COSTS AND OUTPUTS

Armen Alchian

P-1449

3 September 1958

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

COSTS AND OUTPUTS

by

Armen Alchian[1]

Obscurities, ambiguities and errors exist in cost and supply analysis despite, or because of, the immense literature on the subject. Especially obscure are the relationships between cost and output both in the long run and in the short run. Propositions designed to eliminate some of these ambiguities and errors are presented in this paper. More important, these suggested propositions seem to be empirically valid.

I. Costs

Costs will be defined initially as the change in equity caused by the performance of some specified operation, where, for simplicity of exposition, the attendant change in income is not included in the computation of the change in equity. Suppose that according to ones balance sheet the present value of his assets were $100, and suppose that at the end of the operation one year later the value of his assets were expected to be $80, not counting the sale value of the product of the operation. The present value of $80 a year hence (at 6 per cent) is $75.47, which yields a cost in present capital value (equity) of $24.53. Because of logical difficulties in converting this present value concept into a satisfactory rate (per unit of time) concept, we defer a discussion of this problem

---

[1] University of California, Los Angeles, and the RAND Corporation. Indebtedness to William Meckling of the RAND Corporation, who gave many long hours to discussion of the points raised herein, even before the first of several drafts, is very great. Although my egoism prevents sharing the authorship with him, I cannot absolve him from responsibility for any errors that still remain and likewise for any merit the paper may have.

and (for convenience) measure costs in units of present value or equity. Hereafter the unmodified expression "costs" will always mean the present worth, capital value concept of cost, i.e., the change in equity.

II. Output

All the characteristics of a production operation can affect its cost. In this paper we want to direct attention to three characteristics: (1) The rate of output is typically regarded in economic analysis as the crucial feature. But it is only one feature and concentration on it alone has led to serious error, as we shall see. (2) Total contemplated volume of output is another characteristic. Is a cumulated output volume of 10,000 or 100 or 1,000,000 units being contemplated? Whatever may be the rate of output, the total volume to be produced is a distinct feature with important effects on cost. Of course, for any rate of output the larger the total cumulated volume to be produced, the longer the operation will continue. Hence, incorporated in this description of total output is the total time length of the programmed production. Will it span one month or one year, or (at the other extreme) is the contemplated total volume so large that at the rate of output an indefinitely long time is allowed to the production run? (3) The programmed time schedule of availability of output is a further characteristic. For a point output the programmed date of the output availability is sufficient, but for outputs which continue over time the time profile (delivery schedule) of the output replaces a single date. We shall call these three distinct aspects the output rate, the contemplated total volume, and the programmed delivery dates.

These three characteristics can be summarized in the following definition, which also defines a fourth characteristic, m, the total length of the programmed schedule of outputs:

$$V = \int_{T}^{T + m} x(t)dt.$$

In this expression V is the total contemplated volume of output, x(t) the output rate at moment t, T the moment at which the first unit of output is to be completed, and m the length of the interval over which the output is made available. Of these four features only three are independently assignable; the fourth is then constrained. Unless specific exception is made, in the following we shall always discuss changes in only one of the features, V, $\chi(t)$, and T, assuming the other two constant and letting the full compensatory adjustment be made in m.[2]

III. Propositions About Costs and Output

Our task is now to make some propositions about the way costs are affected by changes in these variables. Letting C denote costs (i.e., the change in equity) - we have

$$C = F\ (V,\ x,\ T,\ m)$$

subject to the definition of V, which constrains us to three degrees of freedom among the four characteristics.

PROPOSITION 1:

$$\left.\frac{\partial C}{\partial x(t)}\right|_{\substack{T = T_o \\ V = V_o}} > 0 \qquad (1)$$

---

[2] We note that time or dating enters in a multitude of ways. There is the date at which the delivery of output is to begin. There is the period of time used as a basis for the measure of the rate of output, i.e., so many units per day, per week or per year. And there is the total time over which the output is to be made available.

The left hand expression is the derivative of the costs with respect to x, when T and V are held constant, letting m make up the adjustment. It shows the change in costs when the rate of output is increased without increasing V and without changing the delivery date but with an appropriate reduction of m. Proposition 1 states that the faster the rate at which a given volume of output is produced the higher its cost. We emphasize that cost means the change in equity, not the rate of costs.

PROPOSITION 2:

$$\left.\frac{\partial^2 C}{\partial x^2}\right|_{\substack{V = V_0 \\ T = T_0}} > 0 \qquad (2)$$

The increment in C is an increasing function of the output rate. This is a proposition about increasing marginal cost in present value measure, and is usually derived as an implication of efficient allocation of scarce heterogeneous resources among alternative uses.

