Towards the Concept of the Long Run Average Cost Curve

David D. Whitehead
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/economics_etds

Part of the Finance Commons

Recommended Citation
Whitehead, David D.. "Towards the Concept of the Long Run Average Cost Curve" (1968). Master of Arts (MA), Thesis, Economics, Old Dominion University, DOI: 10.25777/2t6s-zh67
https://digitalcommons.odu.edu/economics_etds/8

This Thesis is brought to you for free and open access by the Department of Economics at ODU Digital Commons. It has been accepted for inclusion in Economics Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
OLD DOMINION COLLEGE

TOWARDS THE CONCEPT OF THE LONG RUN AVERAGE COST CURVE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the degree of

MASTER OF ARTS IN ECONOMICS

BY

DAVID D. WHITEHEAD

Norfolk, Virginia

1968
TOWARDS THE CONCEPT OF THE LONG RUN AVERAGE COST CURVE

A THESIS

APPROVED FOR THE DEPARTMENT OF ECONOMICS

BY

H. M. Becker, Thesis Director

R. C. Burton, Committee Member

J. W. Nickson, Jr., Committee Member and Chairman, Department of Economics
INTRODUCTION

The shape of the long run average cost curve has historically been one of the most important concepts in economic theory. The long run average cost curve has traditionally been assumed to take the shape of a "U." This shape would allow for only one optimum scale of a firm in the long run. The "U" shape of the long run average cost curve is based on the concept of returns to scale. The firm, by increasing its size, would experience increasing returns down to the lowest point on the "U" shaped long run average cost curve. At this point long run marginal cost would equal price, thereby placing the firm in its best cost-revenue position. The firm would then be in equilibrium and would lose by increasing its scale further. This single optimum scale would then permit many other firms to produce in this industry at the same optimum scale, thereby establishing a competitive equilibrium.

The question that this paper will deal with is the shape that this long run average cost curve actually takes. The assumption of a "U" shaped long run average cost curve lends itself very well to theory. But empirical studies on
the actual long run average cost condition of firms have raised some doubt as to the shape this curve takes. The purpose of this paper is to survey both the theoretical reasons for this long run average cost curve to take a "U" shape and the empirical evidence of the shape the long run average cost curve takes under actual market conditions.

The first chapter will be concerned mainly with the concept of cost, and with the presently accepted view of economies and diseconomies of scale. The concept of cost will be put in its proper perspective, especially as it applies to the long run average cost curve. Following this, a description of the commonly accepted view of the relationship between long run average cost, long run marginal cost, short run average cost and short run marginal cost will be given.

The second chapter will deal with the theoretical reasons for increasing returns to scale, constant returns to scale, and decreasing returns to scale. Each of these three scale concepts will be analyzed with the view of proving or disproving their existence in theory.

The third chapter will survey the empirical evidence available and will contain the conclusions reached in this paper as a whole. One empirical study will be explained in detail in order to demonstrate the gap which exists between
theory and actual practice.

Following this a summary review of the major empirical findings will be given in Chapter IV. The conclusion will then bring together the theoretical and empirical findings.
CHAPTER I

The primary purpose of this paper is to analyze both theoretically and empirically the shape that the long-run average cost curve takes. The purpose of the present chapter is to provide the background necessary to accomplish this end. This chapter will first be concerned with putting the concept of cost in its proper perspective, especially as this concept applies to average cost. This will be followed by a brief account of the theoretical shape that the short and long-run average cost curves are assumed to take. This will lay the necessary groundwork for evaluating the theoretical concept of economies of scale, diseconomies of scale and constant returns to scale which will follow in Chapter II.

The neoclassical concept of cost or the real cost doctrine is based on the hedonistic concept of pain and pleasure. Alfred Marshall gives perhaps the best explanation of this concept. In his explanation Marshall uses the terms "sacrifice," "abstinence" and "exertion" which are all subjective measures of disutility. This concept is then based

---

1The explanation given by Alfred Marshall, in his Principles of Economics Test, is given in Appendix A.
determine objectively the nature of the secertttee in question.

perfection or utility, therefore it would be impossible to
the other because the concept involves an interpersonal com-
poor man. But this cannot be stated objectively one way or
cept the rich man would get less for the secertttee than the
very rich and a very poor man, it not long under this con-
>secrettee involved in earning a given amount the same for a
secertttee nature involving the real cost of doing the job in the
subjective nature of the real cost of doing the job in the
a great deal less physical exertion and drilltity. Also the
a great deal less physical exertion and drilltity, a drug in competition however
path to a managerial worker, a job that in competition however
is physical exertion and drilltity, and the very high salary
job of a ditch digger, a job that must require a great deal
very low except consumption. The subjective nature of the concept seed-
noticer the drilltity of labor for the abstraction from con-
work. The cost of land was not considered, since it involved
referred to the unintercetenees involved in that type of
the cost of labor would be measured in terms of the execution
subject to as opposed to the secertttee of present consumption,
percent consumption or the secertttee encountered in present
a farm would then be measured in terms of the abstraction from
to the use of any productive factor. The cost of capital to
on a subjective measure of pain, toll and secertttee attached,
Thus this concept of cost, based on disutility, has been rejected by many economists who believe that the study of economics should be set on a more objective, scientific base.

