Volume Title: Output, Input, and Productivity Measurement

Volume Author/Editor: The Conference on Research in Income
and Wealth

Volume Publisher: Princeton University Press

Volume ISBN: 0-870-14181-3

Volume URL: http://www.nber.org/books/unkn61-1

Publication Date: 1961

Chapter Title: Economic Problems in Measuring Changes
in Productivity

Chapter Author: George J. Stigler

Chapter URL: http://www.nber.org/chapters/c2219

Chapter pages in book: (p. 47 - 78)

Economic Problems in Measuring Changes in Productivity

GEORGE J. STIGLER
UNIVERSITY OF CHICAGO

Labor Productivity as an Index of Productivity

PRODUCTIVITY measurement occupies an unusual and possibly unique place in the history of quantitative economic research.

Most important measurable concepts have a long history of theoretical discussion before the actual measurements begin on a large scale. The national income concept was discussed from the time of the mercantilists, the concept of a price level or the purchasing power of money has a similarly ancient theoretical tradition, and one could say much the same of the terms of trade, taxable capacity, the incidence of taxes, etc. Once the task of measurement became important, new theoretical difficulties were always uncovered, but the measurers had some initial guidance.

A second set of economic measurements appeared without much assistance from earlier theoretical tradition. The studies of the distribution of income by size are an example of this work, and so too is the derivation of Engel (income-expenditure) curves. In these cases the quantitative worker had to formulate his own concepts and do his own theorizing until his results began to attract the attention of the theorists, who then came to aid and thwart him with their refinements.

But only productivity measures of important economic magnitudes arose in the face of a theoretical tradition which denied them any relevance to economic structure or policy. A productivity measure, until recently, was a measure of the average product of some class of productive services. When they began to be calculated on a large scale it was already a basic proposition of economics that one should never look at average products, only at marginal products.

The equality of marginal products in all uses is a necessary condition for efficient use of a resource, and hence for maximum output. The marginal productivities are basic elements of the demands for productive factors. The dependence of marginal products on the quantities of and proportions among productive factors is the essence of the theory of production. These are illustrative statements of the fundamental role of marginal products in economic analysis. So far as I know, not a single theoretical statement of any importance can be made about the average products of factors.[1]

Yet certainly the product, and possibly even the productivity, of calculators of productivity indexes continued to increase. Does this mean that a set of statistically unsound quasi-random numbers were being supplied to careless users, or that the economic theorists were purists who did not recognize the great usefulness of approximate data? The answer is not easy.

An approximate answer depends upon the closeness of the approximation and the question which is being asked. For a lame ant the statement that the height of a house and of the Eiffel Tower are equal is a satisfactory approximation; a pilot might need a closer approximation. The uses of productivity data, however, are infinitely varied, and it does not seem possible to present any objective criterion of the minimum goodness of approximation that is generally required.

We may say a trifle more about the accuracy of a labor productivity measure as an estimate of a capital-and-labor productivity measure. Labor is quantitatively the largest input (in marginal units of measure, i.e., as a share of income) so large changes in labor productivity over time are likely to reflect at least roughly the movements of a properly defined measure of productivity. But in general the labor productivity measure will exceed it by more, the more capital has grown relative to labor.[2] Labor productivity will therefore be a better measure of total productivity, the more nearly proportional the increases of labor and other resources over time, and the smaller the relative weight of nonlabor resources in total input.

The extent to which capital and labor change together over time proves to be fairly close in the manufacturing sector. Using Kendrick's

[1] What propositions there are about average products—such as that average product should be maximized to maximize output if only one factor is scarce—are polar cases.

[2] Let $P = A(t)f(C,L)$, where P is output, C is capital, L is labor, and $A(t)$ is the index of productivity. If technical progress is independent of the proportions between the factors, as this production function assumes, and there are constant returns to scale, one may derive the equation,

$$\frac{d\left(\frac{P}{L}\right)/dt}{P/L} - \frac{A'(t)}{A(t)} = \frac{w_c d\left(\frac{C}{L}\right)/dt}{C/L},$$

where w_c is the share of capital return in the total product, i.e., $w_c = \frac{\partial P}{\partial C} \cdot \frac{C}{P}$. [See R. M. Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, XXXIX (1957), 312 ff.] It follows that the relative change in output per worker will be in excess of the true index of productivity if capital increases relative to labor, and the two estimates will approach equality as w_c approaches zero, for given relative changes in the factors.

data, we may calculate the coefficients of correlation between inputs of man-hours and capital:

All manufacturing, 1869–1953, $r = .984$ ($n = 10$)
Two-digit manufacturing industries
1929–53 change for each industry, $r = .860$ ($n = 20$)
1948–53 change for each industry, $r = .728$ ($n = 20$)
1952–53 change for each industry, $r = .257$ ($n = 20$)[a]
1949–50 change for each industry, $r = .610$ ($n = 22$)[a]
1950–51 change for each industry, $r = .868$ ($n = 22$)[a]
1951–52 change for each industry, $r = .574$ ($n = 22$)[a]

[a] Based upon my data.

Our interest is primarily in the interindustry correlations, which indicate that the industrial pattern of labor input changes is fairly closely correlated with the corresponding pattern of capital input changes over considerable periods, but that the correlation is smaller and more unstable the shorter the period of comparison. The correlation for annual changes is at times so small as to make the labor productivity index highly unreliable as an index of capital productivity.

Even if the movements of man-hours and capital are very different, a labor productivity index will provide a tolerable estimate of total productivity if the weight assigned to capital is small. In manufacturing this proves to be the case: in Kendrick's two-digit manufacturing industries the weight of capital to that of labor is 1 to 3.23, on average, with a range, however, from 1 to 11 (apparel) to 1 to .75 (products of petroleum and coal).[3] The remark seems only partly relevant to these figures,[4] but it should be emphasized that the common but erroneous practice of excluding working capital from capital leads to exaggerated relative weight for labor inputs.

A labor productivity index seems generally to rank the commodity-producing industries correctly with respect to true productivity changes. It does not follow that the actual numerical changes in productivity are reliable. In fact they are biased, and substantially so. Only when changes in labor and other inputs are proportional will the labor index be unbiased.[5] In all manufacturing, the regression of

[3] The apparel ratio is undoubtedly too low because of the omission of rented capital.

[4] Kendrick's weights are based upon the total remuneration to capital and labor, although his index of the quantity of capital is restricted to durable capital plus inventories. I do not know how good an index of total capital this combination is.

[5] Using the notation of footnote 2, if w_c is stable the excess of labor productivity over true productivity changes between periods 0 and 1 will be

$$w_c \log \left(\frac{C_1}{L_1} \Big/ \frac{C_0}{L_0}\right).$$

If $C=a+bL$, the sign of expression is positive if $aL_0>aL_1$, which holds if $a<0$, $L_0<L_1$.

capital on man-hours for the period 1869 to 1953 was

$$C = -34.8 + 1.232L,$$

where C is capital and L is man-hours. Labor increases were accompanied by larger relative capital increases, so labor productivity measures overstated true productivity changes. On the other hand, in the interindustry comparison of inputs for 1929–53,

$$C = 25.9 + .832L,$$

so here the labor productivity index understated the differences among industries in the true productivity changes. In the annual changes, a is negative in two years and positive in the other two, so the direction of bias was unstable.

Are the labor productivity measures valuable approximate answers to an important question, or are they misleading pieces of arithmetic? In comparisons of productivity changes over long periods, and comparisons among industries with widely differing rates of productivity increase, the rankings of productivity increases by labor productivity increases are tolerably reliable. Even this tentative conclusion, which I interpret to be adverse to most short-run uses of labor productivity measures, assumes the accuracy of the total productivity measures, and it is to this, the main problem of this paper, that I now turn.

General Considerations

A pure measure of economic progress measures the increase in the output of given resources, or the decrease in the inputs for a given product.[6] The shifting mixtures of inputs which are responses to changes in their relative prices do not constitute advances in productivity, which come only from changes in the "state of the arts."

The state of the arts is the heritage of technical and economic knowledge which is possessed (by whom we consider later) at a given time. Its relevant content is summarized by the list of technologies which are not inferior in the sense that no technology in this list uses more of some inputs (and no less of others) than any other technology.