Its validity, however, does not depend upon the validity of the premises of the classical model. For example, inventories need not increase in proportion to the rate of output if the variance of random deviations in output rates does not increase more than proportionally to the expected output rate. In this event, a sufficient condition for proposition 2 as derived by the classical model would be upset. But destruction of sufficient conditions does not eliminate the possibility of all necessary conditions being fulfilled; thus, even if the classical model's assumptions are upset, the proposition could still be true. Whether or not it is in fact true cannot be settled by an examination of the model from which it is derived. For present purposes proposition 2 can be regarded, if one wishes, as a postulated proposition.[3]

PROPOSITION 3:

$$\left.\frac{\partial C}{\partial V}\right|_{\substack{x = x_o \\ T = T_o}} > 0 \qquad (3)$$

C increases with V for given x and date of initial output, T. At a constant output rate, for example, this will require a longer program of production - a larger m.

PROPOSITION 4:

$$\left.\frac{\partial^2 C}{\partial V^2}\right|_{\substack{x = x_o \\ T = T_o}} < 0 \qquad (4)$$

Increments in C diminish as V increases, for any rate of output, x, and initial output date, T. Thus, for any constant rate of output, as the total planned output is increased by uniform increments, costs (changes in equity) will increase by diminishing increments. The "reasons" for this proposition will be given later.

---

[3] See Whitin, T. M., and M. H. Peston, "Random Variations, Risk and Returns to Scale," Quarterly Journal of Economics, LXVIII, Nov., 1954, pp. 603-614, for a longer discussion of some forces that could reverse the inequality of proposition 2. Some of their suggested forces, e.g., relation between stocks of repairmen and number of machines, are circumvented by the ability to buy services instead of the agents themselves. Another weakness is the association of size of output with the number of independent random forces.

The above proposition also implies decreasing cost per unit of total volume, V. We shall state this as a separate proposition.

PROPOSITION 5:

$$\left.\frac{\partial\, C/V}{\partial V}\right|_{\substack{x = x_0 \\ T = T_0}} < 0 \qquad (4a)$$

IV. Graphic and Numerical Illustrations of Propositions 1-5

1. Graphic Illustration. The above properties are shown by the cost surface in Figure 1. Proposition 1 describes the slope of a slice on the cost surface where the slice is parallel to the Cx plane. Proposition 2 states that the slope of the path of such a slice on the cost surface increases with x. Proposition 3 is portrayed by the slope of a slice along the surface parallel to the CV plane - going back into the page. The slope of this slice decreases as V increases. Proposition 4 describes the decreasing rate at which this surface of costs increases along this slice. Movements in other directions can be visualized. For example, one possible path is to start from the origin and move out some ray. This gives costs as a function of proportional increase in both the rate and the total output for a fixed interval of production, m. But the behaviour of the cost slope of this slice, except for the fact that it is positive, cannot be derived from these propositions.