The opportunity cost doctrine replaced the real cost doctrine as the economist's basic theoretical explanation of cost. The opportunity cost doctrine states that the real cost of any given course of action is the benefit or amount of gain which could have been obtained by undertaking the next best alternative course of action. 2 This concept of cost is looked on as a more objective theory of cost. The cost to society of producing a given product, A, is the value of the next most desirable product, B, which could have been produced with the same resources, in the same quantities, used to produce A. This analysis can then be turned to yield the cost of labor in a given use. We must assume first that a man is given a choice between at least two jobs, A or B. If there was only one job that the man could perform, then the cost to society of using him in this job would be zero because his best alternative use is unemployment. We must further assume that job B is the best alternative to job A. Now if this man decided to take job A, his opportunity cost in taking this course of action is the value that he gives up

The cost of capital in a given use becomes the
cost of land in a given use to the best attainable
to the cost of land in a given use to the benefit that
one could extract out of the cost of any other factor of production
saving with that to be lower. During the same time of reasoning
the cost will be lower than the manager’s and for this reason the
profit per dollar of the street dumper possesses—the opportunity cost for the situation
which constitutes the same amount of effort and education that the
street dumper. The opportunity job open to a street dumper are those
costed a foregoing cost of accepting the job of dumper
attendant for the street dumper and the expenses cannot be
specific characteristics. Thus the job of manager is not an
expected, more ambitious, born into money, or in general he
created, more ambitious, born into money, or in general he
manager because of the fact that the manager is better off
attendant. A street dumper gives a lower wage than a
solution to the problem exists on the words, could have been
street dumper. The wages of the wages of a managerial worker. The
and when one attends to analyze why the wages of a street
found in the street dumpy concept of cost, which becomes apparent
value this concept of cost, one can solve the ambiguity
benefit that he could have accepted from job b. Now by
By not taking job b, then the real cost of job a is the
benefits that could have been derived from that capital in its best alternative use. Thus, it appears that this concept gives objectivity to the concept of cost—but does it really?

George T. Stigler suggests one of the more serious problems with the alternative or opportunity cost theory. He assumes that an independent mechanic, self-employed, will receive $1,200 a year in wages and a mechanic who is employed by someone else will receive a wage of $1,500. Under these conditions Stigler assumes an equilibrium because the value of $300 is placed on the privilege of working for oneself. Thus using the opportunity cost theory, the cost of a hired mechanic becomes $1,200, although the hired mechanic actually receives $1,500. How does one explain this $300 discrepancy? The only way seems to be to add in the cost of the pleasure that one gives up by working for someone else instead of working for one's self. If this is correct, then the theory of alternative cost must be reformulated in this way: "The cost of productive service $X$ in making $A$ is equal to the amount $B$ that $X$ could produce plus (or minus) the nonpecuniary returns (or cost) attached to producing $B."^3 And in Stigler's words, "The notion of objective cost must then be abandoned."^4

---


assess the first component, both concepts are of a very subjective

case, therefore, both concepts are of a very subjective

case, therefore, both concepts are of a very subjective

case, therefore, both concepts are of a very subjective

case, therefore, both concepts are of a very subjective

case, therefore, both concepts are of a very subjective
nature. Statements made by economists about the cost of one course of action, made in terms of the benefits which could have been derived through an alternative course of action, usually narrow down to the author's value judgment—even though the statement has been well researched and put in very scientific language.5

In adopting a subjective theoretical concept of cost, one must also accept the problems created by its subjectivity. The problems created are numerous but an example should serve to point out the problem area, especially as it involves the topic of this paper. Assume for the moment that one firm has a monopoly, that its fixed costs are zero, and that its unit costs increase as its output increases. Thus, average total cost is equal to average variable cost, as shown in Chart I, below.