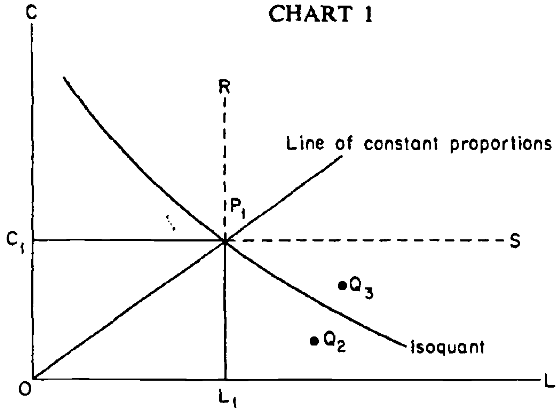
This set of noninferior technologies is displayed graphically by an isoquant—a curve showing the minimum combinations of inputs necessary to produce a given quantity of a given product. (Inferior technologies lie above the curve.) In Chart 1 we draw a curve displaying all combinations of (say) capital (C) and labor (L) which in the existing state of the arts will produce a specified amount of a given product. This is our isoquant; under the normal assumptions of

[6] The two are equivalent only if the production function is linear and homogeneous; this problem is discussed later.

diminishing returns to each productive factor (and, in the multifactor case, general complementarity of inputs), this curve is convex to the origin.[7] The entire quadrant above the curve represents less efficient methods of production, in the sense that the given product is there made with more of one input and no less of the other than would be required by available alternative techniques.

Suppose now that we observe a new process of production, Q_2 or Q_3, at a subsequent time; how can we tell whether it is more or less efficient than that indicated by P_1? In formal theory we postulate a known production function, and we would unhesitatingly say that Q_3 is an inferior technique—knowledge has retrogressed—and Q_2 a superior technique. Lacking this knowledge of the production function, we can say only that points in the rectangle whose northeast

CHART 1



corner is at $P_1(OL_1P_1C_1)$ represent techniques superior to P_1, and those in the infinite rectangle whose southwest corner is at $P_1(RP_1S)$ represent techniques which are inferior to P_1. Even this weak conclusion depends upon the assumption that the qualities of the inputs have remained constant: if they have deteriorated, the rectangles are no longer well-defined.

So far, of course, we have not used one part of our observational information: the relative prices of inputs in periods 1 and 2. Under

[7] Let the production function be $P=f(C,L)$. Then the slope of an isoquant is given by $dP=0$, or

$$\frac{dC}{dL} = -\frac{f_l}{f_c}.$$

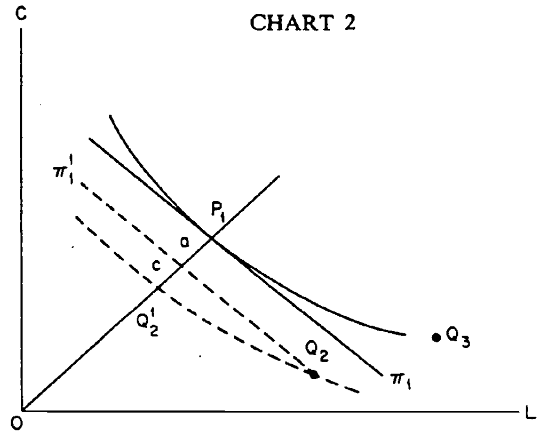
Then
$$\frac{d^2C}{dL^2} = -\frac{f_c^2 f_{ll} + f_l^2 f_{cc} - 2 f_c f_l f_{cl}}{f_c^3}.$$

Since $f_{ll}<0$ and $f_{cc}<0$, the second derivative is necessarily positive if f_{cl} (or a sum of such terms in the general case) is positive.

competition (we look at monopoly later), the entrepreneur who minimizes costs in each period operates where marginal products are proportional to factor prices, or, in terms of Chart 2, where a factor price line is tangent to the relevant isoquant. Any process falling below the price line (π_1) was unattainable in period 1 or it would have been adopted because it would lower costs. In our illustrative example, we may say that relative to P_1, Q_2 is certainly a superior technique and Q_3 possibly a superior or inferior one. The process Q_2 represents an improvement of approximately a/OP_1, for the inputs would have fallen (in period 1 prices) in the proportion, a/OP_1.

These approximations may be rather poor if price relatives in the second period were substantially different. There may exist another process (Q_2^1) which would have been even cheaper at period 1 relative

CHART 2



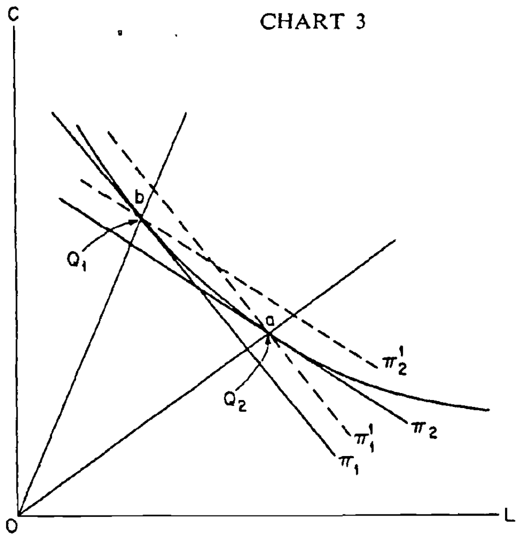
prices, so that the increase in efficiency is actually $(a+c)/P_1$. Whether this is true, or contrariwise the approximation exaggerates the gain in efficiency (and Q_2^1 lies above π_1^1), depends upon the shape of the new isoquant relative to the earlier isoquant.[8] The two isoquants can even intersect if once-known techniques are forgotten, or the qualities of some inputs deteriorate. Advances in technology do not have to lead to constant displacements of the isoquants; only one input may be economized by a given advance.[9]

[8] There are obvious parallels between our ratios and consumer index numbers, but the analogy is far from complete. The cost of living indexes rest upon the assumption of constant tastes, and lose their meaning if tastes change. In productivity analyses, on the contrary, the essence of the problem is the measurement of the change in the state of the arts, the analogue of tastes.

[9] Only if the isoquants bear a very special relationship to one another will the gain in efficiency measured on a ray from the origin through P_1 equal that measured on a ray through Q_2. If the production function has a very simple form, such as $Q_2(C,L)=$

Of course a change in the relative prices of inputs will also lead to a change in inputs, even if the state of the arts does not change. Thus, in Chart 3 the maximum profit position is at Q_1 if the price line is π_1, and at Q_2 if the price line is π_2. If combination Q_1 is valued at π_2 prices, a relative rise in efficiency from period 1 to period 2 of a/OQ_2 is indicated; and if Q_2 is valued at π_1 prices, a relative fall in efficiency from period 1 to period 2 of b/OQ_1 is indicated. In fact if these relationships failed to hold, a change in technology would be demonstrated. A price line with the average of the slopes of π_1 and π_2 will not necessarily pass through both Q_1 and Q_2, so changes in relative prices will, with the usual index number formulas, lead to some change in

CHART 3



reported efficiency, although none has taken place. The ambiguity could be eliminated only if one knew the production function.

If the total output of the commodity changes when the new technique is adopted, as is usually the case, along which isoquant should we measure productivity changes? If the production function is linear and homogeneous, no choice is necessary: then to produce m times as much as P_1, we originally needed mL_1 and mC_1 and the family of all possible isoquants has the same slope at all points along the line of fixed proportions, OP_1[10]. Nothing but a scale factor is involved, and we can assert that real inputs fell in the proportion of the

$kP_1(C,L)$, with each production function homogeneous of degree one, the measurements along the two rays will be equal.

[10] For $dC/dL = -f_l/f_c$, and if f is homogeneous of degree 1, its derivatives are homogeneous of degree zero, i.e., changing inputs by a multiple m leaves each marginal product unchanged.

fall in inputs necessary to produce P_1. If the production function does not have this simplifying property, we must recognize a new element, which (under competition) is the presence of external economies. We discuss this problem in the final section.

Let us recur to the question of what a given state of the arts means. In formal theory it is usually taken to be the sum total of existing knowledge, no matter how much of this knowledge any one entrepreneur possesses at a given time, because in long-run equilibrium in a stationary economy, everybody eventually learns everything relevant. In the changing economy this will hardly do; what is relevant is the frequency distribution of knowledge among entrepreneurs. So we must recognize that even if there are no ideas new to an economy, there can be large increases or decreases in the average knowledge of entrepreneurs in an industry, and this is one reason why the state of the arts cannot be defined by an inventory of technical knowledge.

The relevance of the foregoing apparatus depends upon the existence of long-run competitive equilibrium in each period. If a given firm is not in equilibrium, it will not be operating with the lowest possible expenditure for a given output, and the marginal products will not be proportional to prices. Hence the firm will be elsewhere on a short-run isoquant than the point of tangency of the price line, and the measure of a technical change will be a mixture of gains or losses from any such technical change plus those from moving closer or farther from the long-run minimum cost condition.