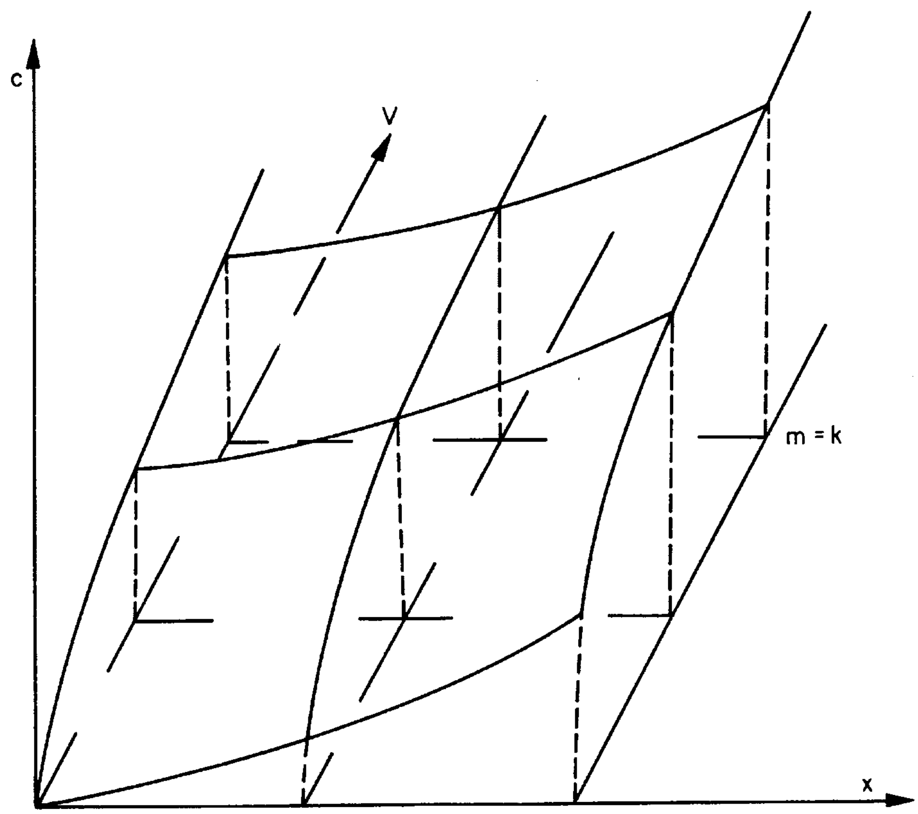

Fig. 1 — Cost surface as function of x and V

2. Tabular, Arithmetic Illustration.

Table 1

Costs, Volume of Output, and Rates of Output

		Volume of Output				
		1	2	3	4	...
Rate of Output, x	1	100	180	255	325	
(per year)	2	120	195	265	330	
	3	145	215	280	340	
	4	175	240	300	355	

For an output rate, x, of one per year, beginning at some specified T, production must continue for one year to get a total volume, V, of 1, for two years to get 2, three years for 3, etc. For a production rate of 2 per year, production must last one year to get 2 units, two years to get a total of 4, etc. The present value of costs for an output rate, x(t), of 2 a year for a total, V, of 4 in two years is $330 (which, at 6 per cent, is equal to a two-year annuity of $180 a year).

Proposition 1 is illustrated by the increase in the numbers (costs) in cells down a given column. Proposition 2 is illustrated by the increases in the differences between these cell entries. These differences increase as the rate of output increases, for a given total output. This represents increasing marginal costs (remember that cost is a present value capital concept) for increments in the rate of output. Proposition 3 is illustrated by the behaviour of costs along a row (given output rate) as total volume of planned output changes. Proposition 4 states that the increment in C is a diminishing increment as one moves across a row, *i.e.*, as total volume of output is larger. For example, in the first row, the output *rate* is one a year. The first cell is therefore an output operation lasting one year because only one is produced, at the rate of one a year. The total cost is $100. For a total contemplated output of 2 units, at a rate of one per year, the operation will take two years and the cost is $180. The marginal cost of one more unit of total volume of output - not of one unit higher rate of output - is $80. For a total output of 3 units in three years, the cost is $255, an increment of $75, which is less than the previous increment of $80. Here the increments in cost are associated not with increments in rates of output, but with increments in total volume of output. Proposition 5 is illustrated by dividing the cell entries in a row by the output quantities at the head of each column. The quotient, cost per unit of output quantity, decreases as V increases.

3. Economic Illustration. A comparison that could be made is the following. Imagine a person to contemplate a total volume of output of one unit at the rate of one a year. But he subsequently revises his plans and produces one more in the next year at the rate of one a year - again planning to produce a total volume of just one unit. Compare the total costs of that operation with an operation in which 2 units of total output were initially planned at the rate of one a year. Both take two years, but the cost of the latter is $180 while the former's present value is $100 plus $100 discounted back one year at 6 per cent, or a total of $194. Thus it is cheaper to produce from a plan for a two-year output of two units at the rate of one a year than to produce two by repetition of methods which contemplate only one total unit of output at the same rate of one a year.

From this example it would appear that a reason for proposition 4 is that better foresight enables one to see farther into the future and make more accurate forecasts. But this is not the reason, however helpful better foresight may be. A larger planned V is produced in a different way than a smaller planned V. A classic example is the printing press. To get three hundred copies of a letter in one day may be cheaper with mimeograph than with either typewriter or offset printing. The mimeograph method may be so much superior that even if the rate of output were stepped up to 300 in an hour (instead of in a day), mimeographing might still be cheaper than typing. This does not deny that higher rates of output imply higher costs, as for example that 300 in an hour will cost more than 300 in two hours. The method of production is a function of the volume of output especially when output is produced from basic dies. And there are few if any methods of production that do not involve "dies." Why increased expenditures on more durable dies should result in more than proportional increase of output potential is a question that cannot be answered, except to say that the physical principles of the world are not all linear (which may or may not be the same thing as "indivisible").[4] Different methods of tooling, parts design and assembly is the usual explanation given in the production engineering literature.

Proposition 4 seems not to be part of current economic principles. And yet it may be the key to seeing the error in some attempts to refute proposition 2, which applies to increased rates of output for constant total volume of output (or, as we shall see later, for perpetuity durations of output). Propositions 2 and 4 refer to two counter forces, rate of

---

[4] Could it be that the term "indivisibility" has been meant to cover this phenomenon? A yes or no answer to that question is impossible because of the extreme vagueness and ambiguity with which the term has been used. Furthermore the question is probably of little if any significance.

output and total planned volume of output. What would be the net effect of increases in both cannot be deduced from the present propositions. All that can be said is that if the rate of output is increased for a given total contemplated volume of output, the increment in cost will be an increasing function of the rate of output. Proposition 4 on the other hand implies diminishing increments as V increases and it implies a lower per unit cost for a larger total volume of output. Thus we have the possibility that higher rates of production might be available at lower unit costs if they are associated with a larger volume of output, because this latter factor may be sufficient to overcome the effects of the higher output rate.

