Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey

M. Ishaq Nadiri


Your use of the JSTOR database indicates your acceptance of JSTOR’s Terms and Conditions of Use. A copy of JSTOR’s Terms and Conditions of Use is available at http://www.jstor.org/about/terms.html, by contacting JSTOR at jstor-info@umich.edu, or by calling JSTOR at (888)388-3574, (734)998-9101 or (FAX) (734)998-9113. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of Economic Literature is published by American Economic Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aea.html.

Journal of Economic Literature
©1970 American Economic Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©1999 JSTOR
Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey

By M. Ishaq Nadiri
Professor of Economics, New York University
Senior Research Staff Member, National Bureau of Economic Research

I am grateful to my colleagues at the National Bureau, Professors V. K. Chetty, Solomon Fabricant, John Hause, Christopher Sims, Finis Welch, and Mr. Paul Wachtel for reading an earlier draft of this paper. I also am indebted to Professor Marc Nerlove of the University of Chicago and Mr. Edward Denison of the Brookings Institution for their comments on a previous draft of this paper. Thanks are due, for their helpful comments, to Professors Zvi Griliches and Romesh Diwan of Harvard University and Rensselaer Polytechnic Institute respectively. Many thanks are due to Miss Veena Bhatia for her competent assistance and comments which made the writing of the paper possible and to Mrs. Rose Ferro for her help in the preparation of this manuscript. This paper is closely related to a study of interrelated factor demand functions which Sherwin Rosen and I are currently undertaking at the National Bureau.

Productivity change is both the cause and the consequence of the evolution of dynamic forces operative in an economy—technical progress, accumulation of human and physical capital, enterprise, and institutional arrangements. Its measurement and the interpretation of its behavior at the microeconomic and macroeconomic levels require the untangling of many complex factors; it is a task that has been a major challenge to economists and of extreme interest to entrepreneurs and government policy-makers. There has been a recent upsurge of interest in the measurement and explanation of productivity changes because of the development of new theoretical models, the availability of new and better data and estimation techniques, and the advent of large scale computers that have made possible the testing of refined hypotheses. However, the issues involved are too numerous and too complex, and the available empirical evidence too sketchy (and often inconsistent) to allow bold conclusions about the measurement—and the determinants—of changes in productivity.

Our task in this survey is to concentrate on some basic theoretical hypotheses and empirical evidence recently reported. No effort is made to be conclusive; limitations of time, space, and ability have made it necessary to confine the discussion to a few topics. The emphasis of the paper is on the determinants of total factor productivity in the U.S. economy (with only a limited discussion of the productivity changes in the various industries). No attempt is made to discuss explicitly the research related to short-run factor productivity (short-run employment functions) and international comparisons of factor productivity.

1 Interested readers are referred to the important works on technological change and factor productivity: M. Brown [15, 1956], J. Kendrick [78, 1961], L. Lave [86, 1966], and the NBER volume The theory and empirical analysis of production functions edited by M. Brown [16, 1967].
The plan of the paper is as follows. Section I comprises a preliminary review of the indices of productivity, some empirical evidence, and a brief discussion of the forces which may govern the behavior of productivity changes. Two major topics are covered briefly in Section II: (i) the problems of aggregation underlying the aggregate production function, and (ii) some theoretical issues on the nature of technical change and problems of its diffusion. These topics are discussed both to indicate the types of dynamic factors involved in shaping the course of productivity movements, and also to consider the forces which are generally abstracted from in the discussions of total factor productivity and the aggregate production function.

The next two sections are devoted to the calculation of technical change based on the aggregate production function. In Section III an attempt is made to discuss some of the main controversial issues concerning the specification and estimation of the aggregate production function. The main emphasis in this section is on the attempts to generalize the form of the aggregate production function (and some issues associated with estimating the parameters of the aggregate production functions). Section IV is devoted to a discussion of the ways technical change is linked to the conventional inputs as well as the use of the aggregate production function as a schema for growth accountancy to calculate the contribution of various factors to the growth of output.

Some concluding remarks and suggestions for future research are presented in the final section.

**SECTION I**

**Indices of Factor Productivity, Some Stylized Facts and a Preliminary Look at Technical Change**

I.1 Measures of Factor Productivity

Productivity is often measured as a ratio of output to inputs. There are as many indices of productivity as there are factors of production. While each index has its own use, the most important and most often used are the partial productivity indices of labor and capital and the total or multifactor productivity index. The former indices are simply the average products of labor, or capital, while total factor productivity, often referred to as the "residual" or the index of "technical progress," is defined as output per unit of labor and capital combined. Symbolically, these indices are:

\[ AP_L = \frac{Q}{L}; \quad AP_K = \frac{Q}{K}; \]

(1)

(b) Total productivity index:

\[ A = \frac{Q}{(aL + bK)} \]

where \( Q, L, \) and \( K \) are, respectively, the aggregate level of output, labor, and capital inputs; \( a \) and \( b \) are some appropriate weights.

There are many ways of measuring total factor productivity, but the two indices most often used in empirical research are Kendrick's arithmetic measure [78, 1961] and R. Solow's geometric index [138, 1957]. Kendrick approaches measurement of \( dA/A \) using a distribution equation. He implicitly assumes a homogeneous production function and the Euler condition to obtain the following measure.²

\[ \frac{dA}{A} = \frac{Q_1/Q_0}{(wL_1+rK_1)/(wL_0+rK_0)} - 1 \]

² The weights in this measure are changing over time and the aggregate production function consistent with this index is

\[ Q = \frac{tKL}{(cL^\rho + dK^\rho)^{1/\rho}} \]

which is a linear homogeneous production function with constant elasticity of substitution, \( \sigma = 1/(1+\rho) \), and \( c \) and \( d \) are the efficiency parameters, \( \rho \) is the elasticity parameter, and \( t \) is the disembodied neutral technical change. Kendrick and R. Sato [80, 1963] report \( \sigma = .8 \) and the production function

\[ Y = A^{0.911} \frac{KL}{(cL^{2/3} + dK^{2/3})^{1/3}} \]

for the U.S. economy over the period 1919–60.
where \( w \) and \( r \) are the wage rate and the rate of return on capital respectively, and variables with the subscript 1 refer to the current period and those with the subscript 0 refer to the base period. In empirical estimates the weights for calculating (1a) are often permitted to change smoothly over time [Kendrick, 78, 1961]. Solow’s measure is based on the Cobb-Douglas production function with constant returns to scale and autonomous and neutral technological change, i.e.,

\[
\frac{dA}{A} = \frac{dQ}{Q} - \left[ \alpha \frac{dL}{L} + \beta \frac{dK}{K} \right]; \\
\beta = (1 - \alpha)
\]

where \( \alpha \) and \( \beta \) are the shares of labor and capital and \( dQ, dL, \) and \( dK \) are the time derivatives of \( Q, L, \) and \( K. \) Under the assumption of competitive equilibrium the Kendrick measure (1a) can be stated as

\[
(1c) \quad \frac{dA}{A} = \frac{Q_1/Q_0}{\alpha_0 \left( \frac{L_1}{L_0} \right) + \beta_0 \left( \frac{K_1}{K_0} \right)} - 1
\]

This is equivalent to Solow’s measure for small changes in the quantities of inputs and outputs [89, 1966].

1.2 Some Facts about Factor Productivity

More than a decade ago M. Abramovitz [1, 1956], S. Fabricant [46, 1959], Kendrick [78, 1961], and Solow [133, 1957] established on the basis of measures such as (1a) and (1b) that the conventionally measured inputs, capital and labor, leave a large portion of the growth of output unexplained. Since then considerable research on the measurement, determinants and consequences of factor productivity has been undertaken. Just to give the reader some idea about the empirical behavior of factor productivity and some related issues, we present the following facts based on Kendrick’s recent work [79, 1970] on the U.S. economy.

(a) The rate of growth of total private output increased from 2.8 percent during 1948–66, but with considerable variation over the period due to cyclical factors.

(b) The growth rate of labor input was fairly steady, though with some fluctuation. However, the growth rate of capital services varied considerably, mainly due to fluctuations in its rate of utilization. In the postwar period the growth rate of capital was considerably higher than that of labor.

(c) Though capital accumulation increased since 1910, the rate of return on capital remained constant and the share of labor rose considerably. Both of these were made possible by the advancement of technical change, permitting substitution of capital for labor.

(d) The rate of increase in labor productivity was generally low in the first phase and very high towards the end of contractions and the beginning of expansions. Short-run production functions generally indicate increasing returns to labor alone.

(e) Total factor productivity increased by about 2.1 percent per annum for the whole period—but with considerable yearly fluctuations. It averaged, for example, 2.7 percent per year in expansion phases and only 0.2 percent in cyclical contractions.

(f) There was substantial variation among the rates of growth in various industries. The dispersion exhibited wide fluctuations, depending on the growth rate of the overall economy. Communications and transportation generally showed greater productivity growth than mining, manufacturing, and services. Within manufacturing the rate of productivity growth was highest in rubber, transportation equipment, chemicals, and a few other industries—mainly durables.

(g) There was a mild acceleration in the rate of growth of factor productivity after World War II, perhaps due to the stability of the economy. In addition, the dispersion in the rates of growth among
industries has declined. Technological change has been more widely diffused. The costs of both capital and labor have decreased and the capital-labor ratio has increased in most of the industries during the postwar period.

Thus the picture is one of diversity. The economy grew at different rates in different periods, and different industries have been responsible for larger or smaller contributions to the growth of the aggregate economy. There is also evidence that suggests an interdependency between the overall growth rate and the industrial structure of the economy. The factors responsible for these empirical results are difficult to determine precisely, but some of the most frequently mentioned are stated on pages 1141–42 and discussed at length in subsequent sections.

I.3 Technical Change and the Production Function: A Preliminary Look

As mentioned earlier, the productivity indices are deduced either from an explicitly defined production function or from a distribution theory where the production function is implicit. Thus the accurate specification of the form and estimation of the parameters of the production function, such as $\alpha$ and $\beta$ in (1b) are crucial to the measurement of these indices. Consider the aggregate, two-factor, twice-differentiable production function

\[ (2) \quad Q = AF(L, K) \]

where $A$ is a measure of disembodied technical change and the function $F$ is homogeneous.\(^4\) Differentiating with respect to time and dividing by $Q$, we get

\[ (3) \quad \frac{dA}{A} = \frac{dQ}{Q} - \left[ \frac{dLF_F}{Q} \frac{dL}{L} + \frac{dKF_F}{Q} \frac{dK}{K} \right] \]

where $F_L$ and $F_K$ are the partial derivatives of output with respect to $L$ and $K$ and the variables preceded with a $d$ refer to time derivatives of the variables of (2).

It is clear from (3) that the magnitude of the residual ($dA/A$) and its stability over time depend upon: (i) the form of the production function that governs the behavior of $F_L$ and $F_K$, (ii) proper measurement of $L$ and $K$ and adjustment for their quality changes, and (iii) the importance of variables other than $K$ and $L$ (such as the entrepreneurial ability, or inventories) that are left out of the production function.

Suppose the production function (2) is $Q = AL^\alpha K^\beta$, then equation (3) reduces to (1b)—if the share of labor, $\alpha$, is assumed to be invariant with respect to $dL/L$ and $dK/K$ and constant returns to scale prevail ($\beta = 1 - \alpha$). Thus, any errors due to misspecification of the form of the function will spill-over into the measure of $dA/A$.\(^5\) Similarly, if the inputs $K$ and $L$ are measured erroneously, say, by multiplicative factors $v_L$ and $v_K$ denoting the quality improvement of $L$ and $K$, then it is easy to show that

\[ (3a) \quad \frac{dA}{A} = \alpha \left( \frac{dv_L}{v_L} \right) + (1 - \alpha) \left( \frac{dv_K}{v_K} \right) \]

That is, the “residual” becomes a weighted sum of the growth rates of the quality

\(^3\) The weights $\alpha$ and $\beta$ are often measured by the share of the inputs using market prices. These weights need not remain constant over time or from industry to industry. In a dynamic context, these productivity indices also suffer from the usual index number problems, i.e., selection of the base year.

\(^4\) For simplicity, time subscripts have been omitted throughout the paper except where necessary.

\(^5\) This can be observed in an example given by R. Nelson [109, 1969]. Suppose the underlying production function is CES and not Cobb-Douglas and the input shares are assumed constant. We may write the equivalent of (3) as

\[ (3a') \quad \frac{dA}{A} = \frac{dQ}{Q} - \left[ \alpha \frac{dL}{L} + (1 - \alpha) \frac{dK}{K} \right] - \frac{1}{2} \alpha(1 - \alpha) \left[ \frac{\sigma - 1}{\sigma} \right] \frac{dK}{K} - \frac{dL}{L} \]

where $\sigma$ is the elasticity of substitution between $K$ and $L$. Thus the measure of technical change (1b) differs from (3a') by the term \(-\frac{1}{2} \alpha(1 - \alpha)(\sigma - 1)/\sigma d[K/K - dL/L] \). If $\sigma < 1$ and $dK/K > dL/L$, then $Q$ must grow at a lower rate than the index of $K$ and $L$ combined.
changes "embodied" in the conventional inputs. Similar results are obtained when a third factor is left out of the production function (9). Suppose the function is defined as 

\[ Q = AL^aK^bE^c \]

then the corresponding productivity relation would be

\[
\frac{dA}{A} = \frac{dQ}{Q} - \left[ \alpha \left( \frac{dL}{L} \right) + \beta \left( \frac{dK}{K} \right) + \gamma \left( \frac{dE}{E} \right) \right]
\]

(3b)

where \( E \) is the omitted variable. If every factor is paid its marginal product and output is exhausted, the contribution of \( E \) to total productivity is \( dA/A \) of equation (3).

