.

Leontief, Wassily, et al, 1953, *Studies in the Structure of the American Economy*, New York: Oxford University Press.

Miller, R. E., and P. D. Blair, 1985, *Input-Output Analysis: Foundations and Extensions*, Englewood Cliffs: Prentice Hall, Inc.

Moon, Douglas, 1981, "Technological Change and Productivity in Input-Output Analysis and the Potential of Sectoral Optimization Models," in Ali

Dogramaci and Nabil R. Adam (ed.), *Aggregate and Industry Level Productivity Analyses,* Boston: Martinus Nijhoff Publishing, pp. 31-51.

Morrison, C. J., 1985, "Productivity Measurement with Nonstatic Expectations and Varying Capacity Utilization: An Integrated Approach," *National Bureau of Economic Research Working Paper,* # 1561, February, 39 pp.

Morrison, C. J., and E. R. Brendt, 1981, "Short-Run Labor Productivity in a Dynamic Model," *Journal of Econometrics,* Vol. 16, pp. 339-365.

Nadiri, M. I., 1970, "Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey," *Journal of Economic Literature,* Vol. 8, no. 4, December, pp. 1137-1177.

Nadiri, M. I., and M. A. Shankerman, 1981, "Technical Change, Returns to Scale, and the Productivity Slowdown," *American Economic Review,* Vol. 71, no. 2, pp. 314-319.

Nelson, Richard R., 1981, "Research on Productivity Growth and Productivity Differences: Dead Ends and New Departures," *Journal of Economic Literature,* Vol. 19, September, pp. 1029-1069.

Norsworthy, J. R., Michael Harper, and Kent Kunze, 1979, "The Slowdown in Productivity Growth: Analysis of Contributing Factors," *Brookings Papers on Economic Activity,* Vol. 2 1979, pp. 387-445

Norsworthy, J. R., 1984, "Growth Accounting and Productivity Measurement," *Review of Income and Wealth,* Vol. 30, no. 3, September, pp. 309-329.

Ozaki, Iwao, 1970, "Economies of Scale and Input-Output Coefficients," in A.P. Carter and A. Brody (ed.), *Applications of Input-Output Analysis,* vol. 2, Amsterdam: North Holland Publishing Company, pp. 280-304.

Pratten, C. F., 1971, *Economies of Scale in Manufacturing Industry,* Cambridge: Cambridge University Press.

Rymes, T. K., 1971, *On Concepts of Capital and Technical Change,* Cambridge: Cambridge University Press.

Salter, W. E. G., 1969, *Productivity and Technical Change,* second ed., Cambridge: Cambridge University Press.

Sato, R., 1980, "The Impact of Technical Change on the Holotheticity of Production Functions," *Review of Economic Studies,* Vol. 48, pp. 767-776.

Sato, R., and P. S. Calem, 1981, "Lie Group Methods and the Theory of Estimating Total Productivity," in Thomas C. Cowing and Rodney E. Stevenson (ed.), *Productivity Measurement in Regulated Industries,* New York: Academic Press, pp. 145-167.

Simon, H. A., 1951, "Effects of Technological Change in a Linear Model," in Tjalling C. Koopmans (ed.), *Activity Analysis of Production and Allocation,* New York: John Wiley & Sons, Inc., pp. 260-277.

Solow, R. M., 1957, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics,* Vol. 39, no. 3, pp. 312-320.

Sudit, E. F., and Finger, N., 1981, "Methodological Issues in Aggregate Productivity Analysis," in Ali Dogramaci and Nabil R. Adam (ed.), *Aggregate*

and Industry Level Productivity Analyses, Boston: Martinus Nijhoff Publishing, pp. 7-29.

Watanabe, T., 1971, "A Note on Measuring Sectoral Input Productivity," *Review of Income and Wealth*, December, pp. 335-340.

Wolff, E. N., "Industrial Composition, Interindustry Effects, and the U.S. Productivity Slowdown," *Review of Economics and Statistics*, Vol. 61, no. 2, May, 1985, pp. 268-277.