A larger volume of output could, of course, be obtained by both longer time and faster rates of production, but the relationship between time and volume should not be allowed to mask the fact that it is total contemplated volume of output - not the longer duration of output - that is here asserted (maybe erroneously) to be the factor at work in propositions 3 and 4.

If both the volume and the rate of output change in the same direction, the two effects on costs are not in the same direction. And neither the net effect on the rate of change of increments in the cost, nor even the effect on the costs per unit of total volume of output, is implied by these or any other accepted postulates. It has been said for example, that if some automobile manufacturer were to cut V, the volume of cars produced of a given year's model, one million to a half a million, costs per car would increase. This statement could refer, either to a reduction in V achieved by producing for half the number of months at an unchanged monthly rate of output, or to a simultaneous and parallel reduction in both V, the volume, and x, the monthly rate of output. If it refers to the former, it is a restatement of Proposition 5; if it refers to the latter it is a statement that cannot be deduced from our propositions, which imply merely that costs would be lower if both V and x were reduced than if V alone were lowered.

Even returns to scale seem to have been confused with the effect of size of output. It is conjectured that a substantial portion of the alleged cases of increasing returns to scale in industries or firms is the result of ignoring the relation of costs to Volume (rather than rate) of output. The earlier discussions of automobile production and printing costs are simple illustrations of how this confusion can occur.

How many of the cases of alleged decreasing costs to rates of output are really decreasing costs to volume of output is an open question. Is it too much to expect that all of them can be so explained? Or that the realm of such cases can be greatly reduced by allowing for V, instead of letting x be the only variable? But that dirty empirical task is left for later.

The observed concentration on a standardized model, *e.g.*, four or five different sizes of tractors as distinct from a much greater possible range is explained by the effect of volume of output on cost. Although an infinite range is possible the concentration on a fewer smaller set of alternatives is more economical for most problems. The only way economic theory heretofore could explain this apparent anomoly was to invoke a falling cost curve for small output rates, in turn dependent upon some kind of unidentified indivisibility or returns to scale. Now the explanation may be contained in propositions 4 and 9.

V. More Propositions

Four more propositions remain. Proposition 6 is given in a footnote because its implications will not be suggested in this paper.[5]

---

[5] Proposition 6:

$$\left. \frac{\partial^2 C}{\partial x \, \partial V} \right|_{T = T_0} < 0 \qquad (6)$$

This says that the marginal present value-cost with respect to increased rates of output decreases as the total contemplated output increases. This can be regarded as a conjectural proposition whose implications will not be developed in this paper. And the same proposition can be re-expressed as

$$\left. \frac{\partial^2 C}{\partial V \, \partial x} \right|_{T = T_0} < 0 \qquad (6)$$

This states that marginal present-value costs of increased quantity of output decrease as the rate of output increases.

Of interest is the relationship between these postulates and the implied shape of the production possibility function where the rate and the volume of output are the two output alternatives. The cost isoquant with x and V as the arguments can be convex or concave. Usually a concave function is implied when rates of output of two different products are the arguments. However, Hirshleifer, J., "Quality vs. Quantity; Cost Isoquant and Equilibrium," Quarterly Journal of Economics, LXIV, Nov., 1955, pp. 596-606, has pointed out that convex production possibilities are implicit in many engineering cost functions when quality and quantity are the alternative outputs. Hirshleifer, as it seems from his context, is really discussing cases where his quantity variable refers to volume and not rate of output. Had he really meant rate of output rather than volume, his results might not have been so "reasonable." The convexity or concavity of the cost isoquant, it may be recalled, is given by the sign of

$$\frac{d^2x}{dV^2} = \frac{F_{xx}F_y^2 - 2\,F_{xv}F_xF_v + F_{vv}F_x^2}{F_v^3}$$

Substituting our postulated conditions shows that the expression may be of any sign, hence the indeterminacy of the concavity or convexity property. However, concavity of the cost isoquant where the two arguments are rates of production for two different products is still implied.

Propositions 7 and 8 concern the effects of changes in T, the time between the decision to produce and the delivery of output.

PROPOSITION 7:

$$\left.\frac{\partial C}{\partial T}\right|_{\substack{x = x_0 \\ V = V_0}} < 0 \qquad (7)$$

This is not shown in the graph or in the table, but it says that the longer the time between decision to produce and delivery of output, the less the cost.