The efficient product is limited by the average total cost curve. The economy
second party appears to make no product except a normal
due to the negative of the first cost factor has changed the shape
the introduction of the second cost factor. Also notice that
due to the introduction of the second cost factor. The average
cost and average variable cost are no longer the same, contrary
took the chart XX (see next page.) How average total
cost and average position of the firm will be affected to
a second cost to the second party. The configuration of the firm
be second product. The second product becomes
and assume further that the centroid to be changed will be
second party is exactly an extension of the first party
second one must assume that the agreement to take over the operation of the firm under the
price one must assume a second party can be found who to
do so he decides to lease the monopoly to a second party.
that the same as the firm decides to set off and in order
will be second product. The average total cost curve is limited by the average total cost curve that the
where the product by equilibrium marginal revenue with marginal
second curve. Thus the firm is a monopoly to market
the demand curve. But since the firm is a monopoly it must
the marginal cost curve is limited by the shape that the

8
now appears to be operating at less than full capacity as shown by the average total cost curve. Cost could be reduced by increasing output, whereas in Chart I the firm was operating with a constantly increasing average total cost function. The firm still produces and sells GA units of output with the same marginal revenue function, MR. The consumer still pays the same price for the given output. Although monopoly profits are not shown, they still exist. But because the term rent has been substituted for the term profit, the monopoly profit can no longer be found. Thus this superficial and subjective concept of cost can be made to include a monopoly profit as a rent or cost of production. 6 Nothing has really

changed except the definition of cost. The first party re-
received a profit only because he chose to call it a profit
instead of a rent paid to himself for productive factors which
he owned and could have rented to a second party as an alter-
native use for these factors. Thus, it can easily be seen
that in dealing with the concept of cost, and especially the
concept of average total cost, one is dealing with a very sub-
jective concept and must not put complete trust in cost con-
ditions as they appear on the surface. Today many large and
small companies alike capitalize rent and monopoly profits
into what appears to be a legitimate cost production. Capi-
talization of incomes made possible by contrived scarcity,
such as permits for the operation of taxicabs and liquor
stores, or the zoning of urban land sites for skyscrapers,
or simply the capitalization of goodwill, are all examples
of the form this takes.

The purpose of this chapter is to put the concept of
cost in its proper perspective, not to develop a new and more
objective concept of cost. Thus in recognizing the limita-
tion of this cost concept an effort can then be made to limit,
as much as possible, the subjective nature of one's analysis.
It will be impossible to remove the ambiguities mentioned
above in the alternative cost doctrine. These ambiguities
do not invalidate the theory of cost as a useful tool in
In this chapter, we examine various empirical works that have been undertaken to determine the cost of production. Some of the more important of these studies are:

- Empirical cost and circumstantial evidence of a firm's costs
- extent of empirical cost and circumstantial evidence of a firm's costs
- extent of empirical cost and circumstantial evidence of a firm's costs
- extent of empirical cost and circumstantial evidence of a firm's costs
- extent of empirical cost and circumstantial evidence of a firm's costs

In the process of determining the cost of production, we can place in the evaluative actual cost condition of the firm, but they do
The cost concept (in that the cost to society has not been
examined) is not one of cost or production (the act-
cation-making process, etc. cost concept or cost is differenti-
cated costs within the firm are separate in this de-
need in money terms, and costs are allocated to only those
with the addition of the explicit cost. Cost will be meas-
of the cost of production is the equivalent of accounting cost
affect of the firm made to secure the "true de-
production and the equivalent to the total mon-
place of the implicit value of all productive services used in
determination of cost of production is stated as "the purchase
Jeetvée nature as the other cost concept of "normal product"
and, as it is understood, would suffer from the same sub-
account would correspond to the concept of "normal product"
value as determined by other uses. e.g. therefore, the money
native use—whereas the de facto by the best after
the money account would then be measured by the best after-
because these inputs give rise to an implicit money account
and this does not actually correspond for pay for the de-
taken into account that arise from inputs which the firm owns
output. However, under this devaluation, cost must also be
of course, is a fluctuations time period used for analysing the impact of one or more productive agencies as existed. To this, the short run refers to that period of time in which the average cost curve takes in the short and long run.