Two procedures are now used to deal with this problem of disequilibrium. The more common procedure is to pick prosperous years for one's calculations. It is hard to place confidence in the precision of this procedure, especially when it is applied to individual industries which may be in a state of depression while business is generally good.[11] Over long periods the estimates should not be seriously biased, but this is a further reason for distrusting short-run estimates.

The other procedure is, in effect, to scale down the nominal amount of inputs which are not fully used—presumably capital, as a rule.[12] If the technical proportions between the inputs are rigid, this procedure is clearly (half) correct: the effective input of the surplus factor varies proportionally with the nonsurplus factor.[13] But even then the

[11] For example, in Kendrick's terminal year (1953) for two-digit manufacturing industries, apparel products were earning only 2.6 per cent on assets (3.1 per cent in 1947 prices) after taxes, a lower rate than in any preceding year since 1938.

[12] One of the first to do this, in an oblique way, was J. M. Clark, "Inductive Evidence on Marginal Productivity" (*American Economic Review*, 1928), reprinted in *Preface to Social Economics* (New York, 1936).

[13] Of course, any one input then yields a correct estimate of changes in "technical" efficiency.

procedure is half wrong: it measures technical possibilities, not the efficiency of an economic organization. If entrepreneurs generally keep too much of one input around, whether for reason of esthetics or stupidity or nepotism, this is a source of inefficiency to be treated on a full parity with "technical" inefficiency.[14] When the proportions between inputs are variable, the adjustment is also incorrect in estimating the efficiency of available technical methods. Greater capital per laborer increases the marginal productivity of labor, and only a portion of the capital is, in an esoteric sense, excessive.[15]

The Presence of Monopoly

Prices equal marginal costs under competition, and they do not under monopoly. Hence decisions of entrepreneurs are guided by quantities which are (in principle) reported and observable under competition, but are guided by quantities which are not observable or reported under monopoly. What difference does this make in our ability to interpret the evidence on economic progress? We consider first the effects of monopsony in the factor market, and then monopoly in the selling markets. They affect, alternatively, the measures of input and output.

The monopsonist is guided in his combination of productive services by their marginal costs. His condition for maximum profits is:

$$\frac{\text{Marginal product of } C}{\text{Marginal cost of } C} = \frac{\text{Marginal product of } L}{\text{Marginal cost of } L}.$$

In terms of our isoquants, the line of equal outlay (π_1 of Chart 2) becomes a curve, which is usually concave to the origin.[16] In Chart 4,

[14] Of course, to the extent that the unemployment of capital is due to factors outside the control or reasonable anticipation of the entrepreneur, the procedure aims at a meaningful question: How efficient are the entrepreneurs within the area where their own decisions are determining? But I doubt that one can give a useful answer, since this area of self-determination is ambiguous analytically as well as empirically.

[15] The issues involved seem to be identical with those encountered in the determination of "excess capacity."

[16] Total cost is $C = cp_c + lp_l$, and the slope of the equal outlay curve is given by

$$\frac{dC}{dl} = 0 = c\frac{dp_c}{dc}\frac{dc}{dl} + p_c\frac{dc}{dl} + p_l + l\frac{dp_l}{dl}$$

or

$$\frac{dc}{dl} = -\frac{p_l + l\frac{dp_l}{dl}}{p_c + c\frac{dp_c}{dc}} = -\frac{(MC_l)}{(MC_c)}$$

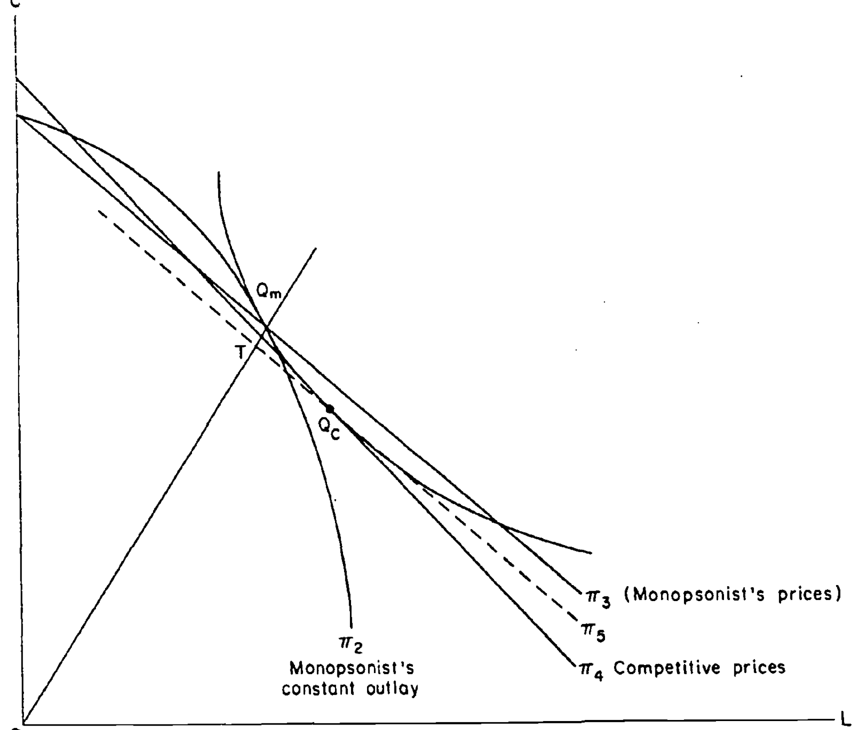
$$\frac{d^2c}{dl^2} = -\frac{MC_c\left(l\frac{d^2p_l}{dl^2} + 2\frac{dp_l}{dl}\right) - MC_l\left(c\frac{d^2p_c}{dc^2}\frac{dc}{dl} + 2\frac{dp_c}{dc}\frac{dc}{dl}\right)}{MC_c^2}.$$

This expression is negative if the first and second derivatives of the supply functions of the factors are positive or zero.

the monopsonist operates at Q_m, where the equal output and equal outlay curves are tangent.

Let us assume that the monopsonist has control over the wage rate he pays, but not over the price of capital. Then the marginal cost of labor will exceed its price (on the reasonable assumption that the supply curve of labor is positively sloping), and the slope of the constant outlay curve will be steeper, at the indicated inputs of labor (OL_m) and capital (OC_m) inputs, than the corresponding price ratio.[17]

CHART 4



That is, the prices paid by the monopsonist generate a price line (π_3) which is flatter than the isoquant at its point of tangency (Qm) with the constant outlay curve (π_2).

[17] Let P^o be the specific price of capital services of which C_m units are purchased, and P_l^o the price of labor when L_m units are purchased. Then the slope of the monopsonist's constant outlay curve will be

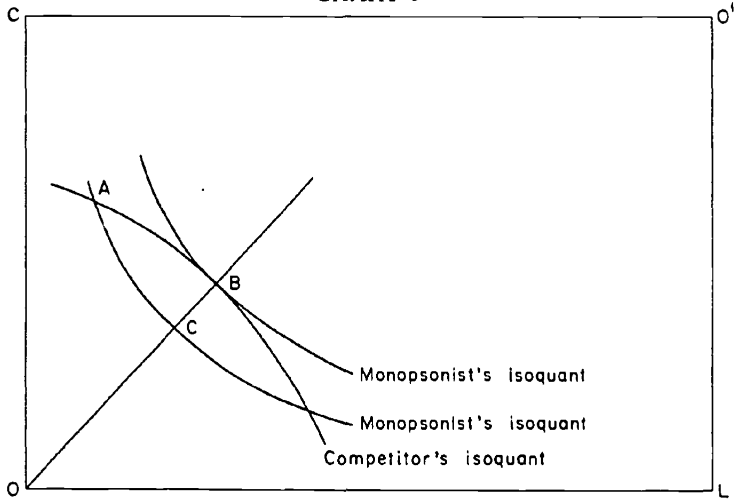
$$-\frac{P_l^o + L_m\left(\frac{dp_l}{dl}\right)_{l=L_m}}{P_c^o}$$

and the slope of the corresponding price line is $-P_l^o/P_c^o$, which is algebraically larger, for the second term in the numerator of the previous expression is also positive.

If the monopsonist is replaced by a large number of firms so that competition is established in the labor market, the individual competitor will face the factor price line π_4.[18] The price of labor will be higher relative to that of capital because more labor will be used; more labor will be used because the marginal cost of labor to a competitive buyer does not exceed its price. The price line π_4 is accordingly steeper than π_3. It is tangent to the isoquant at Q_c. The competitive inputs, valued at the monopsonist's prices (π_5 parallel to π_3), indicate that the removal of monopsony increased efficiency in the proportion TQ_m/OQ_m.