If we think of a single output point, then T is relatively unambiguous. If the output is to be made available over a period of time, then T could be defined as the beginning moment of output. But many different output programs are possible, even when they all extend over the same interval. One might be tempted to use some sort of average T, say, the date of output weighted by the rate of output. But such an average T cannot be used for our purposes, because any particular value of T can be identified with an infinite variety of output patterns. Since we are talking about partial derivatives, the whole problem is easily avoided. All we need do is to state that if one moves any output program or schedule closer to the present (or farther into the future) by a simple time shift, T will have decreased (or increased). Whatever the shape of the output schedule, a reduction of the interval between the present moment and the beginning of the output date (a sort of uniform time-wise shifting) will increase cost. A more deferred output schedule (whatever its unchanged shape) will mean a lower cost.

Proposition 7 is really a corollary of proposition 2. The slower the rate at which inputs are purchased, the lower their price because the lower are the costs to the seller, when proposition 2 is applied to the seller.

Not only do the supply curves of inputs fall (or shift to the right) as more time is allowed but the rates of shifting differ among inputs. The supply curves of some inputs are more elastic than those of others; and the rate at which the price elasticity of supply increases with T differs among inputs. Thus, while in an immediate period the price elasticity of supply of input x may be low relative to that of input y (and it may always be lower than that of y), the <u>ratio</u> of the costs of increments in y to the costs of increments in x may change with deferred purchase. If the ratio decreases, deferred purchases of y relative to purchases of x will be economical. In other words, it is not merely the slope of the supply curve or the price elasticity of supply that determines which inputs are going to be increased earliest. Rather it is the rate at which these price elasticities <u>change</u> with deferred purchase that is critical. Thus, as stated earlier, the input x with a very low price elasticity of supply will vary more in the immediate period than the input of y with a higher price elasticity, if the deferment of purchases by, say, a month would lower the cost of y more than that of x. As an extreme, if the supply curves of two inputs x and y were both horizontal, the input of one of them would be increased less if with deferred purchase the price or supply curve would become lower - though still horizontal. That input whose price would become lower with a deferred purchase would be increased in quantity later with the relatively heavy present increase concentrated on that input whose deferred purchase price would not be as much lower.

PROPOSITION 8:

All the derivatives in propositions 1-5 are diminishing functions of T but not all diminish at the same rate. This proposition asserts a difference in the extent to which inputs will be varied in the immediate, the short and the longer period.

Short and long run. Statements to the effect that certain inputs are fixed in the short run are frequent and characteristic. In fact there is no such fixed factor in any interval other than the immediate moment when all are fixed. Such statements may represent a confusion between revealed choice and technological constraints. There are no technological or legal restraints preventing one from varying any of his inputs. Even in Viner's classic statement of the short- and long-run cost curves, the short-run is defined in terms of some fixed inputs and other inputs which can be varied as desired.[6] He stated that the long-run is the situation in which all the inputs are "freely" variable. One need only ask, "What do the desires to adjust depend upon in the short-run?" And what does "freely" variable mean? The first is answered by "costs" and potential receipts of the variations, and the second by noting that "freely" does not mean that costs of changes are zero. The fact is that the costs of varying the inputs differ among inputs, and the ratios of these costs vary with the time interval within which the variation is to be made. At any calendar moment, T, the producer will choose which input to vary according to economic costs and not because of technical or legal fixities that prevent the changing of some inputs.[7]

---

[6] Viner, J., "Cost Curves and Supply Curves," Zeitschrift fur Nationalokonomie, (1931), Vol. III, No. 1, 1932, pp. 23-46.

[7] The nearest, but still different, presentation of the immediate, short-run and long-run found by the author is that contained in Friedman's unpublished lecture notes. Other statements may exist; an exhausting search of the literature failed to clarify exactly what is meant by the long run and short run.

Debate over definitions or postulates is pertinent only in the light of their purpose. The purpose of the short and long run distinction is presumably to explain the path of prices or output (x or V?) over time in response to some change in demand or supply. The postulate of fixed inputs, and others more variable with the passing of time, does imply a pattern of responses that seems to be verified by observable evidence. On that count the falsity of the postulate is immaterial. But if there are other implications of such a postulate that are invalidated by observable evidence, the postulate becomes costly. The question arises therefore whether it is more convenient and useful to replace the fixity postulate by a more general one that yields all the valid implications that the former one did and more besides, while at the same time avoiding the empirically false implications. It appears that the proposed alternative is cheaper in terms of logical convenience, more general, and more valid in its implications. But that is a judgment that is perhaps best left to the reader.

The differences between a short-run (near T) and a long-run (distant T) operation imply differences in costs, and these costs are pertinent to an explanation of the path of prices or costs over time in response to a lasting change in demand or factor availabilities. For example, for a lasting increase in demand, the output made available at more distant dates is produceable at a lower cost; this means that the supply at a given cost will be larger and price lower in the more distant future as the longer-run operations begin to yield their output. These outputs, having been planned for a later T date, are lower in cost. Output will be larger for a given price, thus reducing price in the market. This longer-run lower cost is the phenomenon whose explanation has usually been sought by resort to fixity of some particular inputs in the short run. The above argument suggests that this phenomenon can be explained without the fixity assumption that in turn leads to other, empirically wrong, implications.