It is clear then that any mis specification or errors in estimating the parameters of the aggregate production function—errors in measuring the variables, errors due to omission of relevant inputs—will spill over to the measure of total factor productivity. If these sources of bias are successfully removed, the remaining portion of \( dQ/Q \) unexplained by the combined rate of growth of all the factors of production is the measure of "true" total factor productivity or technical change. A substantial portion of the literature on factor productivity, as we shall see in the following pages, is devoted to removing these biases and to explaining the determinants of the "unbiased" rate of technical change in the economy.

I.4 The Sources of Factor Productivity and the Nature of Technical Change

Assuming for the moment that the aggregate production function exists and is specified accurately, and that the inputs are "properly" measured, what can we say about the forces which explain movements of the purely descriptive factor productivity indices, (1a) and (1b)? Numerous forces are involved and the task of sorting them into well-specified categories is a difficult and hazardous one. Nonetheless, two major sets of factors have been suggested as the determinants of factor productivity: the technical characteristics of the production process and the movement of the relative factor prices. The often mentioned technical characteristics are:

(i) The efficiency of production, i.e., reducing the unit cost of all factors of production equally by applying better techniques;

(ii) the bias in technical change, i.e., the nature of the new technique is such that it leads to a greater saving in one input than in the other;

(iii) the elasticity of substitution, which measures the ease of exchanging factors of production in the course of the production process;

(iv) the scale of operation of the production process, i.e., economies (diseconomies) that arise due to changes in the scale of operation of the economy;

(v) the homotheticity of the production function, i.e., whether the returns to scale are evenly distributed among all factors of production.

These concepts can be illustrated by the usual isoquant diagram. The first three characteristics can be depicted as shifts of the unit isoquant \( Q_1 \) toward the origin over time as a consequence of technical progress, each isoquant representing the same level of output. Consider Figure 1. Neutral technical change is measured by the extent of the parallel shifts of the isoquant (e.g., from \( Q_1 \) to \( Q_3 \)) toward the origin. A change in the position of the isoquant (more toward one axis than the other) represents bias in technical change—the transition from \( Q_1 \) to \( Q_3 \) for example, leads to proportionately greater saving of labor for all techniques.

When technical advance leads to a change in the curvature of the isoquants such that capital and labor are more specific, then \( \sigma \), the elasticity of substitution, is reduced. This is illustrated by the movement of the isoquant from \( Q_1 \) to \( Q_4 \).  

\* From an individual firm's viewpoint, the \( \sigma \) is a datum. But the design of the new technique in capital
Figure 1

Note that the ratio of relative prices represented by the slope of the lines, \( L_1K_1 \ldots L_4K_4 \) is not held constant. If the relative prices remained constant over time (implying parallel shifts of \( L_iK_i \)), then the locus of points of tangency on the homothetic isoquants \( Q_iQ_j \) will be a straight line through the origin.

The scale effects can also be illustrated using Figure 1—but note that the isoquants now represent different levels of output. Part of the scale effect is similar to that of neutral technical change—it describes the saving of both inputs due to an increase in the scale of operations of the economy. This effect is exhibited in Figure 1 by the movement of the isoquant like the one from \( Q_1 \) to \( Q_2 \) provided \( Q_2 > Q_1 \). The bias due to scale change or nonhomotheticity of the production function is conceptually analogous to the bias due to technical change. It can be visualized as a movement like the transition from \( Q_1 \) to \( Q_2 \) in Figure 1. (Again, \( Q_2 \) must be greater than \( Q_1 \).) That is, as the scale of operations of the economy increases, given all other technical properties of the production process, unit requirements of labor will decline compared to those of capital.

Besides the solely technical aspects, movements in relative prices influence factor productivity via their effect on \( K/L \), i.e., increasing the employment of one factor of production at the expense of the other. Of course, the effectiveness of changes in factor prices depends upon the elasticity of substitution, \( \sigma \). For example, if \( \sigma = 1 \), the changes in relative prices have substantial effects on factor productivity; and if \( \sigma = 0 \), they have no effect.

Unfortunately, these characteristics of technical progress are highly interdependent and cannot easily be distinguished except for analytical purposes; and over and above these there are other conceptual problems that deserve consideration:

(a) there is no agreement on the precise definition of bias in technical change. There are several ways of defining technological bias [139, J. Stiglitz and H. Uzawa, 1969]. Change in relative shares of the inputs is often used as a measure of technical bias. The Hicksian definition measures the bias along a constant capital-labor ratio; the Harrodian definition measures the bias along a constant capital-output ratio; and Solow’s definition measures the bias along a constant labor-output ratio. Symbolically:

\[
\left( \frac{\partial (F_K K)/(F_L L)}{\partial t} \right)_{constant} \geq 0
\]

Hicks

\[
\begin{align*}
\text{capital-saving} & \quad \text{labor-saving} \\
\text{neutral} & \quad \text{neutral}
\end{align*}
\]
\[
\left( \frac{\partial (F_RK)}{\partial t} \right)_{K/Q \text{ constant}} < 0
\]

Harrod labor-saving
neutral capital-saving

\[
\left( \frac{\partial (F_LK)}{\partial t} \right)_{L/Q \text{ constant}} < 0
\]

Solow labor-saving
neutral capital-saving

depends upon the ratio \( \lambda_1/\lambda_2 \); technical change is Hicks neutral if the ratio is
constant, Harrod neutral (labor-augmenting) if \( \lambda_2 \) is constant, and Solow-neutral (capital-augmenting) if \( \lambda_1 \) is constant. The bias in technical change, \( B \), can be defined as:

\[
B = \left[ \frac{d\lambda_1}{\lambda_1} - \frac{d\lambda_2}{\lambda_2} \right] \left( 1 - \frac{1}{\sigma} \right)
\]

The embodiment effect should be

The labor-augmenting bias of technical change is

\[
C = (1 - \sigma)d\lambda_1/\lambda_1
\]

and the capital-augmenting bias is

\[
D = (1 - \sigma)d\lambda_2/\lambda_2.
\]

See Solow [136, 1967].

clearly distinguished from the augmentation effect and quality-correction of the inputs. The augmentation effect means that the productivity increase of an input due to technical advances is expressed as equivalent to a specific increase in its quantity. Embodiment of technical change in capital for example, could perfectly well produce purely labor-augmenting (but nonetheless capital embodied) technical change. Nor should all quality improvement in an input be considered equivalent to the embodiment effect. The latter refers only to quality improvement associated with vintage of capital or cohort of labor. For example, productivity increases due to sex and race characteristics (at a point in time) are not part of the embodiment effect, while improvements due to age and education are part of it.\(^8\)

(c) Finally, these technical characteristics do not remain constant over time or over different productive units. This raises the inevitable problems of aggregation—the subject of the next section.

SECTION II

A Few Necessary Digressions

Productivity relations such as (3) are derived from aggregate production functions which rest on certain strict assumptions about the behavior of microeconomic production units and the properties of the inputs and outputs. The production functions are equilibrium concepts which abstract from the dynamic forces that lead to technical change. It is important to note the aggregation problems in order to interpret

\(^8\) Embodiment is operationally distinct from other sorts of quality change only for fixed and quasi-fixed inputs and not for variable inputs. The age-mix of variable inputs can be readily changed in response to the changes in relative prices and new technology. That is, there is no cost of adjustment in hiring and firing such inputs. See C. A. Sims [131, 1967] for further discussion of these points.
the results based on aggregative relations and to consider the factors which lead to the development, adoption, and transmission of new techniques among firms and possible structural changes in the economy. We shall discuss these two sets of problems briefly before returning to the discussion of total factor productivity—the central theme of this paper.

II.1 Aggregation of Heterogeneous Inputs and Production Functions

In the previous discussion we assumed that the aggregate production function (2) exists, inputs $L$ and $K$ are homogeneous aggregates and, implicitly, that the parameters of the production function remain fairly constant over time. Why should we be concerned with these assumptions? The importance of the problem simply lies in the fact that without proper aggregation we cannot interpret the properties of an aggregate production function, which governs the behavior of total factor productivity.

Labor and capital are aggregates of elements that are basically heterogeneous with divergent characteristics; they differ in their longevity, impermanence, productive qualities, mobility, etc. These heterogeneity properties are the main cause and consequence of technical progress in an economy. Most of the controversy on aggregation has emerged from the problem of aggregating the different types of capital goods; however, the issues are equally applicable to the aggregation of heterogeneous units of labor and output. How should various types of capital goods be grouped? The neoclassical approach assumes a competitive economy, perfect foresight and that the quantity of capital is independent of both relative prices and the distribution of income. The necessary and sufficient conditions for grouping variables are: (a) that the rate of substitution between capital goods of different types be independent of the quantity of labor used with them, and (b) that the marginal rate of substitution between any two types of capital must be constant, i.e., the two types of capital are perfect substitutes. These conditions will ensure the malleability characteristics of capital goods of the neoclassical production function.

Joan Robinson [120, 1965], N. Kaldor [77, 1963] and others have argued that it is impossible to construct an index of the quantity of capital; capital is essentially a value concept that is affected by changes in the relative factor prices, the interest and wage rates, unless we assume a simplified economy of a one-type machine with no technical change. Also, the technological facts are such that different types of machines are complementary; therefore, they are not perfect substitutes as required by the neoclassical aggregation principles. The notion that capital is a value concept has led to the current controversy on double-switching, which suggests that if the relative price frontier is not a straight line but curvilinear, the same method of production can be the most profitable at more than one rate of profit. If this is true, capital cannot be measured in physical units as required by the neoclassical production function. Thus, the concept of “factor intensity” becomes meaningless and the neoclassical production function will not exist [68, G. Harcourt],

9 In fact, each element of capital and labor is the outcome of a series of production processes; for example, labor services are the product of a household production function with different environmental and economic characteristics [8, G. S. Becker, 1964]. It is neither possible, nor is it necessary, to know all these attributes. What is important is to identify the most important ones by using the “hedonic” or “characteristic” approach suggested by Z. Griliches [59, 1964] and K. Lancaster [58, 1965] in reference to consumer goods and to aggregate and embody them in the measures of physical inputs.

10 Condition (a), often known as Leontief’s functional separability theorem, requires “that the marginal rate of substitution between any two variables in a group shall be a function only of the variables in that group, and therefore, independent of the value of any variables in any other group.” See H. A. Green [58, 1964, pp. 2 and 93]. Condition (b) is required for the aggregate to be a simple sum of different elements in the group.
1969]. A possible alternative is to measure capital goods in terms of the labor time required to produce them. The labor time is valued at the highest wage rate consistent with the highest rate of profit.\(^{11}\)

Besides aggregation of variables into homogeneous groups, there is the problem of aggregating a number of technically different microeconomic production functions. Even with homogeneous capital goods and a neoclassical production function, F. Fisher [58, 1969] has shown that labor, capital, and output aggregation over all the production units require very stringent conditions.\(^{12}\) He demonstrated that with constant returns to scale and only two factors of production, the necessary condition for aggregation is that all capital is perfectly substitutable and all technical changes are \textit{capital-augmenting}. But if the returns to scale are not constant, capital aggregation is possible only under the restrictive assumption that the individual firm’s production function can be made to yield constant returns after suitable “stretching of the capital axis.” Similar conditions have to be specified for aggregation of labor [113, M. Nerlove, 1965].

K. Sato [125, 1969] recently extended the Solow-Fisher aggregation principle by showing that, if capital and labor are in efficiency units, the nature of technical change at the microeconomic level is preserved at the aggregate level. That is, if technical change in microeconomic production functions is purely capital-augmenting (Solow neutral), purely labor-augmenting (Harrod neutral) or \textit{equally} labor and capital-augmenting (Hicks neutral), so is technical change in the aggregate. However, a serious problem remains. The shape of the aggregate production function depends on how heterogeneous capital is distributed in efficiency units, which then suggests that the aggregate production function does not remain invariant and, therefore, cannot be used as a basis for the long-run growth theory.

A different problem arises when a third factor is introduced into a two-factor production function like (2). Unless it is a true “public good,” the third input must be allocated optimally to all vintages and cohorts for the aggregate production function to exist. Also, there is the question of complementarity between the third and the two original inputs. The possibility of output exhaustion may also arise if the inputs receive their marginal products. However, if the third factor can be “embodied” in the two conventional inputs, or if we have made the suitable adjustment in the quality of output, then the “adding up” problems will disappear [119, Nerlove, 1965].

Moreover, there is the conceptual problem of the appropriate time unit for measuring changes in the production process. How long does it take for change to be felt, a year, a quarter? Does the availability of a particular body of data dictate the time unit of measurement in estimating the parameters? Also, the practice in most empirical studies of assuming that the rate of technological change and elasticity of substitution, \(\sigma\), remain constant over a period of thirty or more years is based on fragile grounds. These issues raise the problem of aggregation over time, which has only recently received some attention.\(^{13}\)

The conclusion to be drawn from this brief discussion is that aggregation is a serious problem affecting the magnitude, the stability, and the dynamic changes of total factor productivity. We need to be cautious in interpreting the results that depend on the existence and specification of an aggregate production function. Aggregation may not be “necessarily bad”; nor is it necessarily

\(^{11}\) For further discussion of this measure, see Harcourt [68, 1969].