The average cost curve takes in the short and long run.

beneficial to explain the cost book concept of what shape the average cost curve concept of the cost of production in mind. It will now be taken in the area of actual cost conditions with the deficit closer to the data presently used in empirical research under.

take this determination of the cost of production is also much possible shapes that the long-run average cost curve could be announced. From a much better understanding of the money cost will allow for a much better decision on how to allocate consideration to be taken of only the actual expenditures

best alternative uses. Using a determination of cost which points the social cost of inputs as measured by the pyramid.jeetive to be of any particular use in the paper—that area examine from consideration in an area which seems to be too sub-

need in the productive process. This determination of cost willi

takes to be the market value of self-owned sector inputs to the market value of self-owned sector inputs—actually had to pay for the inputs, and what the firm can—by the market price of the inputs, the price of the firm

taken into account. These productive costs are measured costs are
The shape of the short run average cost curve, and for this reason, the shape of the long run average cost curve, are variable. The long run in the long run, the producer can change the factor prices and the quantity of the factor used. The long run can be divided into a period of time in which all inputs are variable and into a period of time in which only fixed inputs are variable. In the latter case, a change in the input of the factor does not change the quantity of the variable input. In the former case, the quantity of the variable input changes as well. In the short run, the quantity of the variable input cannot be changed. The producer does not have inputs which can easily be varied within the short period. In the short period, the producer is a price taker and the marginal cost is equal to the price. In the long period, the cost function is constant. The cost function is constant, so great that the enterprise would not consider the price of the product of the variable input. If the price is below the marginal cost of the variable input, the short run cannot operate an enterprise. If the price is above the marginal cost of the variable input, the short run can operate at a profit, no matter the purpose. Most inputs can be varied at will, no matter the purpose.
In the theory of profit (1952), pp. 317-318.


In economics, the theory of profit (New York: The

texture of the state of technology could be deduced as the state
determine separately and show how they apply to the short run.

It may prove helpful at this time to take each of these con-

serve: combination of the production function in which the various produc-
ted, the law assumes the possibility of varying 
measure whose quantity is held constant.

Second, it is necessary that there be production 

First, the state of technology is given.

Why it is to hold true:

Fourth, these conditions which are necessary to this relation-

good. In the interpretation of the law, price

duction product with the elasticity of production is a constant.

The price of output varies with output, and the elasticity of output is a constant.

answer the question of what the relationship between the factors of production is.

Elasticity of the demand for labor: the law of diminishing returns has been stated as

the law of diminishing returns as a law of growth of productive capacity

When the case for the shaped short-run aggregate cost curve

and that the rate of output. From this law one can deduce the

SUCCEm.

15
of the industrial arts. The state of technology is given in the production function of the firm—as long as the firm is working under a given production function the state of technology is held constant. If there is a change in technology, then a new production function must be derived. Put in the context, changes in technology must be taken in historical sequence—at least for the firm. To illustrate this point it is necessary to show the relationship between technological change and the production function. A production function shows the relationship between physical inputs and physical outputs of the firm, and can thus be said to show the maximum attainable output which can be produced from a quantity of input. The production function shows a flow of inputs resulting in a flow of output for a specific time period. Thus the production function shows the varying amounts of output which can be produced from varying quantities of inputs. Technological change, on the other hand, will have the effect of changing the production function, thus showing a changing quantity of output for the same quantity of inputs. Note the difference in the input-output relationship. Production functions show flows of inputs which produce different levels of output; technological change is confined to the same quantity of inputs producing a greater, or in some cases a lesser, output. Thus the firm under a given production
function can produce varying levels of output by changing the quantity of inputs used, but for any of these levels of output to change without a change in the quantity of inputs would mean a change has taken place in technology used.\(^{15}\)

Thus in the short run the firm is limited to a given production function but can produce anywhere on its production function without changing technology. Its production, of course, will be limited also by the scale of plant it is operating with in the short run.\(^{16}\)

Stigler's second condition was that there be at least one productive factor whose quantity is held constant. Thus the law would not apply if all factors of production were variable.\(^{17}\) The concept of the short run is based on the assumption that at least one productive factor is fixed, just as the law of diminishing returns is based on this assumption. If all factors are capable of being varied then we are again working with the long run concept. In the short run, then, we assume that the size of plant is not capable of being changed, the one fixed factor then becomes the size of the

---

\(^{15}\) Stated in this way, anything that would increase or decrease the productivity of a given input or all inputs combined, without a change in the physical quantity of the input, would be lumped together in the term "change in technology."


plant. Thus the fixed factor is fulfilled in the short run by both the size of plant and the level of technology. In the long run both the scale of the plant and the state of technology can change, in which case the law would not hold.

Stigler's third condition was that the "law premises the possibility of varying the proportions in which the various productive services combined." This condition must be fulfilled. If it were not the entrepreneur could not expect to produce more or less output if he added to just one of his factor inputs. Assume for the moment that it takes one unit of A and one unit of B to produce one unit of output, C. No other proportion of A and B added together will produce a unit of output, C. Under this condition, it would be impossible to hold either the quantity of A or B constant and increase the quantity of the other with the expectation of increasing output. If input A is held constant at one unit, under this condition, no matter how much the other input is increased in units, the resulting output can only be one unit of C. Thus the marginal product would, of course, be zero. In order for the law to hold true the proportion in which A and B combine to produce C must have the quality of being variable. If this were not the case and the input factors combined in only a rigidly fixed proportion then both inputs must be increased

in like proportions in order to achieve a greater output. This would make it impossible to hold one factor fixed and thus impossible for the firm to meet the short run requirements. This would apply not only for two variables but for any number of variables.