This calculated increase in productivity as a result of eliminating monopsony is essentially misleading. It is similar to, although not

CHART 5



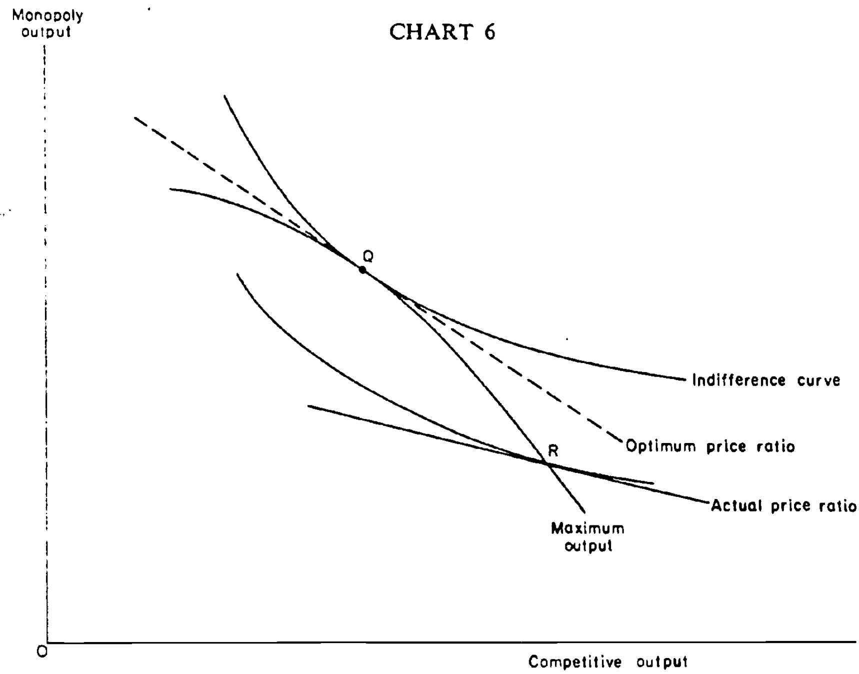
really identical with, the apparent changes in productivity that arise in the competitive case when relative input prices change. Yet it is true that monopsony leads to an inefficient allocation of resources among industries, so its elimination should lead to an increase in the productivity of the economic system. The conventional argument to this end may be illustrated with the familiar "box" diagram in Chart 5. The sides of the box measure the total quantities of capital and labor, the output isoquants of the monopsonized sector are drawn with respect to 0, and those of the competitive sector with respect to 0′. The original situation of the monopsonist and the competitive sector is at A. If the monopsony is eliminated, the new

[18] With n competitors, and no economies or diseconomies of scale, the isoquant of the competitor will be the same as that of the monopsonist except that corresponding points on the input scales will be $1/n$ times as large.

position B is reached, with the monopsonist's output the larger by BC/OC. (The actual equilibrium position will lie on a contract ourve of which B is one point; a knowledge of demand conditions is necessary to determine it.)

The increase in efficiency is not due to any change in the state of the arts, as this phrase is customarily used. But if we include in the arts the economic organization of the economy, as we certainly should, there has been an advance when monopsony is eliminated.

CHART 6



The increase in efficiency cannot be attributed to any one industry, however.

Monopoly power in selling markets is much more important. The value of the marginal product of a factor will be higher in the monopolistic than in the competitive industries, so the elimination of monopoly would allow a redistribution of resources such that real income of the community could be larger. How does the presence of monopoly affect the estimates of economic progress?

Consider a situation in which all resources are fully employed in either the monopolistic or competitive sectors of the economy. There will be a production possibility curve which displays the maximum outputs of the two sectors, given in Chart 6. Subject to difficulties much discussed in welfare economies, there will be a collective consumer indifference curve and an output given by Q.[19] With monopoly the outputs will be at R. The quantities of the resources are the same in both situations and the relative prices of the various resources may also be the same,[20] so efficiency varies between the two situations in proportion to the measure of output.

The monopoly combination R falls on a lower indifference curve than Q, so efficiency rises if monopoly is eliminated. This result is certain to be revealed if the conditions of our diagram are met, for R falls on a lower price line than Q at either monopolistic or competitive prices. But if monopolistic and competitive sectors have constant costs, the production possibility curve coincides with the optimum price line and R falls on this line; in this event removal of monopoly does not lead to an increase of output measured in competitive prices.

The most interesting point raised by the presence of various forms of monopoly is the aggregation problem their effects pose. When these inefficient forms of market organization diminish in scope or quantitative importance, no one industry deserves credit for the resulting increase in productivity. Even if the productivity of each industry is correctly measured, the sum of their advances is not necessarily equal to the social advance in productivity. This can be restated as the proposition that general economic structure is also an input. Unless it is explicitly introduced—and I happily leave this task to the input specialists—there will be unexplained increases or decreases in productivity in the economic system. This is similar to but distinguishable from the effects of external economies (to which I now turn), because this component of the efficiency of an economic structure is independent of its size.

The Economies of Scale

We have so far assumed that the activity we are measuring is subject to constant returns to scale, so an increase of K per cent in every input will lead to a K per cent increase of output—in the absence of economic progress. This assumption, which is explicit or implicit in all productivity calculations, is not obviously correct, and yet its abandonment could have radical influence on these calculations. If a K per cent increase of each input leads to an M ($>K$) per cent increase of output even with a given state of the arts, the conventional calculations overstate the rate of progress, and conversely if $M<K$.

Let us approach this problem through a concrete example. The

[19] We can avoid the difficulties in the concept of community indifference curves by assuming that consumption is independent of the distribution of income.

[20] They will be the same if the monopolistic and competitive sectors use various resources in the same proportions.

real quantities of labor, capital, and value added in American manufacturing industries are given in Table 1. We may fit to these data the so-called Cobb-Douglas production function;

$$P = aC^{\alpha}L^{\beta},$$

where P is product, C is capital, L is labor, and a is a scale constant.

TABLE 1

Employment, Capital, and Value Added in Manufacturing Industries, 1900-48 (1929 prices)

Year	Capital (millions of dollars)	Laborers Employed (thousands)	Value Added (millions of dollars)
1900	17,452	5,063	9,275
1909	31,734	7,226	13,674
1919	46,094	9,665	18,042
1929	63,022	10,502	30,591 (ca.)
1937	55,319	10,619	30,581
1948	82,427	15,322	49,801

SOURCE: Daniel Creamer, *Capital and Output Trends in Manufacturing, 1880–1948*, p. 18.

When Douglas fitted such a function, the sum of the exponents $\alpha+\beta$ was approximately unity, so there were approximately constant returns to scale.[21] This result depended partly upon the period he covered (1899–1922), partly on the deficiencies in his data. When we use Creamer's data (in Table 1), we obtain the equation,

$$P = .19C^{.46}L^{.90},$$

and the sum of the exponents is 1.36.

Before we turn to the discussion of this result, let us simplify it by expressing all inputs in terms of capital, and thus put aside the question of changes in the proportions of inputs. This simplification can be made, at least approximately, by expressing labor in terms of capital by use of the rate of substitution between labor and capital in 1919.[22] The production function of manufacturing for the period

[21] In *The Theory of Wages* (New York, 1934), β was set equal to $(1-\alpha)$; in "Are There Laws of Production?" *American Economic Review*, March 1948, the sum of the coefficients was 1.04 with the same data.

[22] With the production function in the text,

$$\frac{\partial P/\partial C}{\partial P/\partial L} = \frac{.46aL^{.90}C^{-.54}}{.90aL^{-.10}C^{.46}} = \frac{46L}{90C},$$

and substituting the values of capital and labor for 1919,

$$\frac{\partial P/\partial C}{\partial P/\partial L} = \frac{1}{9.33},$$

if we multiply the labor inputs by 9.33, the total "capital" (input) was \$64.7 billions in 1900, \$225.4 billions in 1948.

can then be written,

$$P = bC^{1.36}.$$

The equation summarizes the basic finding that the increase in "capital" (inputs) from 1900 to 1948 of 248.4 per cent led to an increase in output of 435.5 per cent. The conventional measures of progress say that the increase was (535.5/348.4−1) 100=53.7 per cent, or 1.08 per cent per year on average.[23]

This estimation of the rate of progress obviously rests upon the assumption that in the absence of progress the production function would have been $P=bC$. But suppose that there were increasing returns to scale, so in the absence of progress the production function would have been $P=bC^{1.1}$. Then if capital increased from 1 to 3.384 the product would have risen from $P_{1,900}$ to $(3.48)^{1.1}\times P_{1,900}$, or to 3.947 times the initial output, and the measure of progress over the period would be (535.5/394.7−1)100=35.7 per cent. If the function in the absence of progress had been $P=bC^{1.2}$, progress would have been only 19.7 per cent over the period. With constant returns to scale, the estimate was 53.7 per cent, so these alternative production functions reduce the estimate of progress by almost one-third and three-fifths respectively. Any considerable economies of scale would have a large effect upon our estimates of technical progress.