\(^{12}\) There is a considerable body of literature on the subject of aggregation which is beyond the scope of this paper. For recent estimation approaches and important references to previous work in the area see J. B. Edwards and G. Orcutt [42, 1969].

\(^{13}\) Some specific aggregation problems are discussed on pp. 1133–54.
good. That the use of the aggregate production function gives reasonably good estimates of factor productivity is due mainly to the narrow range of movement of aggregate data, rather than the solid foundation of the function. In fact, the aggregate production function does not have a conceptual reality of its own; it emerges as a consequence of the growth processes at various microeconomic levels and is not a causal determinant of the growth path of an economy. To understand the dynamic nature of technological change, the diffusion of new techniques from firm to firm, from industry to industry, the changing linkages among economic units through externalities, etc., it is necessary to study the disaggregates. We shall turn to a brief discussion of some of these issues below.

II.2 Some Theoretical Issues on the Nature of Technical Change

Decomposing total factor productivity into various components is not the same as explaining it. To do that requires knowledge of the dynamic factors that determine the nature of technological changes illustrated by the transitions noted in Figure 1.

There are many stages of knowledge that precede technical innovations. They range from pure principles of science to applied science to technical knowledge and then specific embodiment of technical know-how in the form of better organizational structures, new equipment, better skills, etc. Exploration of this process gives rise to important questions: What determines the stock of pure knowledge in a society? How and when does part of this knowledge take the form of innovations? Which industries are likely to initiate adoption of the new techniques? What are the characteristics of the transmission mechanism that underlie the diffusion of new technology throughout an economy? What are the external economies (dis-economies) of employing the new techniques? These are difficult questions to explore fully in this paper. We shall limit the discussion to only a few of these questions.

II.2a Autonomous and Induced Technical Change

In most of the literature on factor productivity and production functions, it is assumed that technical change is autonomous, neutral and growing at a constant rate; the purpose of these studies is mainly to determine the contribution of technical change to growth of output. This essentially Schumpeterian view of technological change purports to show that the supply of technical change is determined by the state of knowledge and autonomously supplied inventions. Promoters simply discover the commercial use of some of these techniques. Expectations of future profit determine the adoption rates of the new techniques but economic considerations do not determine the nature of the techniques themselves.

S. Ahmad [2, 1966], W. Fellner [50, 1969] and J. R. Hicks [69, 1964] have explicitly considered the effect of relative prices on the direction of technical change. The basic argument is that when entrepreneurs anticipate a relative increase in real wages, they will, in the short run, substitute capital for labor (provided the elasticity of substitution is positive) and then concentrate on innovations which are labor saving. They may do this by allocating resources to adopt the available labor-saving techniques or through research to design their own labor-saving methods.

The more important question is whether one can say a priori, that there exists an inherent labor-saving bias in technical progress itself or that the observed capital intensity is simply due to a substitution effect induced by the cheapening of capital. A labor-saving bias implies that labor scarcity stimulates a search for completely new knowledge which is labor saving or that labor-saving designs are adopted from the
existing knowledge [123, Salter, 1960, p. 43]. The former is unlikely because the firm is interested in reducing total unit costs and will not devote resources specifically to reduce particular costs such as those of capital or labor. If the latter is true, it is equivalent to the substitution effects due to change in relative prices. Salter correctly pointed out that the observed labor-saving character of modern technology is more apparent than real: the new techniques are the product of factor substitution in the first place; they also save capital, which is often underestimated due to faulty measurement and, finally, labor-saving designs are more visible!

The Kennedy-type models of "induced" technical change are formulated in terms of an "Innovation Possibility Curve." The IPC is defined as the locus of all techniques available at a given time. It is also considered exogenous in the sense that no resources are devoted to generating the new techniques per se. The main question is what factors determine the direction of the bias in the new techniques. Assume that the technology is purely factor augmenting, i.e., each input is "more input" after technical change has occurred than before. Let $\mu_1 = d\lambda_1 / \lambda_1$ and $\mu_2 = d\lambda_2 / \lambda_2$ represent the proportional reductions in the requirements per unit of output of L and K, respectively, due to the new technology. The IPC is defined as the locus of all techniques available at a given time. It is also considered exogenous in the sense that no resources are devoted to generating the new techniques per se. The main question is what factors determine the direction of the bias in the new techniques. Assume that the technology is purely factor augmenting, i.e., each input is "more input" after technical change has occurred than before. Let $\mu_1 = d\lambda_1 / \lambda_1$ and $\mu_2 = d\lambda_2 / \lambda_2$ represent the proportional reductions in the requirements per unit of output of L and K, respectively, due to the new technology. The objective function that must be maximized is:

$$r = \mu_\nu S_L + \mu_\kappa S_K$$

where $r$ is the proportionate reduction in total unit costs and $S_L$ and $S_K$ are, respectively, the relative factor shares of labor and capital. Thus the choice of technique depends upon both economic variables ($S_L$ and $S_K$) and technological considerations ($\mu_1$ and $\mu_2$). The bias in technical change depends upon whether $\mu_1 > \mu_2$, e.g., if the relative share of labor exceeds that of capital, i.e., $S_L > S_K$, then $\mu_1 > \mu_2$ implies labor-saving bias in technical progress. The stability of the equilibrium point $E$ on the IPC requires that $\sigma$ should be less than one [124, P. Samuelson, 1965].

Kennedy conducts his analysis in terms of factor shares and, by assuming that relative prices are constant, therefore concludes that "changes in relative factor prices are not essential for a theory of induced bias in in-
novations" [81, Kennedy, 1964, p. 542]. As it was mentioned on page 1148, the elasticity of substitution (σ) plays an important role in determining the direction of technical progress.\(^{16}\)

Some serious questions need to be explored about the induced technical change models. First, it is likely that technological change may not be factor augmenting. It is possible that most of the technical changes may come about by a learning process through sequencing of production and investment activities without any identifiable expenditure of resources or influence of relative prices (see discussion below on page 1162 on Arrow's learning theory [5, 1962]). Second, little is known about the stability and shape of the IPC—it will shift with different types of knowledge. Third, the bias due to technical change and the substitution effect due to changes in factor prices may not be identifiable and may offset each other. Fourth, the implicit assumption of instantaneous adoption of new techniques is unrealistic. There is considerable delay in the adoption of new techniques as indicated by learning curve studies [71, W. Z. Hirsch, 1956].

In fact, as Atkinson and Stiglitz [4, 1969] have pointed out, technical knowledge is often specific to particular production processes. Therefore, technical progress may be "localized" in one technique with minimal "spillovers" to other techniques. The productivity of the technique that is selected is further increased through learning. This discourages switching to a different technique unless current and future price expectations are such as to warrant abandoning the existing one. In contrast to the view of technical change in growth theory, history becomes very important when technical change is localized. This view of technical progress helps to explain why the bias in technical change may seem to continue in one direction and provides a rationale for the difficulties and delays in adopting completely new techniques.

II.2b The Endogenous Theories of Technical Change

It is a mistake to assume that the level of technical knowledge is exogenously determined outside the economic system. That is, the shift in the position or shape of the IPC over time may not be due solely to a "natural drift" but determined by the amount of resources (R & D expenditures) allocated to the production of new, or modification of the existing, techniques. Firms, nonprofit organizations, and government agencies devote considerable resources to produce new basic knowledge, applied techniques, and improvement of the existing designs. The empirical evidence suggests that government agencies and universities concentrate on pure scientific research while industrial research, the fastest growing segment of R & D, is mainly geared to improving the existing designs and creation of new products [110, R. Nelson, M. Peck, and E. Kalachek, 1967]. Foreign inventions are also a major source of new ideas: the individual inventors have been an important, though steadily declining, source of technological inventions.

But what is unique about the production of knowledge that makes it different from any other form of capital accumulation? What economic factors determine its behavior? Until recently little attention was given to the theoretical aspects of endogenously determined technical changes. The main distinguishing feature of technical knowledge (stressed in the works of J. Schmookler [137, 1966], Nordhaus [117, 1969], K. Shell [130, 1965], Arrow [5, 1962], E. Mansfield [95, 1968], R. Nelson [108, 1968]) is that it is highly durable, its potential impact is extremely uncertain and it is subject to large external economies. Once it is produced, its cost of transmission is almost zero, i.e., it becomes a public good very soon afterward.

\(^{16}\) For a detailed discussion see C. Ferguson [52, 1969, pp. 36, 352–54].
These features make its production very costly for a profit maximizing firm. Nordhaus tried to integrate the "indivisibility" and "inappropriability" characteristics of inventions in an otherwise neoclassical production model and shows that the amount of research undertaken by profit maximizing innovators varies positively with the life of the invention and inversely with the interest rate. The time horizon is an index of the technological lead-time. Patents give some monopoly power to the firm, or at least protect the firm from losing its "inventor's cost." The discount rate and relative prices can affect the rate and timing of technical change [91, R. Lucas, 1967; 117, W. Nordhaus, 1969]. The optimal life of an invention is likely to be very sensitive to the discount rate, the ease of invention, the elasticity of demand for the output, and the elasticity of output with respect to the amount of research.

The limited empirical results suggest a high output elasticity of R & D about .08 to .12 [99, J. Minasian, 1969; 96, E. Mansfield, 1969; 60, Z. Griliches, 1964]. Similarly, the marginal social product of R & D expenditures is considered to be more than twice its private marginal return [60, Griliches, 1964; 117, Nordhaus, 1969]. A considerable benefit of research takes the form of new products. Although there are indications that large firms account for most of R & D, large size is not a necessary condition for technical inventions [96, Mansfield, 1969]. Most industrial R & D seems to be concentrated in defense and space industries, where the government underwrites a substantial portion of the R & D expenditures.

II.2c Diffusion of Technology

Besides the production of new techniques there is the question of their transmission from one firm or industry to another. The speed and timing of diffusion depend strongly upon economic conditions and vary considerably among firms [123, W. Salter, 1960]. Factors that induce one firm to adopt new techniques faster than another are: high expected profitability, short payoff period, less concentration of the industry, low ancillary costs of adopting the new techniques and lower age of the existing capital stock of the industry. Interfirm variation of costs and prices due to variations in abilities (age and education) of the managers and labor force, monopolistic barriers such as patents, geographical location, inertia and uncertainties about the outcome of new techniques, etc., also influence the rate of adoption of the new techniques [95, Mansfield, 1968].

However, the most important impediment to diffusion of the new techniques is the existence of old capital stock and product. New equipment must compete with the existing capital stock, which depends upon the firm's obsolescence policy. In principle, any given piece of equipment should be replaced when its operating cost exceeds the total cost of new equipment. The strong complementarity among elements of the existing stock of capital goods makes it difficult to replace only part of the plant; thus the determinants of gross investment govern the speed of adoption of new techniques. Since the distribution of the existing stock of capital and the rate of accumulation differ considerably among different firms, the rate of acceptance of the new techniques will also differ greatly among them [123, W. Salter, 1960]. The size and structure of the existing capital stock also affects productivity (in-
directly) through its influence on the learning ability of the managers and workers. The adoption of new techniques is, therefore, partly dependent on the adaptability and learning capacity of these groups.\textsuperscript{18}

**II.2d Industrial Structure and Total Factor Productivity**

In addition to diffusion of technical change among the firms in a given industry, the interaction among various industries is an important source of increased productivity [97, B. Massell, 1961; 29, E. F. Denison, 1967]. In fact, increase in total productivity is accompanied by some retardation of existing industries and the appearance and growth of new ones.\textsuperscript{19} The industries are linked together by the input-output relationship which evolves over time or changes greatly when there are important technological breakthroughs or shifts in pattern of final demand.

Productivity increases in one industry are generally transmitted to other industries in the form of improved quality of materials or external economies (diseconomies). In some industries, such as the construction industry, a substantial part of the productivity increase may be due to better quality materials supplied by other industries. Unfortunately, in most industry productivity studies, material inputs are not considered.\textsuperscript{20} Another important consequence of changes in industrial composition and growth of the economy is the rapid depletion of the stock of environmental inputs such as water, or air. Different industries use different amounts of these inputs. Looked at another way this depletion of natural resources is a negative contribution to society's output. Relative prices do not take account of these negative consequences of growth.\textsuperscript{21} If they were properly taken into account, the estimated growth of factor productivity in various industries and in the total economy would be lower than that indicated by the empirical results.

To examine the "forward" and "backward" linkages among various industries and changes in patterns of consumption and output, a disaggregate interindustry approach on the production side and an inter-commodity analysis on the consumption side is needed. It would then be possible to examine the feedback between changes in consumption and productivity and assess the positive and negative contribution of technical change.

**SECTION III**

**Aggregate Production Functions and Estimation of Factor Productivity**

We now set aside the conceptual problems of Section II and consider the various attempts to estimate factor productivity using the aggregate production function. The magnitude and stability of $dA/A$ depend upon how accurately the production function is specified and estimated and the various inputs measured. In principle, if all the inputs are properly measured and the function governing their interactions is precisely specified, then the residual $dA/A$ should be zero or nearly so. In this section we consider the issues relevant to the specification and estimation of the production function. The efforts to explain technical change by adjusting the conventional inputs for quality change are discussed in Section IV.