Under the law of diminishing returns if a producer continues to add equal increments of a productive input, holding at least one other productive input constant, that a point will be reached at which the marginal product will decrease. This law is extremely general, even given the condition upon which it is based. The law applies to almost all production functions, in almost every line of business. Yet the law makes no estimate about how fast the marginal product will diminish, nor does it predict whether marginal product will become negative or stay positive. But the law does imply what shape the marginal and average product curve will take, and for the purposes of this paper that is all that is really necessary. A numerical example of this law, followed by a graphical description of the shape that the implied marginal product and average product curve takes will prove helpful in the transition from the average product curve to the average cost curve. The most common example, used here, is one applied to agriculture.
### Table 1

**The Law of Diminishing Returns**

<table>
<thead>
<tr>
<th>Variable Input</th>
<th>Fixed Input</th>
<th>Total Produce</th>
<th>Average Produce</th>
<th>Marginal Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor (Man Days)</td>
<td>(Land)</td>
<td>(curve)</td>
<td>(curve)</td>
<td>(curve)</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>12</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>19</td>
<td>4 3/4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>25</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>30</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>34</td>
<td>4 6/7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>37</td>
<td>4 5/8</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>39</td>
<td>4 3/9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>40</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>40 1/2</td>
<td>3 15/22</td>
<td>1/2</td>
</tr>
</tbody>
</table>


This table, if converted to a graphical presentation, would appear as the curves in Chart III.

### Chart III

![Chart III](chart.png)

**NOTE:** No attempt will be made to make this chart represent the actual scale of the table—for the purposes at hand only the shape of the approximate average and marginal product curves are important.
The average product curve, as implied in Chart III, shows that the average product increases, reaches a maximum, and thereafter diminishes. It should be noted that the marginal product curve cuts the average product curve from above at the average product curve's highest point. To make the transition from product curves to cost curves requires but one simple assumption. It must be assumed that the firm under consideration is so small that as it purchases more and more of the variable input factor, it can in no way affect the price at which its factor inputs are sold in the market. Thus the price paid for the variable input or inputs must remain constant. If this assumption holds true, then the transition is an easy one to make, because a cost curve is but a "geometrical illustration of the relationship between the rate of output of a firm and the rate of expenditure on various inputs." 19 Thus, if the product curve is simply a geometrical illustration of the relationship between the rate of quantity input and the rate of output—and the price of the input is held constant—then it is a very simple matter to derive the total price of the inputs by multiplying the total quantity of the inputs by the input prices. The relationship can be shown as follows:

\[
\text{Total Cost} = \frac{\text{Average Cost} \times \text{Output}}{\text{Quantity of input factors}}
\]

\[
\text{Marginal Cost} = \frac{\text{Change in Total Cost}}{\text{Change in Output}}
\]

\[
\text{Change in quantity of factors} = \frac{\text{price of factor}}{\text{Change in Output}}
\]

From this point the following terms will be abbreviated in this manner:

- TFC = Total Fixed Cost
- TVC = Total Variable Cost
- TC = Total Cost
- MC = Marginal Cost
- AFC = Average Fixed Cost
- AVC = Average Variable Cost
- ATC = Average Total Cost
- SRAC = Short Run Average Cost
- LRAC = Long Run Average Cost
- SRMC = Short Run Marginal Cost
- LRMC = Long Run Marginal Cost

The transition to cost curves from product curves is then a very simple process, assuming input prices are fixed. The cost curves are but a mirror image of the product curve. The law of diminishing returns tells us that after some point the addition of equal increments of one input added to a fixed input will lead to decreasing return, which suggests that the average product will increase to a point, reach a maximum, and then decrease. The marginal product curve will increase to a
point, reach a maximum higher than average product curve, and cut the average product curve at its highest point. Thus the average product will decrease to a point, reach a minimum, and then increase. The MC will, being the increase of the marginal product curve, decrease to a point, reach a minimum below average cost, and then increase, cutting the SRAC curve at its minimum point. A geometric illustration of the two sets of curves could be approximated as follows:

**CHART IV**

![Diagram A](chart-a) ![Diagram B](chart-b)
The SRAC curve is usually shown in textbooks as a "U" shaped curve because of the law of diminishing returns.