A larger economy should be more efficient than a small economy: this has been the standard view of economists since the one important disadvantage of the large economy, diminishing returns to natural resources, has proved to be unimportant. The large economy can practise specialization in innumerable ways not open to the small (closed) economy. The labor force can specialize in more sharply defined functions; we can have economists who specialize in national income estimation, tax avoidance techniques, measurement of productivity, or writing textbooks. The business sector can have enterprises specializing in collecting oil prices, in repairing old machinery, in printing calendars, in advertising industrial equipment. The transport system can be large enough to allow innumerable specialized forms of transport, such as pipelines, particular types of chemical containers, and the like.

The argument, familiar since Adam Smith's time, surely is valid. The question is: how important are the economies of scale of the economy (or industrial subsets of the economy)? We cannot use time

[23] The output per unit of input is

$$\frac{P}{C} = bC^{.36}$$

and if $C=ke^{rt}$, i.e., if input grows at the rate of r per cent, "efficiency" grows at the (percentage) rate of $.36r$.

series on the inputs and outputs to answer this question because we have no independent measure of economic progress.

It might appear more promising to fit a cross-sectional production function to numerous industries, and estimate the extent of increasing returns from this function. Such an estimate would be too low; it would at best measure the extent of economies of size of individual industries and therefore ignore economies which are due to the growth of the entire economy, which are shared by all industries. Thus the gains from specialization in large industries would be measured, those from improved transport and banking systems would be excluded.

Such a function has of course been calculated several times; the function for American manufacturing in 1909, for example, was found to be

$$P = .90L^{.74}C^{.32}.\ [24]$$

This function displays increasing returns (.74+.32=1.06). But a function derived in this way is essentially meaningless; under competition the return per unit of labor or capital (product measured in value terms) will be the same in large and small industries. At best, therefore, the cross-sectional functions measure only monopoly returns or short-run disequilibria. To use this approach would require a measure of output that did not wash out the superior efficiency of large scale.

The method would be more attractive if the interindustry comparisons were international, for then the measure of the effects of scale would not be obscured by the forces of competition (unless mobility of resources between the nations were high), and the measure would not exclude the economies of scale of the entire economy which are shared by all industries. Unfortunately, there are apparently no countries that have approximately the same state of the arts as the United States and also possess satisfactory data on inputs and output.[25] A rough comparison of the United States with Great Britain will have to serve as an example.

We may compare the physical outputs of corresponding industries in the United States and Great Britain in 1947 and 1948, using Frankel's data.[26] Let us assume that the production function (not technique) of an industry is the same in both countries, so if the function for Great Britain is

$$P_e = aL^{\alpha}C^{\beta},$$

[24] M. Bronfenbrenner and P. H. Douglas, "Cross-Section Studies in the Cobb-Douglas Function," *Journal of Political Economy*, December 1939.

[25] The Canadian economy, which meets these conditions, is so closely related to our economy that it reaps a large part of our gains of specialization.

[26] Marvin Frankel, *British and American Manufacturing Productivity*, University of Illinois Bulletin (Urbana, 1957).

that of the United States is $P_a = b(\lambda_l L)^{\alpha}(\lambda_c C)^{\beta} = k\lambda_l^{\alpha}\lambda_c^{\beta}P_e$, where λ_l and λ_c are the ratios of labor and capital of the American industry to the corresponding quantities for the British industry. We may fit a function,

$$\frac{P_a}{P_e} = k\lambda_l^{\alpha}\lambda_c^{\beta},$$

and calculate $(\alpha+\beta)$. The function, estimated from twenty-three industries,[27] is

$$\frac{P_a}{P_e} = 1.45\lambda_l^{.827}\lambda_c^{.444} \quad (R^2 = .806),$$

so $(\alpha+\beta)=1.27$.[28] This equation displays extremely powerful economies of scale: applied to the 48-year period covered by Creamer's data, it would reduce the effects of technical progress to an 8.5 per cent increase in output, or to one-sixth of the amount yielded by a constant-returns-to-scale assumption.[29]

This estimate of the economies of scale is no doubt biased upward. The ratio of American to British inputs is very probably understated. This is especially likely for capital, where we follow Frankel in using fuel consumption as an index of capital. The American data suggest that this index does not vary in close proportion with capital.[30] Even the labor figures, which are simple counts of employees, may have the same bias because our labor force has benefited from much more educational investment.

The conclusion to be drawn, aside from the inevitable one that more work should be done, is that economies of scale are potentially of the same order of magnitude as technical progress. I consider the problem of establishing the approximate magnitude of these economies a major one, not merely of productivity calculations, which are not especially important, but of the theory of economic growth.

[27] One can obtain P_a/P_e and calculate λ_l from Table 10 of Frankel; from Table 5 one can obtain λ_l/λ_l and hence λ_c.

[28] The standard error of α is .19; that of β, .17. The algebraic values of the residuals are not correlated with the absolute size of the American industry, but there is a moderate positive correlation of the absolute values of the residuals with absolute size of industry.

[29] J. B. Heath has recently made a similar comparison of Great Britain and Canada in "British-Canadian Industrial Productivity," *Economic Journal*, December 1957. The chief difference is the use of horsepower as a measure of nonlabor inputs. For fourteen industries he obtains the equation

$$P = 1.55\lambda_l^{.73}\lambda_c^{.31} \quad (R^2 = .9705),$$

where the variables represent the ratio of Canadian to British quantities. The main objection to this comparison is given in footnote 25.

[30] The 1954 capitals (in 1947 prices) may be correlated with fuel purchases for the twenty-two two-digit American manufacturing industries; the Pearsonian coefficient is only .536.

COMMENT

ROBERT M. SOLOW, Massachusetts Institute of Technology

I had almost come to the end of Stigler's paper, nodding agreement as he laid it on the line for the statisticians, when I suddenly realized that I hadn't thought of anything to say. And in our profession to be struck dumb is considerably more damaging to a reputation than simply making mistakes. But fortunately for me, two pages before the end I found an important point on which Stigler and I disagree. It bears on what we both think to be one of the essential and neglected aspects of the problem—the question of increasing returns to scale.

The natural places to try to catch the effects of increasing (or decreasing) returns to scale are aggregate time series like Stigler's Table 1, or their microeconomic analogues. But then we run into a familiar kind of identification problem. Over long periods of time, capital, labor, and output grow. The effects of increasing scale and of technical change are mixed. The data provide no sure and simple way of segregating those increases in output per unit of input which would have occurred through the mere passage of time and increase of knowledge, even if inputs were constant, from those increases which would in fact have been available earlier if only the system had been larger. It is true that over shorter periods one can find decreases in employment and even in capital inputs so that any increases in output should be attributable to some kind of improvement in technique or efficiency (compare 1929 and 1937 in Table 1). But no one likes to hang an argument on these depression observations, on the ground that the equilibrium relationships we seek are likely to be disturbed in such times of rapid, short-run adjustment.

I leave aside all the additional difficulties, such as the inputs we don't measure at all (entrepreneurship is the standard example), the inputs whose measures are systematically biased by quality changes, and shifts between market and nonmarket activity.

One turns next to the possibility of some sort of cross-section measure which will hold time, and therefore the state of the arts, constant. Stigler considers for a moment the standard sort of Cobb-Douglas cross-section study applied to a single economy and finds it "essentially meaningless." I consider this much too generous an evaluation. He goes on to suggest that the cross-section production function makes more sense if it is used to compare two similar but quite separate economies, for then competition need not require factor returns to be equal in the two countries even if the state of the arts is identical in identical industries. This idea is followed up in an ingenious way and leads to estimates of the Cobb-Douglas elasticities which add up to 1.27, a strong case of increasing returns to scale.

This is where I have to find fault. I don't think this device really dodges the nonsense in the cross-section approach. The results suffer from much the same difficulty of interpretation and may in fact mean something quite different from what they appear to say. I think I can show that under plausible (though not necessarily true) assumptions, universal constant returns to scale might lead to just such an appearance of increasing returns.

The difficulty goes back to the ordinary cross-section technique of Douglas and his followers. Suppose there *is* competition everywhere, and suppose every industry operates under constant returns to scale. Suppose you plot output per man-hour against capital per man-hour, one point for each industry, which is essentially what the cross-section technique does. Then what you get is not meaningless, but it is simply not a production function. The points should lie on or near a straight line whose vertical intercept should be the real wage and whose slope should be the real return on capital. This line will be the envelope of the production functions of the several industries. You can fit a log-log function to it if you're in the mood, but you're wasting your time. This theoretical point is surely well known. There are some remarkable recent results by Tibor Barna which show that nature imitates art, and what is theoretically expected actually shows up in the data.