The implication of our proposition is worth emphasizing here. It is that we define a "short run" and a "long run" not as differing in the fixity of some inputs. Instead we use T as the length of the run, and then from proposition 8 derive the implications that were sought by the fixity assumption.

Most important, however, proposition 8 makes it clear that there is not both a "long-run" and "short-run" cost for any given output program. For any given output program there is only one pertinent cost, not two. Unambiguous specification of the output or action to be costed makes the cost definition unambiquous and destroys the illusion that there are two costs to consider, a short- and a long-run cost for any given output. There is only one, and that is the cheapest cost of doing whatever the operation is specified to be. To produce a house in three months is one thing, to produce it in a year is something else. By uniquely identifying the operation to be charged there results one cost, not a range of costs from immediate to short to longer run costs. There is a range of operations to be considered, but to each there is only one cost. The question is not, "What are the long-run or short-run costs of some operation?" but instead, "How do total, average and marginal costs vary as the T of the operation is changed?" Answer: "They decrease as T increases, according to propositions 7 and 8."

The significance of this should be evident in the debate about marginal cost pricing policies for "optimal" output. Also the use of short-run and long-run costs as alternatives in public utility pricing appears to be a ripe area for clarification of concepts.

What the relationship is between the presently suggested effects of T which we have just called a short or long-run effect and the common short run or long run in the standard literature is not entirely clear. Rather vague and imprecise implications about short and long run are available. Hence, rather than assert that the T effect is here being proposed as a substitute for the standard short-run analysis, the reader is left free to supply his own interpretation of the convention "run" and to supplement or replace it, however he chooses, with the present proposition.

PROPOSITION 9:

The preceding propositions refer to costs of outputs for a given distribution of knowledge, F, at the present moment, to situations where technology is held constant.[8]

Proposition 9 is "As the total quantity of units produced increases, the cost of future output declines." The cost per unit may be either the average cost of a given number of incremental units of output or the cost of a specific unit. This is not identical with the earlier proposition 4 referring to the effects of a larger planned V. There the effect was a result of varying techniques of production, not of changes in technology. Here we are asserting that knowledge increases as a result of production - that the cost function is lowered. It is not simply a matter of a larger V, but rather a lower cost for any subsequent V, consequent to improved knowledge. This distinction should not be necessarily attributed to all the explanations of the learning curve. Some describers of the

---

[8] Technology, the state of distribution of knowledge, is different from techniques of production, which can be changed at any time even with a constant technology.

learning curve bring in the effect of different techniques consequent to different sized V. Others also mention that as output is produced and experience acquired, improved knowledge is acquired. Thus even if one continually planned to produce small batches of output, so the V was constant but repeated, the costs would nevertheless be falling. But in the present presentation we have chosen to separate these two effects in logic and principle, attributing the first effect, that of technique, to changes in planned V but with a given state of knowledge (as in proposition 4), while the second effect, that of increased knowledge consequent to accumulated production experience, is isolated in proposition 9. A review of industrial and production management literature will show that both effects are adduced and incorporated in the learning curve discussion, contrary to our decision to separate them. This proposition about the rate of change in technology is accepted in industrial engineering. Usually the proposition is known as the "learning curve" or "progress curve."[9]

Several factors have been advanced as a rationale for this proposition - job familiarization, general improvement in coordination, shop organization and engineering liaison, more efficient subassembly production, and more efficient tools. An extensive literature on this proposition has been developed but it seems to have escaped integration with the rest of cost theory in economics.[10]

---

[9] Sometimes the curve is called an 80 per cent progress curve, because it is sometimes asserted that the cost of the 2nth item is 80 per cent of the cost of the nth item. Thus the fortieth plane would involve only 80 per cent of the direct man hours and materials that the twentieth plane did.

[10] See Hirsch, W., "Manufacturing Progress Functions," Review of Economics and Statistics, XXXIV, May, 1952, pp. 143-155. A less accessible, but more complete, reference to the published material is given in H. Asher, Cost-Quantity Relationship in the Airframe Industry (The RAND Corporation, Santa Monica, California), July, 1956. But see Samuelson, Economics, McGraw Hill, New York 1948, p. 473, where it is mentioned but left incorporated.