In Section III.1 we survey the attempts to estimate the parameters of the CES production function. With all its merits the CES function is found to be subject to certain

\textsuperscript{18} See below for discussion of theory of "Learning by doing," pp. 1169–64.

\textsuperscript{19} The growing industries create demand and cost pressures that force the lagging industries to eventually disappear. It also leads to the development of new industries and accelerates the growth of complementary industries [19, A. Burns, 1954; 86, L. Lave, 1966; 133, W. Salter, 1960].

\textsuperscript{20} See below, p. 1153 for further discussion.

\textsuperscript{21} For a very stimulating discussion on spillovers and costs of growth, see E. J. Mishan's recent book [101, 1967].
important shortcomings. Some authors contend that the form of the CES is not sufficiently flexible to adequately identify the sources of factor productivity. Others have emphasized that the fault lies in the inadequacy of current estimation techniques. The outcome has been the development of more generalized production functions and indirect estimation techniques. These efforts are taken up in Section III.2.

III.1 Evidence on the CES Production Function

The two-factor CES production function has become the most widely discussed function in the literature of the last few years. It has all the properties of the neoclassical production function and includes the Cobb-Douglas and the Leontief production functions as special cases. It was derived from the empirical relation

\[ \log \left( \frac{Q}{L} \right) = a + b \log \left( \frac{w}{p} \right) + \varepsilon_0 \]

where \( w \) is the wage rate and \( p \) is the price of output; \( \varepsilon_0 \) is a random stochastic term and \( b \) is the estimate of the elasticity of substitution, \( \sigma \). This relation assumes optimization (or cost minimization) behavior, the existence of an aggregate production function with disembodied technical change, the independence of \( Q/L \) from capital intensity \( K/L \), no measurement errors in the variables, and no adjustment lag between \( Q/L \) and \( w/p \). The usual CES production function derived from (7) is

\[ Q = \gamma [\delta K^{-\gamma} + (1 - \delta) L^{-\rho}]^{-\rho/\delta} \]

where \( \gamma, \delta, \rho, \mu \), are, respectively, the parameters of efficiency (or scaling), distribution, substitution and degree of returns to scale.

According to equation (8) average labor productivity depends on capital intensity, \( K/L \) and the magnitudes of \( \gamma, \delta, \rho, \mu \) and \( \mu \). The elasticity of substitution \( \sigma = 1/(1 + \rho) \); \( K \) and \( L \) are often—but not always—measured in physical units.

The empirical evidence seems to indicate that the parameters of the CES production function are highly sensitive to slight changes in the data, measurement of variables, and methods of estimation.

The point estimates of the most important parameter, \( \sigma \), vary considerably for different sets of data, countries, industries, and levels of aggregation. Furthermore, they are sensitive to cyclical fluctuations of demand. The only tentative conclusion possible is that most of the time-series estimates of \( \sigma \) are below unity, while the cross-section estimates are generally higher than the time-series estimates and close to unity. This can be seen with a simple example. Define

\[ A = \frac{\gamma [\delta K^{-\gamma} + (1 - \delta) L^{-\rho}]^{-\rho/\delta}}{(aL + bK)} \]

Now taking the total differential of this function with respect to time and dividing by \( A \) we get

\[ \frac{\delta A}{A} = \alpha_1 \left( \frac{\delta \gamma}{\gamma} \right) + \alpha_2 \left( \frac{\delta \delta}{\delta} \right) + \alpha_3 \left( \frac{\delta \rho}{\rho} \right) \]

\[ + \alpha_4 \left( \frac{\delta \mu}{\mu} \right) \]

The first four terms on the right-hand side refer to the changes in the total factor productivity due to technical characteristics of the production function while the last two terms indicate those due to changes in the magnitudes of the factors of production [15, M. Brown 1966, p. 101].

Nerlove [114, 1967] has summarized the evidence on the time-series and cross-section estimates of the CES production function for the postwar period, ending in 1965. Generally the time-series estimates of \( \sigma \) are less than unity. Recent papers reporting various estimates of \( \sigma \) which are distinctly less than unity include V. K. Chetty [31, 1960], J. Moroney [105, 1967], Nadiri-Rosen [106, 1969], and others. Christensen-Jorgenson [23, 1969] estimated the average elasticity of substitution between \( K \) and \( L \) as less than one for the period 1929-67, but varying considerably between the pr war and postwar periods; they also found it to be very sensitive to changes in the measurement techniques. Estimates of \( \sigma \)
on the estimates of other parameters, $\delta$, $\gamma$, and $\mu$, is also mixed. $\gamma$ varies widely, depending upon the period of fit and assumptions about the type of embodiment and returns to scale. The estimates of $\mu$ are generally greater than unity in time-series and about unity in cross-section estimates, but are sensitive to the rates of utilization of the inputs and the level of demand [143, L. Thurow and L. Taylor 1966].

Several problems seem to be responsible for the instability and inconsistency of the estimated parameters of this function. An important factor in these problems certainly is of statistical nature, i.e., the data and construction of the variables vary considerably from study to study. However, three specific problems deserve attention: (a) the basic difference between the time-series and cross-section input-output relations; (b) the parameters of the production function often vary together, and their separate effects cannot be identified except under restrictive conditions and unless more information about the production process is available; and (c) estimation problems due to the simultaneity and nonlinearity between the production function and marginal productivity conditions. We shall briefly discuss these three topics below:

(a) Differences in Time-Series and Cross-Section Estimates. The time-series data actually samples a dynamic adjustment process due to a combination of factors such as changes in relative prices, technical change and external shocks, which are generally excluded in cross-section data. The time-series estimates are often biased because of simultaneity between the inputs and their prices, and misspecification of adjustment lags between the inputs and output, and the dominance of cyclical conditions—like under-utilization of capacity.

The cross-section results are also plagued by certain conceptual and estimation problems. In a competitive market there is no reason for relative prices to differ among production units. Any observed difference in inputs is a measure of differences in intrafirm managerial ability and consequently the individual production function is not identified [105, Y. Mundlak, 1964 and 146, A. Walters, 1963]. If there is insufficient variation in the marginal productivity conditions, or if the input differentials are due to differences in skill or in the quality of the inputs, then cross-section estimates of $\sigma$ will be biased towards unity [88, W. Leontief, 1964]. The cross-section estimates also ignore the temporal development of the economy—such as changes in the economy’s general structure, which is also an “input” [138, G. Stigler, 1961]. The usual procedure of choosing years of near full-employment for estimation purposes may bias the estimates upward. The variability of output price and of quality of labor often disregarded in cross-section (regional), studies, tends to bias $\sigma$ toward unity. Further, the estimate of constant returns to scale found in most cross-section studies may reflect external economies related to industry size [147, A. Walters, 1963].

There are several sources of bias common to both time-series and cross-section studies of $\sigma$. (1) The assumption underlying the marginal

---

for different industries vary considerably, but most are less than unity.

The cross-section results vary among industries but are close to unity as reported by Nerlove [114, 1967]. P. Zarembka [163, 1970] recently reported estimates of $\sigma$ for two-digit industries for 1957 and 1958 that are not significantly different from unity. Z. Griliches’ results for agriculture and manufacturing industries [61, 1967 and 62, 1968] confirm the same results.

23 For point estimates of $\gamma$, $\delta$ and $\mu$, see Nerlove 114, 1967]. These estimates are shown to vary from period to period and are generally sensitive to small changes in $\sigma$. Recent work by R. Bodkin and L. Klein [18, 1967], P. Dhrymes [82, 1965], M. Brown and J. S. de Cani [17, 1968], Brown [15, 1966], Brown and J. Popkin [18, 1962], R. K. Diwan [97, 1966], G. Sahota [122, 1966], Thurow-Taylor [143, 1968] and F. Westfield [160, 1966], suggested increasing returns for the U. S. economy during 1929–55. The cross-section estimates of $\mu$, by Griliches [61, 1967] and, recently, by Zarembka [183, 1970], support the hypothesis of $\mu=1$ or close to unity.
productivity condition refers to the relation of the "best practice" factor proportions to input prices, while the data used in estimation refer to the "average practice" factor proportion. The extent of bias depends on the departure of the "average practice" from "best practice" factor proportions [12, R. Boddy, 1967].

(2) The functions estimated using aggregate data are often "hybrids"—for example, a combination of Cobb-Douglas and CES functions. The parameter estimates of these aggregate production functions depend critically upon the relationships among the "interclass" and "intracl ass" elasticities of substitution [11, T. Black, 1969 and 126, K. Sato, 1967]. The constancy of the inter- and intraclass elasticities of substitution does not necessarily lead to constancy of partial elasticities. Thus the stability of the parameters of the aggregate production function depends on the manner in which various inputs or stages of production are aggregated.

(3) Restricting the analysis to only two factors is another source of bias. In most industry studies value-added data are used as a measure of output on the assumption that the ratio of raw materials to total output remains constant [41, E. D. Domar, 1963].

The evidence, however, suggests that this ratio is not constant for the whole economy nor for the various industries. Generally, it has declined due to improvements in technology, better inventory management, and substitution of both raw materials and primary inputs. The omission of materials from the production relation often leads to a positive bias in estimates of returns to scale and affects the elasticity of substitution between \( K \) and \( L \).27

(b) Identification Problems. Often there are not enough degrees of freedom in the available data, especially if it is time-series, to identify and isolate the separate effects of the parameters of the production function. Aggregate data vary in such a narrow range that it is very difficult to distinguish one production function from another.

Various factors are, at least in part, responsible for identification problems. The aggregate data surely include both old and new capital. It is seldom possible to distinguish between embodied and disembodied technical change as long as old capital remains productive and new investment is being undertaken [75, D. Jorgenson, 1966]. The existing capital stock may represent disembodied technological change, while the deflator of new investment may be considered as the index of quality improvement.

As R. E. Hall [66, 1968] pointed out, if the rate of embodied technical change is a constant function of vintage, the rate of disembodied technological change a constant function of time, and the depreciation rate a function of the age of capital, then these effects cannot be empirically identified [65, T. Haavelmo, 1944]. There are two variables, time and vintage, and three parameters to identify.28

27 There are other variables which affect total factor productivity but are often excluded from the production relation. For example, government expenditure on research and development and the use of government financed capital goods contributes to the growth of output in the private sector [55, R. J. Gordon, 1969]. Omission of these expenditures understates the true growth rate of capital, and overstates the rate of growth of factor productivity. Another excluded variable is the stock of liquid assets; the neoclassical production function is stated in physical terms and financial assets have no place in it. The question arises as to whether or not any productivity can be attributed to liquid assets. If not, why are they held?

28 Hall [66, 1968] has pointed out that identification
Other reasons for the identification problem are the effect of variations in factor supply upon relative factor prices [152, Wilkinson, 1968] and the variability of the rate of technical change over time [114, Nerlove, 1967]. Further, the interaction of bias in technical change with the other parameters of the production function, such as $\sigma$, nonneutral scale effect, etc., exacerbates the problem.\(^{29}\)

Not much progress has been made so far in satisfactorily untangling the separate effects of various parameters of the production function. Beckmann and Sato [9, 1969] attempted to delineate various types of technical change by studying the relationship of the factor-price ratio with variables such as $Q/K$, $Q/L$ or $L/K$. But their empirical results were inconclusive and somewhat misleading.\(^{30}\) There is still a need for more information on best-practice techniques and for adequate specification of the dynamic nature of the production process in given industries. Imposing identification restrictions on the data in order to estimate the production functions for higher aggregates is neither a very interesting nor rewarding venture.

(c) Estimation Problems. In estimating the parameters of a production function it is possible to fit either the function directly or the marginal productivity conditions. These two sets of relations are, however, interdependent, because of the joint distribution of the stochastic error terms of the production function and marginal productivity relations. Consider the following production function and marginal productivity relations

\[
Q = F(K, L)\mu_0; \quad \frac{\partial f}{\partial K} = \left( \frac{q}{P} \right) \mu_1; \\
\quad \frac{\partial F}{\partial L} = \left( \frac{w}{P} \right) \mu_2
\]

where $\mu_0$, $\mu_1$ and $\mu_2$ are random variables. These terms reflect uncertainties in the production process and about relative input prices, managerial inertia, etc. To estimate the parameters, the joint distribution of $\mu_0$, $\mu_1$ and $\mu_2$ must be considered. Any misspecification will lead to erroneous estimates of the contributions $K$ and $L$, which in turn affects total factor productivity ($dA/A$).

Zellner, et al. [154, 1966] have shown that if (i) entrepreneurs maximize the mathematical expectation of profit, (ii) prices of product and factors are either known with certainty or statistically independent of $\mu_0$, and (iii) the expected values $E(\mu_{01}) = E(\mu_{02}) = 0$, then simple least squares estimates will, at least, be consistent. Moreover, under normality, or with the stronger assumption that $\mu_1$ and $\mu_2$ are independent of $\mu_0$, they will also be unbiased. Dorothy Hodges [72, 1969] proved the same conclusions for the CES production function except that the CES can be estimated by either of two single equation procedures, viz. nonlinear least squares or Kmenta's application of least squares to a linear approximation of the CES (which requires that $\mu_1$ and $\mu_2$ are independent of $\mu_0$).