Now let us turn to the international comparison. In a notation as close to Stigler's as I can conveniently make it, let Q_i^e and Q_i^a be the outputs of the i^{th} industry in England and America. Then if each industry has a Cobb-Douglas production function, the same in both countries, we can write

$$Q_i^e = L_i^{\alpha_i} C_i^{\beta_i}$$
$$Q_i^a = b(\lambda_i L_i)^{\alpha_i} (\mu_i C_i)^{\beta_i}$$

where L_i and C_i are the labor and capital inputs into the i^{th} industry in England and λ_i and μ_i are the ratios of American labor and capital inputs to English in the i^{th} industry. It would, of course, be a gross coincidence if every industry's technology were describable in this constant elasticity form. I suspect the Cobb-Douglas function has been grossly overdone (and I have done my share of overdoing). But at least I am allowing each industry its own elasticity. I am sure Stigler is under no illusions about the meaning to be attached to the single interindustry α and β in his cross-section function.

Suppose there is approximate competition everywhere. If there are strong increasing returns to scale it is hard to make sense out of an assumption of competition—unless the economies of scale are all purely external, which is in itself not an attractive assumption and could in any case pose very difficult analytical peculiarities. But my argument does not depend too much on the assumption of competitive product markets, so I make the assumption for convenience. Then we have the marginal productivity equations:

$$w^e = p_i^e \frac{Q_i^e}{L_i}\alpha_i \qquad w^a = p_i^a \frac{Q_i^a \alpha_i}{\lambda_i L_i}$$

$$r^e = p_i^e \frac{Q_i^e}{C_i}\beta_i \qquad r^a = p_i^a \frac{Q_i^a \beta_i}{\mu_i C_i},$$

where w^e, r^e, w^a, r^a are the wage rate and rental of capital services in England and America and p_i^e and p_i^a are the prices of the i^{th} commodity in the two countries. From these equations it is easily deduced that

$$\frac{p_i^a Q_i^a}{p_i^e Q_i^e} = \frac{w^a}{w^e}\lambda_i \quad \text{and} \quad \frac{p_i^a Q_i^a}{p_i^e Q_i^e} = \frac{r^a}{r^e}\mu_i.$$

Finally we can conclude that

1. $$\frac{w^a}{w_e}\lambda_i = \frac{r^a}{r^e}\mu_i \quad \text{or} \quad \frac{\mu_i}{\lambda_i} = \frac{w^a r^e}{w^e r^a};$$

2. $$\frac{p_i^a Q_i^a}{p_i^e Q_i^e} = \sqrt{\frac{w^a r^a}{w^e r^e}} \quad \sqrt{\lambda_i \mu_i}.$$

The first of these states, remarkably enough, that the ratio μ_i/λ_i (the ratio of capital per worker in the i^{th} industry in America to capital per worker in the same industry in England) is independent of i; that is, it is the same for all industries. A glance at the first column of Frankel's Table 5 indicates at once that the data do not behave in this way. I suggest that among the reasons they do not are the following:

1. English and American industries do not always have the same production function.

2. The degree of monopoly may differ from industry to industry and country to country.

3. Fuel input may be a poor measure of capital services; dentistry may be more capital intensive than marshmallow-toasting.

4. In many industries the production function may be quite different from the Cobb-Douglas, and this result depends on the constant elasticity form.

5. Different industries may vary in the average skill level of their work forces and therefore it is a mistake to measure labor input by employment and another mistake to imagine each industry as paying the same wage rate even under competitive conditions.

The second conclusion above states that if ratios of values of output for each industry are plotted against λ_i and μ_i, the points should fall along a Cobb-Douglas surface of degree one, in fact with equal exponents. But these are monetary values whereas I understand Frankel's figures are attempts at physical output measurement. If value ratios behave in this special way, how might output ratios Q^a_i/Q^e_i be expected to behave? That depends on the observed relation between the price ratio p^a_i/p^e_i on one hand and λ_i and μ_i on the other. If industries with relatively high values of λ_i and μ_i are associated with relatively low values of p^a_i/p^e_i, i.e., if America tends to specialize in commodities for which it has a price advantage; then for high values of λ_i and μ_i, Q^a_i/Q^e_i will tend to be rather higher than the value ratio, and for low values of λ_i and μ_i, Q^a_i/Q^e_i will tend to be rather lower than the value ratio. Thus both the Law of Demand and the Principle of Comparative Advantage seem to suggest that a regression of the output ratio for various industries against λ_i and μ_i will be biased in the direction of increasing returns even if each underlying production function exhibits constant returns to scale.

I am afraid I may have left two misleading impressions, which I would like to correct.

First, I do not believe that the Frankel data or similar figures will bear the refined interpretation I have put on them. I have pointed out one way in which they fail to behave as they theoretically should and I have suggested a number of reasons (all of which I believe to be empirically true) why they may be deficient for this kind of theorizing. This is no reflection at all on Frankel, who never intended the figures for this purpose. What I have been trying to show is that Stigler's attempt to use a purified cross-section production function approach to the measurement of economies of scale escapes some but not all of the pitfalls. The results must be handled cautiously.

Second, I am not trying to debunk increasing returns to scale. As an economic theorist I would gladly pay tithes to a Society for the Preservation of Constant Returns to Scale, and consider the money well spent. But analytical convenience aside, the way of the world is surely quite different. I agree completely with Stigler about the potential magnitude of scale effects, and about their probable actual importance. The problem of measuring economies of scale and distinguishing their effects from those of technical progress is an econometric puzzle worthy of anybody's talents.

REPLY: George J. Stigler.

Because of Frankel's kindness in supplying the value data underlying his study, it is possible to evade one of Solow's criticisms. One may compare relative prices with relative outputs, and thus disregard the multiplicity of inputs (which is an unwelcome complication in this context). The resulting equation is

$$\log\left(\frac{\text{price in U.S.}}{\text{price in G.B.}}\right) = 2.005 - \underset{(.075)}{.263} \log\left(\frac{P_a}{P_e}\right),$$

with $n=24$. The equation implies a production function of degree $1/(1-.263)=1.34$, which is substantially identical with that in the text.

Solow's main criticism, that the demand conditions can produce a spurious finding of increasing returns even when constant returns prevail, is of course disquieting. International specialization is probably not a large source of bias; the commodities are primarily domestic in Frankel's sample. The effects of domestic demands, however, do not seem capable of easy summary. I would welcome a showing of some bias because the present estimates of increasing returns are embarrassingly large.

POSTSCRIPT: Robert M. Solow.

Controversy over the source and nature of increasing returns to scale goes back a long way in the literature of economics. Not many of Clapham's "empty boxes" have been filled. To the extent that there are unexploited economies of scale internal to individual firms, data pertaining to individual firms should be studied. Stigler has suggested to me that it is also worth studying economies of scale of various degrees of externality, and I think he is right. In particular he has suggested one of the classical cases, in which economies of scale are external to each firm in the industry, but internal to the industry as a whole.

In this case, my equation (1) when written for each firm in the i^{th} industry would show constant returns to scale ($\alpha_i+\beta_i=1$), but the multiplying constants a and b would be increasing functions of the output of the industry, i.e., of the sum of firms' outputs. This implies that the aggregate production function for the industry would exhibit increasing returns to scale. This in turn suggests that equations like (2), if they had to be written for the industry, would overexhaust the product.

On these assumptions I think the proper way to proceed is slightly different. Since the economies of scale are external to the firm, it is still permissible to suppose that each firm takes the prices of factors and the output of the industry as parameters. Then for each firm (supposed equal in size, for simplicity) the usual value-of-marginal-product conditions will hold. In addition, if equilibrium is to exist for the industry, the price of the product must equal the unit cost of production. It can then be shown that equations (2) undergo, for the industry, only a slight modification. They become:

$$\frac{w^e}{1-\gamma_i^e} = \frac{p_i^e Q_i^e \alpha_i}{L_i} \qquad \frac{w^a}{1-\gamma_i^a} = \frac{p_i^a Q_i^a \alpha_i}{\lambda_i L_i}$$

$$\frac{r^e}{1-\gamma_i^e} = \frac{p_i^e Q_i^e \beta_i}{C_i} \qquad \frac{r^a}{1-\gamma_i^a} = \frac{p_i^a Q_i^a \beta_i}{\mu_i C_i}$$

where γ_i^e and γ_i^a are respectively the elasticities of *a* and *b* with regard to the industry output.