Nevertheless the proposition is a well-validated proposition and is widely used in industrial engineering. The significant implication of this proposition is that in addition to rate of output, an important variable in determining total costs is the total planned output, for two reasons: first, because of changes in technique via proposition 4, and second, because the larger is the planned and ultimately realized output the greater is the accumulated experience (technology) and knowledge at any point in the future via proposition 9. Thus the average cost per unit of output will be lower, the greater is the planned, and ultimately experienced, output. A more complete discussion of the evidence for this proposition would require a separate paper.

VI. On the Advangages of the Capital Value Measure

Use of capital values enables one to avoid misleading statements like: "We are going to operate at a loss in the near future but operations will be profitable later." "In the short run the firm may operate at a loss so long as receipts exceed variable costs." "A firm operates with long run rather than short run objectives." All of these statements are incorrect if liabilities or assets (other than money) are owned by the enterprise. What seems to be meant when a person talks about expecting to have losses for a while before getting profits in cash flows will be negative for a while. But it is difficult to see how this is in any relevant sense a situation of losses. And similarly when a person talks about expecting losses it appears that he means that he expects future events to occur which are unfavorable; and in this case the changed belief about the future is immediately reflected in current values if not in current money flows...as many a stockholder has learned. Any period during which expectation about future events becomes more favorable are periods of increasing equity--*i.e.*, of profits, even though the period in which the more favorable events will occur is in the future. When a firm reports that it operated during the past quarter at a loss it means simply that the net present value of assets decreased during that period, even though the future cash receipts and outlays have not yet been realized. The profits are in the present moment--the increase in equity--as some stockholders have joyously learned. The presently anticipated increase in *future* receipts relative to future outlays means an increase in *present* equity values, profits.

Statements to the effect that a firm would willingly and knowingly operate at a loss in the short run are consistent only with an identification of costs with money flows--and are certainly inconsistent with the postulates of increasing wealth (or utility) as a goal or as a survival attribute. Such identification of costs with money flows eliminates capital theory from the theory of the firm and from much of price theory. There is no cause to pay this price since it is just as easy not to abandon capital theory and if one retains it more useful implications will be derived.

Yet in economic texts costs are almost always measured as *time-rates*, and only rarely as capital values. At first blush this would seem to be an irrelevant or trivial distinction since capital values are merely the present values of *receipt* or *outlay* streams. But what about going from capital values to time rates of *cost* streams? New problems arise in this effort. Suppose that the outlay stream for some operation is used as the basis for cost calculations. If, and only if, *no* other assets or liabilities are involved, can money flows be identified with costs; otherwise they represent, in part, accumulations of assets or liabilities. As soon as ~~assets and liabilities are admitted, money flows are not~~ synonymous with costs, and changes in values or assets or liabilities must now be included. With the break between money outlays and costs the measure of costs becomes the change in present value of net equity consequent to some action (ignoring receipts, for present purposes).

If a firm signed a contract and committed itself to produce some quantity of output, then the cost it has incurred in signing the contract and obligating itself to produce the output is its decrease in equity, say $E_a - E_b$. At moment a, prior to the contract, the equity or net wealth of the firm is E_a. At this moment the firm considers entering into some production plan. If it does so, what will happen to its equity at the end of the plan, or how will the equity change over that interval? If for the moment we ignore the receipts or income from that plan, the decrease of equity by moment b would be the measure of cost of the output operation which it is obligated to perform. The difference $E_a - E_b$, between the equity E_a at the beginning and the present value E_b of the equity (E_t) at the end of the operation, is the total cost, C, of the operation.

The time-rate of costs (of change in equity) is given by dE/dt, the slope of the line from E_a to E_t, which is quite different from C. The former, dE/dt, is a derivative, a time rate of change. The latter, C, is the integral of the former. It is a finite difference, $E_a - E_t$, obtained from two different points on the E curve, while the former is the slope of the E curve and can be obtained only after an E curve is obtained. What is the meaning of the E curve in such a case? Presumably it means that if the firm decided at any moment to stop further output, under this contract it would find itself with an equity indicated by the height of the line E_aE_t. Ignoring the contractual liability for obligation to produce ~~according to the contract, the equity declines along the E line; but if~~

one does regard the contract performance liability, the equity does not change as output is produced because there is an exactly offsetting reduction in contractual liability as output is produced. The equity of the firm stays constant over the interval if the outlays and asset values initially forecast were forecast correctly.

If the rate of cost, dE/dt, or if the E curve, is plotted not against time, but against the output rate, we do not get a curve similar in interpretation to the usual total cost curve in standard cost curve analysis. The rate of cost, dE/dt can be converted to average cost per unit of rate of output by dividing the rate of cost dE/dt, by the associated rate of output at that moment. And the marginal time-rate of cost is obtained by asking how the slope of the equity curve dE/dt is affected by changes in x, i.e., $d^2D/dtdx$.