The most noteworthy estimation methods recently proposed are

(1) Kmenta's approach is to use the least squares technique directly to estimate the equation

\[
\log Q = \log \gamma + \mu d \log (K/L) \\
+ \mu \log L + \beta \left[ \log \left( \frac{K}{L} \right) \right]^2 + \mu_0'
\]
where $\beta = \rho \delta$, $1 - \delta / 2$. This indicates a departure from the Cobb-Douglas function (where $\rho = 0$). The higher order terms of capital intensity $K/L$ are included in $\mu_\delta$ but have an insignificant effect on Kneta's numerical results for moderate values of $K/L$.

(2) The more common approach to estimation of the CES production function is the stepwise procedure of estimating, first, the ratio of the marginal productivity relations to obtain estimates of the $\rho$ and $\delta$ in equation (8); then using these estimates ($\hat{\rho}$ and $\hat{\delta}$), to estimate the two remaining parameters of the CES, i.e., $\mu$ and $\gamma$. Unfortunately, this procedure has certain shortcomings. It rests on the assumption that marginal productivity conditions hold, which in turn requires that the returns to scale parameter is unity. Furthermore, it assumes that the responses of $K$ and $L$ to changes in relative prices, $w/g$ are identical, that the marginal productivity conditions hold instantaneously, and that the current relative prices are exogenous and good proxies for their expected values.

Kneta's method is consistent not only with the CES function, but also with a quite general class of production functions [92, M. McCarthy, 1965]. Limited use has been made of this procedure [61, Griliches, 1967 and 139, Zarembka, 1970].


The first step is to fit equation

$$
\log \left( \frac{K}{L} \right) = a_0 + b \log \left( \frac{w}{g} \right) + e_0
$$

where $w$ and $g$ are, respectively, the price of labor and capital inputs; $a_0 = \log (\alpha / \beta)$ and $b = 1$ if the production function is Cobb-Douglas $Q = AL^\alpha K^\beta$, $a_0 = \log (\beta / 1 - \delta)$ and $b = 1 + \rho$ if the production function is CES.

The second step is to use either estimates $\hat{\alpha} / \hat{\beta}$ or $\hat{\delta} / (1 - \hat{\delta})$ and $\hat{\rho}$ to estimate the constrained equations:

$$
\log Q = \log A + \alpha \log L + (\hat{\alpha} / \hat{\beta}) \log K + e_1
$$

if the production function is Cobb-Douglas, and

$$
\log Q = \log \gamma + \frac{\mu}{\beta} \log (\hat{\delta} K^{-\hat{\rho}} + (1 - \hat{\delta})L^{-\hat{\gamma}}) + e_2
$$

if the production function is CES.

It has also been found that the estimates of the parameters of the production function are sensitive to the classification of the variables into exogenous and endogenous groups. The usual method of regressing $\log (Q/L)$ on $\log (w/g)$ biases $\sigma$ toward zero, while reversing the procedure gives much better estimates of $\sigma$.

(3) As an improvement over the preceding method, Bodkin and Klein [18, 1967] propose a nonlinear maximum likelihood procedure. They explicitly take into account the joint distribution of the stochastic error terms, and the nonlinearity of the production and marginal productivity conditions. The likelihood function is based on the arguments and parameters of the production function and the variances and covariances of the stochastic terms. Assuming certain initial values for the parameters, the likelihood function is solved iteratively until the lowest sum of squared errors is obtained and the parameter estimates converge on a particular set of values. Their estimation method is, however, very sensitive to the choice of initial conditions.

(4) A relatively recent development is the Bayesian estimation technique, which combines knowledge about the prior distribution of the parameters of the production function and its stochastic error term, using a likelihood function similar to that employed by Bodkin and Klein. The Bayesian approach

and Brown and A. Conrad [16, 1967], have used a distributed lag estimate of $(w/g)$ as a proxy for expected factor prices.

Maddala and J. Kadane [94, 1967], in their Monte Carlo study, examined the effect of the choice of exogenous variables on the estimation of the CES production function. Fixing the values of $\gamma$ and $\delta$ to arbitrary constants and making various assumptions about the "economic" and "technical" disturbances, they examined the bias in estimating $\sigma$ due to an "improper" choice of the independent variables.

Zellner, et al., [134, 1966] have formulated this approach for the Cobb-Douglas function and V. K. Chetty [91, 1969] has applied it to the CES function. The priors chosen by Chetty are $0 < \delta < 1$: $0 < \gamma < \infty$; $0 < \mu < \alpha$ (a is a constant), and a complicated prior on $\rho$, such that $\rho$ is invariant with respect to measurement.
does not use any information about maximization behavior or about the market conditions. It directly estimates the parameters of an average production function, and not the efficient combination of inputs suggested by economic theory. The estimates obtained by this technique are also sensitive to the choice of priors.

(5) Ideally we should be interested in estimating, not the average production function, but the best practice production frontier, since this is the locus of the most efficient techniques available. Differences arise among firms due to differences in technical knowledge, in ability to maximize profits, and environmental conditions, i.e., factors besides relative prices [112, Nerlove, 1963]. There has been little effort besides that of M. J. Farrell [48, 1957] to isolate "technical efficiency" and "economic efficiency"—the former is related to the choice of techniques and the latter refers to efforts to minimize costs of producing the output. One of the advantages of this approach is to find the relative efficiency of various firms, or industries, and examine the contribution of the managerial inputs in selecting the best techniques and minimizing cost. Recently, some attempt was made to use linear programming techniques [3, D. Aigner and S. Chu, 1968] to estimate the best practice production function, but this effort is still in its preliminary stages.

Progress in designing better estimation techniques has been fairly good. However, it is too early to judge the efficiency of the various techniques. For one thing, they are usually applied to different bodies of data and different time periods. What is needed is a controlled experiment using the same set of data to determine the relative merits of these methods in estimating the aggregate production frontier or the average production functions. One attempt along these lines has been made by R. Bodkin and L. Klein [13, 1967]. Their results were inconclusive except to clearly establish the fact that the parameter estimates were very sensitive to different methods of estimation and specification of the production function. Further research along these lines would be of considerable value.

III.2 Formulation of New Production Functions

The instability and inconsistencies of the estimates noted earlier may be partly due to the restrictive nature of the CES function to permit interaction of the various factors which contribute to growth of productivity. The attempts to remove the restrictive features of the CES production function have taken these forms: (a) amendment of the standard two-factor CES function; (b) indirect estimation of the parameters of multi-factor production functions by formulating first, the relevant cost functions; (c) specification of intertemporal production models which account explicitly for the costs of adjusting the level of inputs. These approaches are briefly stated below.

III.2a Amendments to the CES Production Function

It has been increasingly clear that the marginal productivity condition (7) is not independent of the capital-labor ratio and the returns to scale. Estimates of \( \sigma \) derived from the marginal productivity conditions of labor and capital are found to differ significantly [32, P. Dhrymes, 1965; 35, Dhrymes-Zarembka, 1970; and 70, Hildebrand-Liu, 1965]. This has led to the development of the Variable Elasticity of Substitution.
(VES) production function. The procedure is to test the sensitivity of values of \( \sigma \) to changes in capital intensity, skill-mix, and returns to scale or time. When the value of \( \sigma \) does vary with respect to these variables, the relation (7) does not hold and the usual CES production function is no longer valid. For example, the invariance of \( \sigma \) to capital intensity is tested by fitting the relation

\[
\log \left( \frac{Q}{L} \right) = \log a + b \log \left( \frac{w}{p} \right) + c \log \left( \frac{K}{L} \right) + \epsilon_1
\]

(11)

where \( \epsilon_1 \) is the random error term. If \( c \neq 0 \), the labor productivity condition (11) implies the following production function

\[
y = \left[ \beta x^{-\rho} + \alpha x^{-m} \right]^{-1/\rho}
\]

(12)

where

\[
y = \frac{Q}{L}, \quad x = \frac{K}{L}
\]

\[
\rho = \frac{1 - b}{b}
\]

\[
m = \frac{c}{1 - b}
\]

\[
\alpha = \frac{1 - b}{(1 - b - c)\beta^{1/b}}
\]

\[
\beta = \frac{-d(1 - b)}{b\alpha^{1/b}}
\]

and

\[
d \text{ is the constant of integration.}
\]

The elasticity of substitution for this function is

\[
\tilde{\sigma} = \frac{1}{1 + \frac{\rho m \rho}{S_K}}
\]

where \( S_K \) is the share of capital. Relation (12) is a labor-embodied homogeneous CES production function in which labor productivity depends upon capital intensity [90, Y. Lu and L. Fletcher, 1968]; it includes the Leontief, Cobb-Douglas, and the CES production function (8) as special cases. Since \( m \) and \( S_K \) are positive, the relationship between \( \tilde{\sigma} \) and \( \sigma \) depends upon the magnitude of \( \rho \); that is, if \( \rho \leq 0 (\sigma \leq 1) \) and \( \tilde{\sigma} \geq \sigma \). Thus, when \( \sigma < 1 \), \( \tilde{\sigma} \) is larger than \( \sigma \) and tends toward unity and if \( \sigma > 1 \), \( \tilde{\sigma} \) is smaller than \( \sigma \) and tends toward unity, again suggesting an equilibrating process [38, R. K. Diwan, 1970].

The technical bias in the Hicksian sense can be shown to be

\[
g = 1 + \frac{m \rho (S_K - S_L)}{m \rho S_L - (1 + \rho) S_K}
\]

where \( S_L \) is the share of labor. Technical progress will be neutral when \( g = 1 \), which is possible if \( m = 0 \), \( \rho = 0 \) or \( S_L = S_K \). When \( m = 0 \), (12) reduces to the CES function (8) and the labor productivity does not depend upon \( K/L \); when \( \rho = 0 \), the production function is Cobb-Douglas and \( \tilde{\sigma} = \sigma = 1 \). Finally, when factor shares are equal and if \( dK/K > dL/L \) then the growth rate of the marginal productivity of labor has to exceed that of capital in order to establish the equilibrium. In such a case the technology is capital-using. There is evidence that in the United States the share of labor has grown faster than that of capital and capital input has grown faster than labor in the postwar period: suggesting that \( g < 1 \), that is, labor-saving technical change.

The empirical results suggest that \( c \neq 0 \), and \( \tilde{\sigma} > \sigma \) [38, Diwan, 1970 and 114, Nerlove, 1967]. Also, for the U.S. manufacturing industry, Diwan [38, 1970] and G. H. Hildebrand and Liu [70, 1965] showed that generally \( m \neq 1 \) and \( g < 1 \) for most industries. Diwan [38, 1970], using firm data, shows that \( \tilde{\sigma}, m \) and \( g \) vary as the size of the firm increases, reflecting, perhaps, the presence of economies of scale.

A further amendment has been proposed by M. Brown and A. Conrad [16, 1967], making all the parameters of the CES production function (8) dependent upon educa-
tion and research expenditures. The empirical results of their model show significant effects for both expenditures. Finally, A. J. Zellner and N. Revankar [155, 1969] recently developed a generalized production function such that for a given elasticity of substitution, the returns to scale parameter \( \mu \) varies with the rate of output. This function includes the CES and the Cobb-Douglas forms as special cases, but they have estimated empirically only the Cobb-Douglas version of their generalized function. Their results indicate considerable variation in returns to scale with the level of output.\(^{37}\)

On the whole the empirical evidence on the validity of the VES functions is still very limited. There is little scope for distinguishing the VES from other forms of the production function using aggregate data. Microeconomic or industry data are more suitable for testing such relatively complicated functions.

### III.2b New Functional Forms Using Duality Theorems

Another limitation of the CES functions is that it has been confined to two inputs, and attempts to include more than two factors have led to certain problems. As proved by D. McFadden [93, 1963] and H. Uzawa [145, 1964], when three or more inputs are included, strict assumptions about the partial elasticities of substitution, \( \sigma_{ij} \), are required for estimation purposes. That is, all the pairs of partial elasticities of the different classes of inputs must either have the same constant value or should be unity for all subsets of the inputs. Dhrymes and Kurz [34, 1964], W. M. Gorman [56, 1965] and V. Mukerji [104, 1963] have developed \( n \)-factor generalized functions which yield constant ratios of partial elasticities of substitution among all the inputs.\(^{38}\)

Recently attempts have been made to formulate and estimate new types of aggregate production functions by using the duality relation between cost and production functions. The motivation behind these efforts is the problem that the second-order parameters of the production function such as the elasticity of substitution can be estimated with reasonable efficiency from second-order data only. The approach has certain convenient features. Cost functions are easier to formulate and estimate than generalized production functions noted earlier.\(^{39}\)

They permit arbitrary degrees of substitutability between pairs of factors in an \( n \)-factor production process. Finally, the link between short- and long-run cost relationships can be used to integrate short- and long-run production functions or the ex ante and ex post production functions. These features are obviously very important for determining the behavior of factor productivity.

The basic idea is to formulate a long-run cost function \( C_L = f(w_{it}, S, Q_t) \); where \( w_{it} \) is the vector of actual input prices, \( S \) is a vector of inputs chosen \( \alpha \) priori to minimize costs over the planning horizon on the basis of future expected prices and is, therefore,

\[ Q^* = \sum E_i X_i \rho_i \]

where \( X_i \) are the various inputs, \( \rho \) is the elasticity parameter, \( E_i \) are the efficiency parameters and \( \rho_i \) are the elasticity parameters with respect to each input. The ratios of the partial elasticities of substitution are constant in this model, but the function is nonhomothetic and thus the expansion paths have some curvature. The estimates of the elasticities are, therefore, burdened to explain the curvatures of both the isoquants and the expansion paths. Gorman [56, 1965] has gone a step further and developed a general class of production functions which is homothetic, yields constant ratios of partial elasticities and includes (8) as a special case.