Now the question arises, what is the optimum output for the firm given the shape of the SRAC curve? The answer to this is relatively simple if we remember that a competitive firm will maximize its profits if it produces where MC equals price. Under the theory of pure competition it is assumed that the demand curve for the firm is completely elastic, because the firm's output is too small to affect market price. Also under the theory of pure competition one assumes that the entrepreneur is entitled to a "normal profit" which is included in the SRAC of that firm. Therefore the firm under pure competition will be in the situation depicted in Chart V.

**Chart V**
seems to be the factors which govern the shape of the demand curve.

The shape of the demand curve has nothing to do with the

be the lowest point on the demand curve. If
the more elastic point for the law in the short run would
in revenue would be greater than the decrease in cost. Thus
output, output at which the output is at which the average total cost would
output, the elasticity of demand at the point at which average total
total at which the output at the point at which average total
output, the demand curve must be tangent to the price line.

Also since the supply curve is perfectly elastic and the price
the supply curve is perfectly elastic, and the price

The elasticity of demand at the point at which average total
It is assumed, for reasons to be explained later, that by increasing the scale of plant, and adjusting all inputs to their optimum utilization, the per unit cost of production will decrease, reach a minimum, and then increase. The shape of the LRAC curve is formed by a series of short run average cost curves. Each SRAC curve is a different size of plant, therefore each point which lies on the long run average cost curve represents a unique size or scale of plant. Since each point on the LRAC curve is a different scale of plant, the LRAC curve is sometimes referred to as a planning curve. And since all production takes place in the short run, the entrepreneur by necessity must choose between various scales of plants on the basis of the output he desires to produce. An enlarged segment of the LRAC curve would appear as follows:

**CHART VI**
Since these SRAC curves represent a unique scale of plant then each producer must decide which scale he will use in light of the demand conditions facing his firm. If the producer believes that the demand conditions are such that output $GK_1$ is desirable, then he will choose the scale of firm that will allow him to produce $GK_1$ at the least cost per unit—in this case scale SRAC$_1$. If the producer believes that he faces a larger demand, one which will require the output of $GK_2$ units, then he will build scale SRAC$_2$, and so on. But his decision becomes more complex when he foresees the need to produce $GK_1$, because at this output level both scales SRAC$_1$ and SRAC$_2$ will be equally efficient; i.e., their cost per unit of output will be identical. In this situation the producer's decision must be made on the basis of either capital requirements or future demand conditions. The producer may choose scale SRAC$_1$, because it requires less capital investment; or he may choose the larger scale, SRAC$_2$, because he foresees demand conditions as permanently expanding. In all other situations the producer will make his choice of plant solely on the basis of cost per unit of output. In the short run the producer has no option between these scales. Thus if he is producing with scale SRAC$_1$, and requires an output of $GK_2$, he will produce this output with his present scale of plant. But in the long run, if the
If both $x$ and $y$ are negative, then $-x > x$ and $-y > y$. Now, the tangent at the minimum point on the graph curve is tangential to the curve defined by the equation $y = x^2$. In order to show the geometrically existent that the tangent curve at the minimum point on the graph curve is tangent to the equation of a given point, it is practically necessary that the graph curve and the minimum point on the graph curve would be the graph curve and the tangent point between the minimum point on the graph curve, and the tangent point on the graph curve. The geometrically existent that a little when the dx is not constant, the geometric equation that is a tangent to that point shown by the least cost per unit of product at any possible least of the graph curve as a parameter of the minimum point on the graph curve shows a decrease in demand to the forecast as a parameter, the willingness of a
to the optimum point of the LRAC curve at only one point, that being the lowest point on the LRAC curve. To the left of the lowest point on the LRAC curve all the SRAC curves’ optimum points are drawn to the left of the tangency of the SRAC curve and the LRAC curve.\textsuperscript{24} The next question to answer is what happens to marginal cost in the long run. This can also be illustrated in Chart VII. We know that each SRAC curve has a SSMC curve passing through its lowest point, as shown by SSMC\textsubscript{1}. If we draw the SSMC curve for its associated SRAC curve and