Then the first of my two numbered conclusions holds exactly as it stands and the second is modified only by the appearance on the right-hand side of the additional multiplicative factor $(1-\gamma_i^e)/(1-\gamma_i^a)$. If anything, this reinforces the conclusion, since one would expect λ_i and μ_i to be large in those industries for which γ_i^a is large relative to γ_i^e.

MORRIS A. COPELAND, Cornell University

This Conference was started some twenty-odd years ago to afford workers in the field of national income and social accounting, both the producers of the figures and the users, an opportunity to meet and discuss problems of common interest. As time has gone on the scope of topics considered at Conference sessions has gradually expanded. It currently includes about everything that comes under the head of empirical aggregative economic inquiries.

This expansion of Conference interest is a natural and, I think, a desirable one. The statistical measurements with which aggregative economics concerns itself consist of estimates of social accounting magnitudes like GNP and national wealth, as well as other closely related quantities. All things considered the Conference may well take its major objective to be fostering and facilitating statistical investigations of a macroeconomic nature.

It is probably inevitable that the process of expansion of the scope of Conference interest should have gone even farther. Macroeconomic inquiries are not all statistical. There is another kind that I venture to call a priori model analysis because of the tenuous connection between the neoclassical economic models it investigates and the real world. On a number of occasions, participants in Conference sessions have engaged in a priori model analysis. Such analysis was particularly in evidence in the 1958 sessions, and it is this fact that prompts the following comments.

I do not mean to suggest that such analysis should be excluded from Conference proceedings in future. Exclusion would be contrary to the spirit of intellectual freedom. I urge that such analysis be clearly recognized for what it is, and be clearly labeled, because the distinction between a priori analysis and empirical research seems to have been particularly blurred in the 1958 Conference proceedings.

It would require lengthy comment to do an adequate job of grade labeling. In the interests of brevity, I have decided to discuss a single example. The example I have chosen is the second, third, and fourth sections of Stigler's paper. I have made this choice with three considerations in mind. First, among those who have devoted themselves extensively to neoclassical model analysis Stigler stands out as an especially careful thinker. Second, he has made, over a period of years, a particular effort to find some connection between his models and the world of fact. And third, his theoretical pronouncements are often couched in arrogant language that seems to imply there is no competent economist who disagrees with him.

In his first section, Stigler discusses the use of labor productivity measures as substitutes for capital-and-labor productivity measures. He deals with this topic in the spirit of empirical science, and his findings are significant.

The remaining sections of his paper are concerned with the following hypothesis: Ratios of aggregate physical volume of output to physical volume of input are crude in the technical statistical sense. An upward trend in such a crude ratio can be resolved into four main parts: (1) changes in input price relationships; (2) increases in the competitive nature of the economy (or decreases in its monopolistic nature); (3) economies of scale resulting from mere increases in the size of the economy; and (4) "pure" economic progress resulting from a change in the "state of the arts."

Much of his discussion runs in terms of models for which the equations are specified only in broad terms, such as the signs of the first and second derivatives of functions and restrictions that confine the analysis to real values of the variables. Such a priori model-analysis is not specific in the sense that each equation is given a specific analytical form with parameters that can be determined as least squares fits.

Of course Stigler gives us equations of this very specific form, too, and parameters that have been determined as best fits. If this hypothesis is valid and significant, it seems reasonable to insist that

NOTE: These comments were not written until after the Conference sessions.

it lend itself to exploration in terms of specific equations of the best-fit kind and that such equations provide empirical support for thinking that each of the four components has, for some period of years, a value materially different from zero. The question I am particularly concerned to consider is, Does he present statistical evidence that his input-price-change component, his monopoly-vs.-competition component, and his economies-of-scale component do in fact assume values materially different from zero?

In the second section Stigler finds that "changes in relative [input] prices will, with the usual index-number formulas, lead to some change in reported efficiency, although none has taken place." If one grants that a firm can be counted on to behave as if it operated with a production function of the type implied in his Chart 3, this conclusion is logically inescapable. But Stigler might have postulated behavior in accordance with a production function in which capital and labor are perfect complements. On this assumption no change in relative input prices would lead to any change in efficiency. Therefore, the extent of the influence of changes in relative input prices on reported efficiency depends on the nature of the production functions to which business behavior in fact conforms. The more closely they approximate perfect complementarity, the more nearly negligible the influence. But Stigler makes no attempt to show that the influence is material.

In an imperfectly complementary case the influence of a change in relative input prices on reported efficiency could be either plus or minus. Conceivably then, even assuming far from perfect complementarity, the influence on an aggregate measure of efficiency for, say, all manufacturing industries might be negligible because, with changing techniques, the pluses offset the minuses. The second does not consider this possibility.

Stigler refers to input prices as "one part of our observational information." This seems to imply that satisfactory measures for both prices are available. I dare say the difficulties in providing an aggregate statistical measure of the cost of labor are not too serious, but I wish Stigler would give us his solution of the problem of providing a statistical measure of the cost of capital.

Next as to his monopoly-vs.-competition component of an increase in an output-input ratio. In theory at least, this component might be either positive or negative. Hence the exploration of his hypothesis seems to call for some measure of the extent to which there is movement toward more monopoly or toward more competition. Stigler finds that elimination of monopoly or monopsony always improves economic efficiency. As applied to his Charts 5 and 6 this proposition will doubtless command general agreement. As applied to the real world, however, where competition between selling enterprises includes quality of goods, terms of sale, and sales effort and where employer competition mixes wages and working conditions, this proposition is certainly somewhat controversial, yet there is nothing in Stigler's third section that relates to this influence in the real world. The argument proceeds exclusively in terms of a priori models with competition always conceived as "perfect competition."

Stigler may be right in thinking recent increases in aggregate U.S. output-input ratios reflect in part changes in the degree of monopoly or of competition in the economy, but he presents no empirical support, nor does he offer us any measure of this type of change. No doubt he is painfully aware of the difficulties of providing such a statistical measure. At any rate, he does not even attempt to say whether the change in recent years has been toward or away from a more competitive form of organization.

The third section of his paper is purely deductive from start to finish. And it seems fair to conclude that the part of his hypothesis with which it is concerned does not readily lend itself to empirical investigation in terms of the statistical data currently available.

Stigler's third component of increases in crude output-input ratios is the influence of "economies of scale." His hypothesis here is that a production function may be such that an increase of x per cent in each input may, "in the absence of progress," result in an increase of more than x per cent in output. In examining this possibility in connection with time-series data a need is observed to find some way to distinguish between (1) the increase in an output-input ratio that would have occurred with an increase in scale but no change in the production function (i.e., "in the absence of progress") and (2) the part of the actual increase in an output-input ratio that is attributable to "progress" (i.e., to a change in the production function.)[1] Likewise in an intercountry comparison which shows higher output-input ratios for Country A than for Country B there is need to find some way to say how much, if any, of this showing is due to differences in scale and how much is due to differences in production functions.

When Stigler concludes that "economies of scale are potentially of the same order of magnitude as technical progress," he presumably has in mind mainly what I would call aggregative economies of scale. It could be argued that his U.S.-British comparison in large part reflects economies of scale at the plant or enterprise level, and that

[1] Stigler refers to his equation, $P=bC^{1.36}$ (which does not make such a distinction), as a "production function." This use of the term is confusing. It might better be called a "state-of-the-arts function."

American businessmen have generally developed plants and business organizations of optimum size while their British cousins have not. But the hypothesis that differences in the state of the arts have resulted in differences in approximation to optimum size is more likely to appeal to those who do not have a pro-American bias. At all events, economies of scale at the industry level can hardly have contributed significantly to the finding of an $(\alpha+\beta)$ of 1.27. Such so-called external economies imply an input with a supply schedule that descends to the right. And practically the only situation that can yield this kind of schedule under anything remotely resembling perfect competition is one in which the input is itself a product made with an input that has a decreasing supply price. External economies that result from the larger size of the American economy seem a more plausible possibility than the one Stigler has in mind.[2]

Stigler attaches more weight to his intercountry comparison than to his Cobb-Douglas fit to Creamer's data.[3] But there is need in both cases to distinguish between differences in an output-input ratio that involve no change in production function and differences that are due to such a change. We do have a technique of sorts for making such a distinction in a time-series analysis, but no comparable technique for an intercountry comparison.

The technique I refer to requires us to assume that output is a particular analytical function of inputs and time. If for the moment we avoid being specific analytically, we must assume $P=f(C, L, t)$. Presumably we will expect $\partial P/\partial t$ to be >0, for we will take time as an indicator of the state of the arts. In other words, we will expect technological progress. Having fitted a function of this kind to data such as Creamer's, we can then hold time constant and with it the state of the arts, and investigate the way output changes with changes in inputs. Also we can hold the inputs constant and investigate the way output varies with time and the state of the arts.