\(^{38}\) The Dhrymes-Kurz, Mukerji function has the form where \( X_i \) are the various inputs, \( \rho \) is the elasticity parameter, \( E_i \) are the efficiency parameters and \( \rho_i \) are the elasticity parameters with respect to each input. The ratios of the partial elasticities of substitution are constant in this model, but the function is nonhomothetic and thus the expansion paths have some curvature. The estimates of the elasticities are, therefore, burdened to explain the curvatures of both the isoquants and the expansion paths. Gorman [56, 1965] has gone a step further and developed a general class of production functions which is homothetic, yields constant ratios of partial elasticities and includes (8) as a special case.

\(^{39}\) Nerlove [112, 1968] in a well-known article recognized the merits of the cost function approach and some of his ideas have been extended recently by other writers.
unaffected by short-run changes in factor prices. In the second step, the ex post short-run cost is minimized subject to the constraint of a given $S$. The elasticity of substitution among inputs will generally be higher ex ante than ex post, and the ex ante technology will trace the envelope of the ex post techniques [54, A. M. Fuss, 1969]. The functional form for the cost function being given, the underlying production function can be identified. W. E. Diewet [36, 1969] used a second-order approximation to an arbitrary cost function of the form

$$C_L = \sum_{i=1}^{n} a_i w_i h_i(Q_i)$$

(13)

$$+ \sum_{i} \sum_{j} a_{ij}(w_i w_{ij})^{\ell/2} h(Q_i)$$

where the parameters $a_i$ and $a_{ij}$ are functions of the ex ante choice of techniques, $S$, and $h_i(Q_i)$ are functions of output that specify the ex post output expansion. The ex post production function corresponding to (13) is shown to be nonhomothetic (unless $h_i(Q_i) = h_j(Q_i) = h(Q_i)$ for all $i, j$). Also, depending upon the sign of $a_{ij}$, the factors of production are substitutes, complements, or independent (i.e., $a_{ij} > 0$, $a_{ij} < 0$ or $a_{ij} = 0$). The partial elasticities of substitution vary over time and between pairs of inputs.

Problems arise in the econometric estimation of (13). There are too many parameters to estimate and the aggregate data are too highly collinear and of too poor a quality to permit sensible estimation of such complicated functions. G. Hanoch [67, 1969–70], however, has developed a “Constant Difference Elasticities of Substitution” (CDE) function which is a multifactor production function, requiring only a limited amount of data on inputs and factor prices and can be easily estimated.\(^4\) Hanoch’s approach is an important step forward and needs to be further developed and tested empirically.

III.2c Intertemporal Production Functions and Cost of Adjustment

The integration of the short- and long-run production decisions has also been explored recently within the context of dynamic production functions emphasizing the cost of adjustment in moving from one level of production to the next. This approach is essentially an extension and generalization of the work of R. Eisner and R. Strotz [45, 1963] in the theory of investment and Nerlove’s [114, 1967] on the theory of production. Firms are assumed to simultaneously combine the existing levels of input for current production and make preparations for the next period’s production.\(^4\) This approach amounts to maximizing the intertemporal profit function

$$R = \int_{0}^{\infty} \left[ P(X(t)) - C(X(t)) - B(dX(t)) e^{-rt} dt \right]$$

where $R$ is the present value of net receipts,

$$\frac{1 - \mu}{\mu} (b_i - b_l) \log Q$$

where $X_1$ and $X_i$ ($i = 2, \ldots, n$) are the various inputs, $P$ is the price of output, $w_i$ are the prices of $X_i$ and $w_i$ is the price of $X$, $\mu$ is the degree of homogeneity of the production function, $b_l = 1/(1 - \rho_1)$ and $b_i = 1/(1 - \rho_i)$ are, respectively, the elasticity of substitution of $X_1$ and its cross elasticities with other inputs. If $b_l = b_i$ we get the marginal productivity condition for the standard CES function. Note that if all $\rho_i < 1$, i.e., all $\alpha_i > 0$, then all factor pairs are substitutes and if only $\rho_i > 1$ and $\rho_2, \ldots, \rho_n < 1$ then $X_1$ is a substitute to other factors but $X_2, \ldots, X_n$ are a group of complements. The standard simultaneous equation estimation techniques can be used to estimate these equations but appropriate data on factor prices and better grouping of inputs are needed.

$X$ is the vector of inputs; $F(X(t))$ and $C(X(t))$ are, respectively, the production and cost functions; $B(dX(t))$ is an adjustment cost function that depends upon the rate of change of the inputs, and $r$ is the discount rate.\textsuperscript{42} In these models there is no sharp distinction between fixed and variable inputs. The inputs differ only in their cost of adjustment, \textit{i.e.}, their speed of adjustment. The parameters of the underlying production function can be indirectly obtained by estimating a set of interrelated demand functions for inputs. This indirect estimation, as forcefully pointed out by Nerlove [115, 1967], is preferable to the direct method. It permits inclusion of the utilization rates of factors explicitly and a separate distributed lag function of each input. This is crucial for revealing the dynamics of changes in factor productivity arising in the production process [33, Dhrymes, 1967].

Very little empirical work on these models is available. One study on the subject [106, Nadiri and Rosen, 1969] shows that in the short run the rate of utilization of capital stock and hours worked adjust in response to a unit change in output, giving time, first for labor, and then for capital to catch up. It is the omission of the rate of utilization of capital in most of the short-run production studies that is probably responsible for the high return to labor input in the short run.\textsuperscript{43}

\textsuperscript{42} The maximization conditions of (14) are given by the differential equation $F_t(X) - C_t(X) + rB_t(dX) - B_t(dX(t)) = 0$. Given the initial conditions of the adjustment function and making some assumptions about the shape of this function (usually assumed to be quadratic) and price expectations, it is possible to obtain a series of interrelated demand functions for the inputs.

\textsuperscript{43} Nadiri-Rosen [106, 1969] estimated a model like (14) using data for the U.S. total manufacturing. By assuming capital stock and rate of utilization as fixed, the output elasticities of labor and hours worked were, respectively, 1.36 and .12. This suggests substantial decreasing returns to man-hours (.68), in contrast to the usual estimates that range from 1.3 to 2.0 [14, F. Brechling, 1965 and 84, E. Kuh, 1965]. When capital services were allowed to vary, holding only the capital stock fixed, estimates of return to scale became greater than one.

These types of models are related to and are consistent with the progress functions and demand-oriented theories of technical change noted earlier. The firm may set aside $B(dX(t))$ of its resources for research and development purposes in the hope of reducing the uncertainty about its future plans. The initial value of $B(dX(t))$ for a particular type of technique may be very high in the beginning and decline as the process of creating new knowledge by further research and development becomes an integral part of the firm’s activity. If new techniques emerge, as Arrow [5, 1962] has argued, by a sequential pattern of investment or production, then $B(dX(t))$ would be negligible.\textsuperscript{44}

To summarize, an encouraging start towards formulating more generalized production functions and indirect estimation methods is currently under way. These attempts may lead to breakthroughs in the theory and estimation of the aggregate production function. Unfortunately, there is so far only scanty evidence on the empirical performance of the new functions.

However, it should be clearly recognized that the aggregate data are essentially of a first order kind, good for estimating first order parameters only, \textit{viz.} the efficiency and scale parameters. Until better data are available it is unlikely that the use of sophisticated aggregate production functions fixed, estimates of return to scale became greater than one.

There is a substantial body of literature on short-term productivity changes and employment functions which is left out of this survey. The interested reader should consult the works of Brechling [14, 1965], R. C. Fair [47, 1969], T. Hultgren [73, 1965] and E. Kuh [84, 1965].

These models are consistent with and explain Leibenstein’s “X-efficiency” hypothesis [87, 1969], where purchased inputs are not fully utilized because of the prevalence of the “sub-optimal disequilibrium” or “inert areas” in a firm or an economy. This under-utilization of resources is due to managerial inertia which can be removed by training the management and by reducing uncertainties about future demand. These efforts can be considered as adjustment costs.
will lead to much progress in understanding the dynamic forces which affect total factor productivity. The instability—of the parameters noted earlier for the CES function—will remain unresolved. What is needed is to supplement this approach with dynamic input-output type models. This should be a step towards loosening the present growth-theory orientation of production function studies and may eventually provide the basis for more realistic growth models.

SECTION IV
Reassessing the Role of Conventional Inputs

In addition to specifying the form of the aggregate production function, it is necessary to consider the characteristics of inputs. As mentioned earlier (pages 1140–41), if the inputs are not adjusted properly for quality changes, part or all of total factor productivity would be due to growth of the quality indices. Attempts to link technical change with the conventional inputs have varied substantially among writers. Some have argued for the embodiment of all (or most) of technical change in capital goods, while others have favored labor as an important "carrier" of technical change. Still others have reasoned for a more flexible growth accounting procedure.45

IV.1 Capital Embodiment Approach

The link between $K$ and technical change can be established by the vintage models developed by L. Johansen [74, 1959] and, subsequently, by Kaldor and Mirrlees [77, 1962] and Solow [133, 1957]. According to these models, the production function becomes a "choice of design function," i.e., new technology is embodied in new capital goods. Once a machine is constructed, factor proportions determine output per unit of labor, since each machine requires a fixed amount of labor, i.e., there is no possibility of ex post substitution between capital and labor.46 Assuming that the production function shifts upward over time, then output per worker using the new capital depends upon the design of the new machine and the vintage effect associated with time.47

Solow developed a capital vintage model assuming that: (a) machines of the same vintage are identical, but different from those of other vintages; (b) machines of the latest vintage are more productive than those of the preceding vintage by a constant exponential factor; (c) the depreciation rate is constant and uniform for all machines irrespective of their vintage; (d) the marginal rate of substitution between machines of different vintages is independent of other inputs; (e) the production function is Cobb-Douglas with constant returns to scale; and (f) the marginal product of labor in all uses, i.e., over all vintages of capital goods, is equal. These assumptions ensure the existence of an aggregate production function and enables Solow to define:

$$Q(t) = \int_{-\infty}^{t} Q_v(t) dv$$

45 As stated on pp. 1158–59, efforts have been made to allow for both ex ante and ex post substitution of $L$ and $K$.

46 The choice-of-design function requires the investors to decide on the savings rate and the lifetime of real capital, and to choose the factor proportion. As pointed out by E. Phelps [119, 1962], the ex post fixed proportions assumption for existing capital does not usually affect the long-run growth but does affect the short-run behavior of output per worker. The equilibrium mean age of the capital stock will be affected by the investment ratio if $\sigma \neq 1$. If $\sigma < 1$, for example, an increase in the investment ratio will lower the equilibrium level of the capital stock. In the short run, if workers are employed in a combination of the available capital goods such as to maximize output, then output becomes a function of the number of workers. A short-run production function can be formulated with a decreasing slope due to the vintage effect of the capital goods. The implication of such a production function is that output per man should increase with the rate of unemployment, which is contradictory to empirical results. However, if labor hoarding is considered, the contradiction disappears.
\[ L(t) = \int_{-\infty}^{t} \lambda(t) \, dv \]
\[ J(t) = \int_{-\infty}^{t} e^{\lambda v} K_{v}(t) \, dv \]
\[ Q(t) = A e^{\gamma t} L(t) \gamma J(t)^{\beta}, \quad \alpha + \beta = 1 \]

where \( Q_{v}(t), L_{v}(t), \) and \( K_{v}(t) \) are, respectively, the output, labor, and capital of vintage \( v; J(t) \) is capital in efficiency units and \( K_{v}(t) \) is defined as

\[ K_{v}(t) = e^{-\delta(t-v)} I_{v}. \]

After rearrangement we obtain

\[ Q(t) = A e^{\Psi_{2} t} L(t)^{\alpha} \left[ \int_{-\infty}^{t} e^{\Psi_{1} v} I(v) \, dv \right]^{\beta} \]

where

\[ \Psi_{2} = \delta + \lambda \quad \text{and} \quad \Psi_{1} = \gamma - \delta \beta \]

\( I(v) \) is the gross investment in year \( v, \lambda \) is the rate of embodied technical change, \( \delta \) is the rate of depreciation, \( \alpha \) is the share of labor, and \( \gamma \) is the rate of disembodied technical change. If \( \gamma = 0 \), all technical change is embodied in capital; otherwise it is a combination of capital-embodied and disembodied types.48

There are two problems inherent in this procedure. First is the identification of the parameters of the production function, \( \alpha, \gamma, \lambda, \) and \( \delta \). The problem can be solved if some a priori knowledge of parameters \( \gamma, \lambda, \) and \( \delta \) is available. The usual method of estimating \( \lambda \) has been the iterative procedure, assuming trial values for \( \alpha, \delta \) and \( \gamma \). However, the estimates for \( \lambda \) using aggregate data have been too sensitive to values chosen for \( \alpha, \delta \) and \( \gamma \) and the method of estimation [10, E. Berglas, 1965 and 134, Solow, 1960].49 Second

is the restrictive nature of the underlying assumptions of the model, such as constant equal rates of depreciation for different types of capital, absence of any effect of the utilization rate on \( \delta \), no allowance for ex post substitution between \( K \) and \( L \) and, finally, no allowance for any interaction between new investment and existing capital.46 Nevertheless, there have been some empirical studies [20, A. Carter, 1963; 34, P. Dhrymes and M. Kurz, 1964; 57, M. Gort and R. Boddy, 1967; 82, R. Komiya, 1962; and 128, W. Salter, 1960] that indicate higher productivity for the new capital (best-practice technique) than old capital (average-practice technique). But the embodiment hypotheses is not yet completely established due to the sensitivity of the estimates to methods of estimation and nature of the data [151, M. L. Wichens, 1970].