The difference between this curve, where dE/dt is plotted against x, and the usual time rate of cost curve analysis is that our current analysis is based on a larger set of variables, $x(t)$ and V, and hence dE/dt cannot be drawn uniquely merely against the rate of output, $x(t)$. A new curve must be drawn for each output operation contemplated; even worse, there is no assurance that such a curve of dE/dt drawn against the rate of output on the horizontal axis would have only one vertical height for each output rate. The curve might fold back on itself and be multivalued because one value of dE/dt might be associated with a particular rate of output early in the operation and another different value later in the operation, even though at both moments the output rate were the same.

The number of cost curves that could be drawn is greater by at least an extra factor, V. In fact we have at least two families of curves, one ~~for different values of V and one for different time profiles of x(t)~~

And it is not clear what is usually assumed about these in the standard cost curve analysis. One possibility is to assume that the length of the production run m or the contemplated total output V does not affect the rate at which equity changes for any given output rate. The difficulty with this position is not merely that it is logically wrong but that it leads to implications that are refuted by everyday events.

A kind of average or marginal cost can be defined on the basis of the approach suggested earlier. For any given contemplated operation, the change in equity implied can be computed and evaluated in present worths. If this cost is divided by the total contemplated volume of out,put, V, the result is present value cost per unit of product (not time rate per unit rate of output). If the same total output were to be produced at a higher output rate x, and thus within a shorter time interval, m, the total cost (change in equity) would be greater and so the cost per unit of total volume of output would be higher. As noted in the first part of this paper the increase in total present value cost, $\partial C/\partial x$ (not $d^2E/dt\,dx$) is the marginal cost, consequent to an increased rate of output. By varying the contemplated rates of output x, for any given total output plan, (V and T), one can get different total capital costs. These changes in total capital costs can be called the marginal capital costs. But it is important to note again that there are as many such marginal capital value cost functions as there are different possible total output patterns that can be contemplated and these marginal capital costs are not time rates of costs.

VII. Conclusion

Four features have been emphasized in the foregoing pages. First, the distinction between rate and quantity of output; second, changes in technology as distinct from changes in technique; third, the use of calendar time dates of output instead of technical fixity for distinguishing output operations; fourth, the use of capital value concepts instead of rates of costs.

The first and second features and the ones that are emphasized in this paper, enable us to capture within our theory the lower costs attendant to larger quantities of output - not rates of output. Everyday experience where large rates of output are available at lower prices could be explained as a movement down the buyer's demand curve as the seller, in order to sell a larger amount, lowers price. But this seems to be incapable of explaining all such situations. Another explanation usually advanced is the economies of scale - where scale is related to <u>rate</u> of output. However, an alternative explanation suggested here is the lower cost resulting, not from higher <u>rates</u> of output per unit time, but from larger planned volume of total output quantities. An examination of the production management and engineering literature reveals much greater emphasis on batch or lot size as contrasted to the rate of output. Frequently the latter is not much of a variable in each particular firm's decision. This means that the extent to which rate of output <u>is</u> varied may be slight--not that it can't be varied or that its significance is slight. That there has been confusion between the rate of output and the batch size or quantity planned is sure. How much cannot be known.

The third feature--that of identifying each output operation with a calendar date and then postulating that the more distant the date the smaller the change in equity (the smaller the cost)--provides a way to escape the unnecessary bind imposed by the definition of short-run costs as that which results from fixed inputs. The ambiguous idea of two different costs, a short-run and a long-run cost for a given output, disappears and is replaced by one cost for each different program of output.

What must have been assumed in our present literature about the factors mentioned here? Was the rate of output profile assumed to be a constant rate extending into perpetuity? The answer could not be ascertained from an exhausting reading of the literature nor from analogically implied conditions. Certainly the standard cost curve analysis does not envisage a perpetuity output at some given rate--nor does it seem to specify the effects of shorter length runs at any output. For example, Stigler in his well-known paper on the effects of planning for variations in the rate of output imagines one to be moving along a given cost curve appropriate to the case in which output varies. This desirable attempt to modify the cost curve analysis would have been more successful if the output had been further specified or identified in terms of V and T. Then the conventional curves would have disappeared and many logical inconsistencies and ambiguities could have been removed from the standard analysis. But merely drawing the curve flatter and higher does not avoid the problems of appropriate interpretation of costs for unspecified values of the pertinent variables.

Finally, introduction of a new variable, V, complicates the equilibrium of demand and supply, for now there must be a similar element in demand which will determine the equilibrium size of V, if such there be. Suffice it to say here that even though consumers may not act or plan consciously in terms of V, their actions can be interpreted in terms of a resultant aggregative V. Producers, in contemplating the demand for their products, will be required to think of capital value or present value of income with the rate of output integrated into a V - possibly a break-even V, on the basis of which they may make production plans. A simple rate of output, price relationships, will not be sufficient. But this remains to be developed later, only if the present propositions prove valid and useful.