There is an obvious objection to this procedure. The line is drawn according to the specific analytical form assumed for $P=f(C, L, t)$. However, if I propose for Creamer's data an analytical form that gives no suggestion of aggregate economies of scale, and if Stigler wishes to raise this objection, it would seem incumbent on him to offer an alternative form of $P=f(C, L, t)$ that gives at least as good a fit and does exhibit economies of scale.

A comment here on his U.S.-British comparison: He assumes that

[2] "The large economy can practice specialization in innumerable ways not open to the small (closed) economy."

[3] In fact he goes so far as to say, incorrectly I think, that "We cannot use time series on inputs and outputs to answer this question"—viz., "How important are the economies of scale of the economy . . . ?"

the output-input "function (not technique) of an industry is the same in both countries."[4] Since this function includes only C and L as independent variables, with no separate independent variable as an indicator of the more progressive state of the arts in the United States, his anlysis does not distinguish the influence of external economies from that of differences in the state of the arts, and does not indicate whether there are increasing or decreasing returns with scale. His comparison is, therefore, completely irrelevant to the conclusion that he seems to have drawn from it. This leaves the final section without pertinent statistical support.

The hypothesis I propose to fit to Creamer's data is extremely simple. I will take $\partial P/\partial C=0$, so long as there is any excess capacity in the economy. This means my hypothesis exhausts no more degrees of freedom than Stigler's did when he found that $\log P=\log(.19)+.46\log C+.90\log L$. My hypothesis is that output is an increasing function of time (as an indicator of the state of the arts) and of the economy's capacity factor (i.e., the percentage of capacity at which the economy is operating). I will take as an indicator of the capacity factor a deviation-from-trend computation for labor input. Specifically, I assume that $\log P=m\,\Delta\log L+ht+k$, where $\Delta\log L$ is the deviation from its linear trend of $\log L$. With t measured in years and $1900=0$, I find the following best fit:

$$\log P = .396\Delta\log L+.0148t+.9875.$$

This equation gives an appreciably better fit than does Stigler's.[5] And since .396 is markedly less than 1, it clearly suggests decreasing rather than increasing returns with scale.[6]

No doubt Stigler will regard the capacity factor as reflecting short-run disequilibria. My hypothesis certainly emphasizes short-run adjustments, but I protest his normative language. It is principally short-run adjustments that we have learned how to explore statistically. The idea of a functional relation between output and inputs that represents the long-run adjustments for various price situations and for a given state of the arts does not readily lend itself to statistical exploration. If Stigler has a way of exploring such a relation I wish he would tell us about it.

RAYMOND L. RICHMAN, University of Pittsburgh

In his section on "Labor Productivity as an Index of Productivity," Stigler complains that index makers began constructing indexes of

[4] Here again, the confusing term "production function." See footnote 1.

[5] The standard error of my predicted $\log P$'s is .081. The standard error of his predicted $\log P$'s is .141.

[6] It may be convenient to think of my formula in the following form:

$$P = \text{a constant} \times L^m \times C^0 \times f(t).$$

labor productivity when "it was already a basic proposition of economics that one should never look at *average* products, only at *marginal* products." Later he adds, "So far as I know, not a single theoretical statement of any importance can be made about the average product of labor."

These statements are completely misleading. All of the industry, sector, and national productivity measures are measures of long-run changes in productivity. Average product is of more theoretical importance in the long-run case than is marginal product.

First, in the long run under competitive conditions, price tends to minimum average cost. Average cost is composed of the prices of the factors of production divided by their average products.

Second, businessmen have long been pricing on the basis of their average costs. Wages and prices are set over long periods. In labor negotiations businessmen have been concerned with the effect of a wage increase on their average costs. This is not the place to discuss average cost pricing, but it is surely time that economists paid more attention to how the economic system works, although it is not irrelevant how it should work. Further, under some assumptions about the shape of the production function, they might achieve about the same results as they would if they were using marginal costs, e.g., where one assumes a production function (average product) "kinked" at the point of designed capacity.

Third, the production function implies certain precise relations between marginal productivity, average productivity, and total production. From what is happening to average product, what is happening to marginal product can be deduced. As a practical matter, no entrepreneur would ever know his production function by trying to determine marginal product without going through the intermediate stage of learning average product. Cost accountants would still be looking for clues if most economists, including Stigler, had not, perhaps by oversight since it has "no theoretical importance," included drawings of average product when they drew marginal product curves, and if the cost accountants did not know that profits were the difference between average costs and price, not marginal costs and price.

Perhaps most important, the exception which Stigler relegates to a footnote as a "polar case" is of fundamental importance in distributing income among the factors of production, and is pregnant with political and ideological significance. As he phrased it, "What propositions there are about average products—such as that average product should be maximized to maximize output if only one factor is scarce—are polar cases." To translate it to the problem of capital and labor inputs, as the proportion of capital inputs to labor increases, *ceteris paribus*, wages increase and interest decreases until the marginal product of capital is zero, where average product of labor and total product are maximized. A major economic goal is to maximize the average product of labor. The long-run change in the average product of labor is indeed an important measure of economic progress.

Stigler next discusses the accuracy of an index of labor productivity as an estimate of changes in total productivity, that is, capital-and-labor productivity. He draws the obvious conclusion that labor productivity is a better measure of total productivity the more nearly proportional the changes in the various inputs and the larger the labor input is to total inputs. In the individual firm, the capital inputs come largely from outside the firm. (Similarly, changes in the degree of subcontracting, of integration, of power, and other inputs represent changes in inputs from outside the firm.) Under the conditions applying to the individual firm, labor productivity is a good index of total productivity only under the conditions specified by Stigler. However, in the case of a closed economy where all inputs come from within the economy, an increase in capital is not an input independent of labor input. The condition that labor be a large input relative to total input is always satisfied, because most of the capital input is itself a labor input. To represent capital as an independent input is to understate the gain of the economy by double-counting some of the labor input.

ROBERT EISNER, Northwestern University

I shall address myself only to the section of Stigler's paper in which he indicates that certain empirical data collected by Marvin Frankel suggest increasing returns to scale. While I shall call into question the weight to be attached to this suggestion, my concern is not the substantive issue, about which I offer no judgment. Rather, I should like to take Stigler's estimates as a point of departure for an exposure of some pitfalls in presuming to estimate parameters of an economic relation when that relation is not adequately specified.

Utilizing a cross section of twenty-three industries with data as to output and labor and capital inputs in the United States and the United Kingdom, Stigler derives estimates of parameters of a Cobb-Douglas-type production function which indicate that *a* plus *b*, the sum of the labor and capital coefficients, equals 1.27. More precisely, where P_a is output in an industry in America and P_e the corresponding output in England, and λ_L and λ_C are the ratios of American to English inputs in the industry for labor and capital respectively, Stigler finds that

$$P_a/P_e = 1.45\lambda_L^{.827}\lambda_C^{.444}.$$

But now suppose each industry has resources or factors specific to its output and different countries are differently endowed. We may classify these other resources as "land," "know-how," or entrepreneurship and label them R_{ij}, where i refers to the i^{th} industry and j goes from 1 to n. Then, if the production function, or functions, are of the Cobb-Douglas type, they may really be given by

$$Q_i = aL_i^{\alpha}C_i^{\beta}\left(\prod_{j=1}^{n} R_{ij}^{\gamma_j}\right) + u_i,$$

where u_i is the usual stochastic term and the other variables are as Stigler has defined them.

But if these other factors or resources contributing to production do exist and this is the form of the function, it must follow that $\gamma_j \geqslant 0$ for all j and hence that the partial derivatives of output (Q) with respect to each factor as well as all cross partial derivatives must be positive. Thus industries relatively well-endowed with other resources (R_j) would have higher marginal product curves for labor (L) and capital (C). Because of imperfect mobility the remuneration of labor and capital would tend to be higher in industries relatively well-endowed with other factors, and the ratio of labor and capital to these other factors would tend to be relatively low. This is all that is necessary to establish that industries well-endowed with other factors relative to the same industries in another country would attract relatively more labor and capital and have a relatively higher ratio of output to the sum of labor and capital inputs. (This last is analogous to Stigler's demonstration that a relatively higher capital input would raise the average product of labor.)[1] It follows, under these circumstances, that there would be an upward bias to the estimates of the sum of the labor and capital coefficients ($\alpha+\beta$) and to the estimate of the degree of the Stigler-Frankel production function. Until the amount of this bias can be calculated, the evidence that Stigler presents for increasing returns cannot therefore be accepted as valid.

[1] Page 48 and footnote 2.