The linkage between technical change and the investment history of an economy is emphasized by Arrow [5, 1962], Kaldor and Mirrlees [77, 1962] and W. Salter [128, 1960]. Arrow, like Solow, postulates a vintage model where technological progress is embodied in the increasingly productive capital goods, since the labor used to produce them becomes more efficient over time. Acquired knowledge is irreversible and a

would be capital-saving if \( \sigma < 1 \), labor-saving if \( \sigma > 1 \), and neutral if \( \sigma = 1 \). Of course, in the last case, a net capital stock series, which includes an obsolescent rate, generates the same rate of technological change as Solow's embodied capital series, \( dA/A \) [5, M. Brown, 1956]. The assumption of the constant return to scale can be modified to allow for economies of scale [184, Solow, 1959 and 150, F. Westfield, 1966] by defining

\[ J(t) = \int_{-\infty}^{t} I(v)^{\alpha/(1-\alpha)} e^{(\lambda+\beta) v/(1-\alpha)} \, dv \]

where \( \alpha \) and \( \beta \) are the shares of capital and labor and \( \alpha+\beta \neq 1 \). However, construction of \( J(t) \) will not be possible unless the production function is assumed to be what Fisher calls the "capital-generalized constant return" (CGCR) function [53, 1969]; See p. 1145 of this paper.


49 Solow, of course, tried walking on both sides of the street: in 1957 [138] he assumed all technical change to be disembodied and in 1959 [134] all technical change is embodied. Probably the truth lies somewhere between the two limits.
function of all past levels of accumulation. Therefore, when the factors of production receive their private marginal products there is an increasing return due to the accumulation of knowledge, and the social rate of return on investment will exceed its private rate of return.

Kaldor proposes the technical progress function as an alternative to the production function. He rejects the marginal productivity theory of factor payment. The rate of technical progress is assumed to be exogenous and embodied in the latest machines. His model is actually a capital-embodiment model of technical change, i.e., the rate of growth of output per worker operating new equipment depends upon the rate of growth of investment per worker.\textsuperscript{81}

Both the Arrow and Kaldor models are essentially intertemporal dynamic production models. They rest on the notion of a stable progress function in the capital goods industry which relates input requirements to a cumulative index of activities, $Z$. This index can be measured in either terms of cumulative output, or of the time spent in productive activity or a combination of both. If $Z$ is measured by cumulative output, then experience is acquired by doing more regardless of the length of time it takes; when it is measured by time, experience is acquired by doing it longer regardless of the accumulation of output. Symbolically, we may state the first version of the progress function

\begin{equation}
\frac{\int_0^n Ldt}{\int_0^n Qdt} = a \left( \int_0^n Qdt \right) \omega
\end{equation}

where $Z = \int_0^n (Qdt)$; $(a)$ and $(\omega)$ are constants; $\omega$ is the rate of progress. Equation (16) suggests that cumulated input (labor) per unit of cumulated output is a declining function of cumulated output so far. Taking the derivative of (16) with respect to time, we obtain the short-run labor requirement

\begin{equation}
\frac{L}{Q} = a(1 + \omega) \left( \int_0^n Qdt \right) \omega
\end{equation}

which suggests that labor productivity rises with a larger volume of output. If we assume a constant output per unit of time $Z = \int_0^n Qdt = tQ$, then (17) takes the form

\begin{equation}
\frac{L}{Q} = ct^\omega
\end{equation}

where $c = a(1+\omega)Q^\omega$. According to (18) labor productivity increases proportionately with time.

These two versions of the progress function resemble the usual concepts of economics of scale and technical change. They also indicate the inherent identification problem between the scale and technical progress parameters noted earlier; labor productivity may be a function of both cumulated output and time.

The progress function concept can be extended to several inputs, each with a different rate of progress. This would imply a non-homothetic production function, however, since the usual factor substitution theorems of the neoclassical theory do not hold when differential rates of progress are applicable to different inputs. It is also possible to introduce relative prices in the progress functions [118, W. Oi, 1967].

The first version of the progress function, that is, (17) has been mainly applied to industrial processes and has received some

\textsuperscript{81} Kaldor assumes fixed proportions between capital and labor for existing equipment. He assumes $(K/Q)$ to be constant but the profit rate $(\Psi)$ to be a function of an exogenous pure rate of interest $(r)$ and a required risk premium $(s)$ that depends upon $(K/Q)$. That is, $\Psi = r + s(K/Q)$.\textsuperscript{82} $(K/Q)$ serves as an expectational variable in Kaldor’s model and its specification is critical for the prediction and stability of his model. Given $(K/Q)$ the growth rate of $(Q/L)$ becomes a function of the exogenously determined rate of technical progress.
empirical support, especially in the durable capital goods industries [7, A. Asher, 1956 and 120, L. Rapping, 1965]. Asher and Rapping suggest that \( a \) varies with different stages of production. The second version (18) has been applied successfully by Fellner [50, 1969] to estimate learning by doing behavior in athletic contests. However, there is a need for more evidence on the stability and the specific form of the progress functions, the precise measure of cumulative index \( Z \) in various industries, and the problems of aggregating progress functions.

IV.2 Quality Adjustments of Labor

Not all technical change, of course, is embodied in capital goods. Part of it must be transmitted through changes in the characteristics of the labor force, such as skill, education, age, sex, and race.\(^\text{22}\) This approach has been stressed in the works of Denison [26, 1962], Griliches [62, 1968], and J. Kendrick [79, 1970]. These authors derive separate indices for each of the labor qualities, and measure their contribution to the growth of output separately. However, in the final analysis their methods reduce to a quality-corrected series of man-hours.\(^\text{23}\) This approach starts out with the contribution of the aggregate man-hours and then adjusts it for various measurable characteristics of the labor force. Denison identifies reduction in hours worked, age-sex composition, and the educational quality of the labor force as the main factors affecting labor services. Relative earnings are used as measures of the marginal productivity of various types of labor and as weights in constructing the aggregate labor input.

Among the attributes of labor, the one that has been empirically the most significant—and therefore received the greatest attention—is the contribution of education to the growth of labor input.\(^\text{24}\) One rationale for this is that an economy with complex technology and a large scale of operations requires promotion of innovative abilities, as well as flexibility, skills, and a large information network. Since education enhances innovative ability [111, R. M. Nelson and E. Phelps, 1966] it thus contributes toward the maintenance of a large scale economy.

Another interesting aspect of educational investment in that with the phenomenal increase in the level of educational attainment one would expect the rate of return on educational investment to decline. The evidence, however, seems to suggest that it has remained fairly constant [149, F. Welch, 1970]. The reasons for this stability, pointed out by Welch, are the increasing weights in the industrial structure of growth-oriented industries, “nonneutralities” in the production process, and the rising quality of schooling. His empirical evidence for agriculture suggests that the nonneutralities in the production process—shifts toward more skilled and “innovator-allocator” types of workers—are partly responsible for the absence of a diminishing return to investment in education.

The return to education can be obtained either by computing the yield on investment in human capital or by estimating the contribution of the stock of human capital to growth of output. The first approach is adopted by Denison and Griliches, while the latter is developed by T. W. Schultz [128, 1962] and his followers (see J. Mincer [100, 1970] for references). Griliches develops a dis-

\(^{22}\) Griliches [62, 1968] emphasizes both capital and labor embodiment, while E. F. Denison [29, 1967] prefers to classify the contribution of advances in the design of capital goods as a part of “advances in knowledge.”

\(^{23}\) The distinction between the quality correction and embodiment effect discussed on p. 1148 should be kept in mind—not all quality adjustments mean embodiment.

\(^{24}\) Educational attainment has contributed about 40 to 50 percent to the growth of the total labor input of the U.S. economy in the postwar period [29, Denison, 1967, p. 208] while according to Griliches [62, 1968] it contributed about two-thirds to the growth of total labor input in the manufacturing sector for the period 1947–60.
tribution of employed males by occupation and the number of years of schooling completed. The number of school years completed is weighted by average earnings. Griliches makes no allowance for ability, or the hereditary characteristics of the labor force, on the grounds that the school-earning relation reflects these influences. Denson, on the other hand, makes a 60 percent reduction in the yield on education as an allowance for unmeasurable factors such as ability or family background. The main difficulties in using earnings by education as weights are: (a) that the results are highly sensitive to the choice and classification of income groups, and (b) that education is produced mainly outside the market and has strong externalities which are underestimated by the market prices [148, B. A. Weisbrod, 1962].

The stock of human capital approach, on the other hand, emphasizes the expenditure on education and is analogous to the investment models for physical capital. However, there are also some problems with this approach. It is difficult to distinguish between education as productive investment and as a consumption good, and between replacement of the existing stock of human capital and net additions to it. Further, there is the problem of externalities associated with education, which makes the calculation of an aggregate stock of human capital hazardous. Finally, there is the question of complementarity and substitution among stocks of human and physical capital and labor services. These problems suggest a considerable need for further research in these areas.

An important step, at least conceptually, in the right direction has been made by the quality adjustment hypotheses. In addition, empirical results give some support for the existence of both capital and labor embodiment [140, G. Szakolczai and J. Stahl, 1969 and 143, L. Thurow and L. Taylor, 1966]. Unfortunately, these results are not conclusive. In the absence of data with relevant vintage dimensions, it is difficult to distinguish precisely the contributions of these embodied technical changes from those of disembodied technical change and those due to economies of scale.

Moreover, the embodiment models and quality adjustment efforts do not explain why the next period's capital or labor is of better quality. In fact, if the linkage between the inputs and output is properly defined, all progress is necessarily disembodied. At each point in time new ideas are combined with initially given resources to produce better quality outputs in the next period [49, Feller, 1970]. The improvements in the labor and capital of time \( t \) is achieved via the consumption of new commodities and better designed investment goods produced at time \( t-1 \). Consequently, embodied and disembodied technological changes are entwined. Moreover, these models do not sufficiently stress the mutual interdependence of technical progress and capital formation. Each is promoted by the other and they are complementary processes, as we noted earlier. There seems to be a need for linking the embodiment and quality adjustment hypotheses to the optimization theories of investment and employment.

IV.3 Growth Accountancy Approach

In this approach the aggregate production
function is used as an organizing device or accounting format (and not as an estimation framework) to isolate the contribution of various factors to growth of output. The usual procedure is to assume linear homogeneous production functions with relative input prices taken as reasonable measures of marginal products. The works of E. F. Denison [26, 1962 and 29, 1967] and Jorgenson-Griliches [76, 1967] are the most important studies that use this approach. Denison attempts to reduce the magnitude of the "residual" to a pure technological progress effect after making proper adjustments for characteristics of the labor force (along the lines mentioned in Section III.2) and the magnitudes of capital and labor inputs. Jorgenson-Griliches try to explain all technical change by making the appropriate adjustment for aggregation and measurement errors in prices and quantities of the inputs.

Denison's work on factor productivity is extremely rich in content and extensive in coverage. At the expense of violating the spirit of his work, his approach could be summarized by the following relation.\(^{37}\)

\[
dQ = \mu \left[ \sum_{i=1}^{n} \alpha_i dX_i + \sum_{j=1}^{m} y_j + J \right]
\]

where \(dQ\) is the growth rate of national income valued at 1958 prices. \(\mu\) is a measure of economies of scale, \(\alpha_i\) refers to shares of the factors represented by \(dX_i\), \(y_j\) refers to the growth rate of various disequilibrium factors. Denison specifies \(dX_i\) \((i=1 \cdots 7)\) as the changes in employment, composition of employment, level of inventories, nonresidential land, nonresidential structures and equipment, quantity of dwelling and residential land, and the quality of international assets. \(y_i\) refers to adjustment factors due to sectoral misallocation of resources, institutional restrictions, inadequacy of aggregate demand, lags in the adoption of best-prac-

\(^{37}\) This relationship is taken from a mimeographed paper [91, 1969] which Mr. Denison kindly sent to me. A very useful index of total factor productivity has also been designed by Z. Griliches [61, 1967].
differences lies in the nature of reallocation of resources: (i) the contraction of agriculture (this accounts for only 0.1 percent in the U.K., but 0.8 percent in Germany, and 1.0 percent in Italy); (ii) the contraction of non-farm self-employed except for professionals and proprietors of substantial establishments—gains from this transfer range from 0.04 percent in the U.S. and U.K. to 0.22 to 0.26 percent in Italy, France, Norway and the Netherlands; (iii) gains from reductions in trade barriers varying from no reduction in the U.S. to 0.16 percent in Italy and Norway.

For the U.S., Denison deducted the contribution made to the residual by all sources of growth and attributed the remaining 0.76 percent to advances in knowledge. Because he felt that during this period the Western European economies mostly borrowed their new techniques from the U.S., he considered this 0.76 percent as a tolerable estimate of the contribution made by advances in knowledge in all these economies. Any remaining residual is accounted for by growth from sources so far not isolated—reduction in age of capital, a catching up of techniques in France and Italy, recovery from war distortions in Germany and Italy.

Any quarrel with Denison’s findings and approach will be due to a denial of his classification of the sources of growth or his judgment in handling the data. For example, capital-embodied technical change is ruled out (for the U.S.) by Denison on the grounds that the age distribution of the gross capital stock, which should register such change, has not varied greatly nor is it likely to do so in the future. Instead, he classifies this

68 Contrary to Denison’s position, J. Kendrick in his recent study [79, 1970] reports a very important relationship between \( dA/A \) and the age distribution of gross capital in the private domestic economy for 1948–66. He estimates the relation

\[
\log X = 0.20 + 0.30 \log X_1 \quad (9.04)
\]

\[
+ (3.02) \log X_2 - (4.49) \log X_3
\]

where \( X = \) total tangible factor productivity; \( X_1 = \) ratio

type of technical change as part of the “advancement of knowledge.” There is also the question of whether technical change—advancement in knowledge—can be achieved without capital formation [24, J. Cornwall, 1968 and 123, W. Salter, 1960]. Also, the interactions among labor and capital could be quite substantial and their adjustments may not be independent of each other [107, R. Nelson, 1964 and 106, Nadiri-Rosen, 1969]. Thus many of the specific results depend upon selection of a classification schema and handling basic data. But what is important is that Denison’s work opened up new and important areas of research; they need further exploration.

Jorgenson-Griliches [76, 1967] attempted to explain away the very existence of total factor productivity. They claim that “all” technical change can be explained by properly taking into account the aggregation and measurement errors in prices and quantities of the inputs and output. They adopt the usual neoclassical assumptions of competition, a constant returns to scale production function, and producers’ equilibrium (equality of factor-price ratios to their ratio of marginal productivity and equality of marginal rates of substitution of goods to their prices). Technical change is considered as a shift in the production function which is used as an organizing scheme as in Denison’s work.

Starting with the basic national income accounting identity, they use the Divisia index to calculate the rate of growth of total factor productivity \( (dA/A) \) as an index of the rate of growth in outputs, each weighted by its value share in total output less a similar index of rate of growth of inputs. Their

of real stock of intangible capital utilized to real tangible factor input; \( X_T = \) ratio of employment to civilian labor force, and \( X_T = \) average age of fixed reproducible capital stock. (Note that \( X_T \) dropped from 13.0 years to 10.3 years in the same period.)

- The fundamental national income accounting identity is

\[
\sum_{i=1}^{m} q_i Y_i = \sum_{i=1}^{n} p_i X_i
\]
empirical calculation suggests that the rate of growth of total factor productivity has been about 0.1 percent per annum for the period 1948–65.

The Jorgenson-Griliches approach differs from Denison's in the methods of handling the data and in the perception of where adjustments are needed, particularly in evaluating the contribution of the capital input. They argue that capital services—and not capital stock—should enter the production function. Total capital stock is first corrected for biases in deflators of its components, then adjusted by a trend-like rate of utilization. The resulting measure of capital services is further adjusted using the flow price of capital services, \( p_k \), instead of the asset price of capital, \( q_k \). These two

\[
\frac{dA}{A} = \frac{dY}{Y} - \frac{dX}{X} = \sum w_i \frac{dY_i}{Y_i} - \sum v_j \frac{dX_j}{X_j}
\]

or

\[
\frac{dA}{A} = \frac{dp}{p} - \frac{dq}{q} = \sum v_j \frac{dp_j}{p_j} - \sum w_i \frac{dq_i}{q_i}
\]

The weights \( w_i \) and \( v_j \) represent shares of \( i \) and \( j \) in the value of total output \( Y \) and total input \( X \), respectively. The rate of growth of productivity is therefore defined as the difference between the rate of growth of real output and the rate of growth of real factor inputs, each of these terms being a weighted average of the rates of growth of individual products and factors.

They decompose capital into several categories and, whenever possible, correct for the positive bias in investment good deflators. Furthermore, they substitute a new price index for structures. This index is constructed from the Bureau of Public Roads price index for highway structures, the Bell System price index for telephone buildings, and the Bureau of Reclamation price indices for pumping and power plants. This composite index rises faster than the OBE price deflator for structures through most of the postwar years. The BLS implicit price deflator for durables is replaced by an implicit price deflator for consumer durables, which rises less than the deflator for durables in the postwar period. Similar substitution is made for inventories. Capital stock is adjusted for the rate of utilization by using the average kilowatt hours of motors as weights to combine the utilization rates of the component industries and obtain the total manufacturing utilization series. This series indicates the trend in capital utilization and not adjustments—the substitution of different deflators and, especially, the adjustment for the rate of capacity utilization of capital—enable them to force the residual to vanish almost completely.

Their basic point, that if inputs and outputs were correctly measured there would be no residual left, is conceptually correct provided all the contributions of growth factors are faithfully reflected in the prices and quantities used in their study. But this is not necessarily the case. For example, if proper adjustments for the contribution of research and development by the government agencies and government financed capital goods utilized by the private sector [55, R. J. Gordon, 1969], etc., have been made by the authors, the residual might even have become negative, raising the possibility of overadjustment of input prices and quantities. The assumption of constant returns to scale, the notion of producers' equilibrium, and the use of income shares as measures of marginal productivities are rather restrictive. The use of factor shares as measures of factor productivity, in the face of substantial disequilibrium in the economy [142, L. Thurow, 1968], may under-estimate the contribution of resource reallocation in the growth process. Also, increasing the rates of growth of the inputs to explain away total factor productivity requires an explanation of the factors that determine the rate of growth of the inputs in the first place.

the cyclical fluctuations in capacity utilization.

The price of capital services is computed by using the familiar user cost formula,

\[
p_k = q_k \left[ \frac{1 - \mu v}{1 - \mu} r + \frac{1 - \mu v}{1 - \mu} \delta_k - \frac{1 - \mu x}{1 - \mu} dq_k \right]
\]

where \( \mu \) is the rate of direct taxation, \( v \) the proportion of return to capital allowable as a charge against income tax purposes, \( w \) the portion of replacement allowable for tax purposes, and \( z \) is the proportion of capital gains included in income for tax purposes; \( r \) is the interest rate, \( \delta_k \) is the depreciation rate and \( q_k \) the price of capital stock, \( K \). Using the flow price, \( p_k \), instead of the asset price, \( q_k \), for each class of investment goods, Jorgenson-Griliches attempt to remove any bias due to differences in the rates of replacement and of capital gain and loss among different capital goods.
The conclusion reached by Jorgenson-Griliches was amended recently by L. R. Christensen with Jorgenson [23, 1969] using annual data for the period 1929–67. Their calculation differs from the Jorgenson-Griliches study by substituting a measure of relative utilization of capital (derived from series on capacity and actual electricity consumption) and by properly separating compensation by legal forms of organization. Their results indicate factor productivity to have grown in 1948–67 by about 0.31 per annum instead of the reported 0.10 [76, Jorgenson-Griliches, 1967]. It is the use of a different rate of utilization that is responsible for this dramatic change.61 But what is important to underscore, as Christensen-Jorgenson themselves point out, is that the alternative estimates of total factor productivity are “highly sensitive to the choice of conventions for measuring real factor inputs.”

To summarize, the attempt to reduce the magnitude of the residual or explain away its existence has opened up new issues for research. Identification of various attributes of labor, capital, and knowledge, elimination of aggregation biases and better estimation of the factor prices, are clearly major contributions. Unfortunately, the specific results are too sensitive to changes in the types of data and methods of estimation to provide concrete quantitative figures about the contributions of various factors to the growth of output. It will be most useful if Denison’s and Jorgenson-Griliches’ approach is extended to more disaggregated levels. It would be necessary to take account of additional environmental variables (presently excluded) such as air and water, and government services such as law and subsidies, which definitely contribute to the productivity growth of an industry.

**Section V**

**Some Concluding Remarks**

An overall look at the literature surveyed indicates that during the last few years considerable progress has been made in formulating the concept and measurement of technological change. An important contribution towards understanding the determinants of technological change has been made by the theories of endogenous technological change and the attempts to explain the production and transmission of new knowledge. On the other hand, the formulation of more general forms of the production function (particularly the efforts to derive and estimate the production function indirectly) via cost functions has served to widen the scope of research in a new direction. Finally, the attempts to isolate the pure residual (or even to deny its existence) by attributing the growth of productivity to changes in the quality of inputs and their aggregation bias have been of considerable importance. The various controversies involved therein have brought to light the complexity of factors underlying factor productivity.

While the gains from past research are noteworthy, there still remain many open questions that have made it impossible to arrive at any bold and sweeping conclusions about the nature and causes of technological change.

**A. Some conceptual issues:**

1. Serious questions have been raised about the existence of the aggregate production function. On the other hand, the nature of the function, *i.e.*, the magnitude and stability of its parameters, is not yet established unequivocally, leading to strong doubts about its usefulness for empirical research even as an accounting framework. In fact the evidence suggests that the specification of the form of the aggregate production function is of secondary significance and con-

---

61 Denison [30, 1969] had claimed that Jorgenson-Griliches have introduced new measurement errors by using a common rate of utilization for all capital goods. The Christensen-Jorgenson’s result has vindicated Denison’s statement.
tributes very little to the explanation of the "residual."

2. The assumptions (such as perfect substitution between old and new capital, a constant rate of depreciation, etc.) that are required for empirical estimation and identification of the production functions, impose unrealistic constraints. The use of income shares as proxies for the contribution of various inputs to the growth of output is, at best, questionable. Other supplementary measures that can capture certain aspects of productivity which are not reflected in earnings would be a welcome addition. The identification of the separate contribution of disembodied, embodied, and biased technical change of economies of scale, and of changes in the elasticity of substitution to the growth of productivity is still not achieved. The core of the problem is the inherent dynamic interactions of each of these technical aspects with the factor-price relations on the one hand, and among themselves on the other. Working with the already "over-used" aggregative data will not get us out of this impasse. What is needed now is a major reallocation of research efforts towards improving the data and less towards developing highly sophisticated new functional forms or complicated estimation techniques.

3. There is need to develop models that probe further into the determinants of factor qualities, such as education, and the mechanism by which they enter the production function. The mere adjustment of labor and capital data by a constant scalar to account for quality changes is not very informative. For example, the assumption that productivity increases continually with increases in the level of education is unrealistic; after a certain stage, diminishing returns are likely to set in. Efforts to discover the exact linkages between inputs and their various characteristics should eventually lead to the integration of the existing theories of optimal investment and employment with the determinants of technological capital formation. Such an integration will be a welcome step forward.

4. The frequently used measures of technical change like R & D expenditures, the number of patents and of technical workers, etc. measure gross technological capital formation. They often give conflicting results for the magnitude and direction of technical change, since each of them is measuring some particular aspect of it. Further work is needed to distinguish the net and replacement technological capital formation and to develop a composite measure of technical change.

B. The measurement and welfare aspects of technical change:

1. The conceptual problems in defining and measuring the output of the government and service sectors deserve considerable thought. The existing statistics are actually measures of input. We need to know the contribution of government services like laws, regulations, etc. to the growth of factor productivity. The role of the foreign sector as a vehicle of technical change is another issue that has only recently received some attention and needs to be pursued further.

2. There are several difficulties in using GNP or national income in evaluating the role of technical progress. It excludes non-market activities and understates the importance of new products as a vehicle of technological change. It involves double counting, e.g., when technological progress generates pollution or other harmful side effects, some new expenditures need to be made to counteract them. The GNP also includes these expenditures in positive terms and thus overestimates the growth of factor productivity. This phenomenon is particularly important when considered in the context of the rapidly increasing rate of accumulation of negative output in the form of pollutants.

3. The literature often implies the value judgment that all technological change per
se is good and that human beings have an unlimited capacity to adapt to it. However, it is important to investigate the extent to which the novelty, diversity, and acceleration of change associated with rapid technical advance is conducive to human welfare. The costs of adjusting to new technology might be very high and the capacity of human beings to absorb continual and rapid change may be limited. No one would suggest that all technological progress should be stopped and the Luddite experiment be repeated. However, blind acceptance of every technical change just because it works and is profitable is not warranted either. What is needed is explicit recognition of its long-range economic and social costs and benefits.

4. To evaluate correctly the role of technological change, it is necessary to formulate an alternative system of accounting which registers the utilities from different types of assets—not only the conventional inputs, but also factors such as natural resources, physical environment, human skill and knowledge, social and political structures, etc. Such a conceptual framework should be able to accommodate both the purely technical advancement and the social innovations necessary to adapt to the new technology. This approach requires concentration on more disaggregative studies such as microeconomic production functions, inter-industry resource allocation models, location theories, etc. Further, we need to devote considerable attention to the development of models aimed at forecasting and planning the direction of future technical innovations.

These are only a limited number of problems that need consideration in future research. There are, of course, many others but a long story must come to an end some place.

References

13. BODKIN, R. and KLEIN, L. "Nonlinear Estimation of Aggregate Production
1172 Journal of Economic Literature


59. GRILICHES, Z. “Notes on the Measurement of Price and Quality Changes” in *Models of income determination*. Prince-


133. Solow, R. “Technical Change and the