\textsuperscript{24}\textit{Ferguson, Microeconomic Theory\textsuperscript{,} p. 178.}
then drop a perpendicular line from the point of tangency of each SRAC curve on the LRAC curve, the point of intersection between the perpendiculars and the SRMC curve will be one point on the LRMC curve. If this process is repeated for each SRAC curve the locus of points that are derived will be the LRMC curve and will cut the LRAC curve from beneath and at the LRAC curve's lowest point. A simplified explanation of why the point at which the perpendicular, drawn from the tangency, cuts the SRMC curve must be a point on the LRMC curve could be stated as follows: At the point of tangency there is but one output, OQ₁. At this level of output the SRAC must be equal to the LRAC, by definition of tangency. There can be but one level of marginal cost for a given level of average cost. Thus the SRMC of output OQ₁ with the scale of plant SRAC must be the same as the LRMC at this output. Thus by dropping a perpendicular from the point of tangency, where SRAC₁ meets the LRAC curve, we are able to find the SRMC for that level of output which must also be the LRMC for that output level. This is also the reason that SRMC and LRMC are equated at the lowest point on the LRAC curve. With this information one knows why the SRAC curve is "U" shaped, why SRMC rises, and what the relationship is between SRAC, SRMC, LRAC, and LRMC. But this does not explain why the LRAC curve takes the shape of a "U." The basic reason for this, as
explained in text books, is economies and diseconomies of scale.

Economies of scale, or increasing returns to scale, are explained by Paul A. Samuelson in this way: "In many industrial processes, when you double all inputs, you may find that your output is more than doubled." Samuelson also explains that these increasing returns to scale are associated with:

1) the use of nonhuman and nonanimal power sources (water and wind power, steam, electricity, turbines and internal-combustion engines, internal atomic energy);
2) the use of automatic self-adjusting mechanisms (lathes, jigs, servomechanisms);
3) the use of standardized interchangeable parts;
4) the breakdown of complex processes into simple repetitive operations;
5) the specialization of function and division of labor; as well as many other technological factors.

Thus as it is stated, economies of scale are realized in situations where it is possible and desirable to produce very large quantities of an output. Thus as one increases his scale of plant, per unit cost will diminish. Adam Smith was perhaps the first to understand that economies of scale could be realized through specialization and division of labor. He


26 Samuelson, Economics: An Introductory Analysis, p. 27.
realized that if there were any fixed input the opportunities for the division of labor and its specialization would be exhausted very quickly. But he also realized that if there were no fixed factors (the long run concept as defined above) then specialization and division of labor could give very substantial reductions in cost per unit of output. If workers are not specialized, part of their work time is taken up by moving from place to place and changing tools. Specialization will allow each worker to become more proficient at his job and will save on time, in that he will not have to move as much or change tools as often. The reasons given for economies of scale by Samuelson, above, are fairly extensive and will serve the purpose, at this point anyway, of explaining economies of scale, or the negatively shaped portion of the LMAC curve. Other reasons for these economies of scale, from both internal and external sources, will be forthcoming in the next chapter of this paper.

The most commonly used argument for diseconomies of scale, or the positively sloped portion of the LMAC curve, is one that involves the management function of the firm. It is argued that as the scale of operation expands the delegation of responsibility and authority is inevitable. As this

---

27 Ferguson, Microeconomic Theory, p. 181.
have set the stage for the arguments that still need to be made. For this analysis, the most obvious start—of these various types of economies, The most obvious start—of the following chapter to analyze some of the theoretical causes that have also been given. It will be the primary purpose of the scale, economies of scale and consistent return to scale, the chapter an examination of scale. The reasons for this are not expected research in economics. It is noted that only with the subject of the next chapter. will it be the subject of the next chapter.

According to the set forth, we cannot expect the economies of scale to decrease after a certain scale of operation. Some of the arguments are stated by the International Point of View. The production tends to rise, but the point at which the production is affected with the effect of labor with the effect that per unit cost of operation decreases. As cost decreases by distribution by distribution and special.

As the cost of performing the managerial duties of the organization decreases, because of the red tape and paper work that accompanies this decrease, because of the red tape and paper work, the operation of the firm becomes steaded down, or
CHAPTER II

The classical economists were as much concerned over the size-efficiency question as economists are today. Traditionally the question was concerned with whether or not competition was compatible with increasing size; if increasing size and efficiency went hand-in-hand. In the traditional argument the question was put in terms of industries in which economies could either stem from the expansion of the industry as a whole, external economies, or from the expansion of an individual firm, internal economies. Internal economies were not affected by the expansion or contraction of the industry as a whole. The concept of increasing economies created a dilemma. If an individual firm experienced increasing returns, from a continual increase in its size, what force would operate to prevent this firm from becoming a monopoly? The classical economists were not concerned with the external economies since the expansion of an industry as a whole would result in the growth of each individual firm in the industry, and thus could not involve a major change in the relative size of each firm in the industry, and thus could not involve a major change in the relative size of each firm in that industry.
The internal economies were open to all members of that industry, but all the entrepreneurs of these individual firms did not choose to expand—thus allowing the firm that did choose to expand to receive internal economies at the expense of all the other firms in the industry. The expanding firm would increase its size again and again until it had finally monopolized the whole industry. The concept, it is important to remember, was put in terms of industries that showed increasing returns as the firms enlarged their output. The firm in any one of these industries would not, it was assumed, show all three of these tendencies, increasing returns to scale, constant returns to scale and diminishing returns to scale. D. H. Robertson summarized the dilemma between size and efficiency in this way: "The root difficulty about increasing returns has always been to understand how, where they prevail, equilibrium can exist without the whole supply of the commodity in question becoming concentrated in the hands of one producer." The economic question became centered only on increasing return industries, thus leaving the question of decreasing return and constant return industries as being so few in number as to merit no serious consideration.\(^2\) In the words of J. H. Claphan:

---

\(^1\) John M. Blair, "Does Large-Scale Enterprise Result in Lower Costs?"

\(^2\) Blair, "Does Large-Scale Enterprise Result in Lower Costs?" p. 123.
Constant returns, it may be observed in passing, must always remain a mathematical point, their box an empty one. It is inconceivable that a method can ever be devised for so measuring these real by indefinitely subtle and imponderable tendencies toward diminishing and increasing returns that someone will be able to say. So, here a perfect balance.³

This confusion over the cause of constant returns, if it can be called confusion, was believed to have been solved by Alfred Marshall. Marshall believed that the forces causing economies and diseconomies of scale were constantly working against one another. However he did not argue that internal economies were being offset by internal diseconomies, but rather that internal economies were being offset by external diseconomies. Thus if industry B supplied inputs for industry A, and the firms in industry B were experiencing internal diseconomies of scale whereas the firms in A were experiencing internal economies of scale, then the internal diseconomies of scale in industry B would become external diseconomies of scale to industry A. If the external diseconomies just offset the internal economies, the result would be constant return to scale in industry A.⁴ It later became evident that this did not adequately explain the problem of constant returns. But Marshall later laid the ground work which has


developed into the presently accepted view of the form that
economies of scale, diseconomies of scale and constant returns
to scale take. The accepted view of the form economies, dis-
economies and constant returns to scale take was illustrated
in Chapter One.

Marshall by no means wrote in clear precise terms.
It seems, in reading Marshall's Principles of Economics, that
he implies many things but precisely defines very few. An
illustration of this can be found in Marshall's definition of
what "the representative firm" is, or more precisely what it
is not. As he explained it, a representative firm is neither
a young growing nor an old decaying firm; it does not enjoy
unusual advantages nor is it unusually large; and the repre-
sentative firm tends to increase in size in accord with the
industry's expansion, this increase being accompanied by only
the normal economies available to every firm of representative
size within the industry. The firm could then be described as
an average firm, according to Marshall.\(^5\) But how much informa-
tion does this description really relay? The firm is somewhere
between young and old, but where? What are unusual advantages?
And how would one describe and measure normal economies?
Marshall did not answer these questions but he did explain the

nature of the firm, within the context of the industry. And this was the ground work which has developed into the explanation of economies of scale, diseconomies of scale, and constant returns to scale as they are known today.

Marshall divided the economies attributable to increasing scale into those economies which depended on the growth and development of the industry as a whole, external economies, and those economies which depended on the growth of the individual firm and its resources, internal economies. Internal economies were responsible for the constant state of fluctuation which each firm would experience in per unit cost in the long run. Marshall likened the growth of a firm in an industry to the growth of a tree in a forest.6 The firm grows because of the abstinence, desire, energy and good luck of its entrepreneur. As the firm establishes itself in the industry, its capital and lines of credit expand. This leads to further expansion which will result in internal economies of scale. These internal economies will then result in more capital accumulation and easier access to credit. The process repeats itself again and again, limited only by the entrepreneur's energies. As the firm grows older so does the entrepreneur; he loses his energy and desire. His firm will then begin to

---

6Appendix B of this paper is Marshall's description of the growth and decay of a firm within the context of an industry.
decay, like the old trees in the forest. It will not accept changing methods like it did when it was young. Thus this firm begins to die. The old trees in the forest decay and this makes for fertile ground to support the younger and still growing trees. In the same way, the death of the old firms make room for the new and still growing firms in the industry. This cycle is repeated throughout history. The length of the cycle depends on the rate at which the energies of the entrepreneur decline, and finally fail. Thus Marshall explained that the industry would be made up of many different size firms, some very small and growing, some very large and failing, and some which would never grow over a medium size. Over the years a single firm would not dominate an industry because sooner or later its entrepreneur's energy would fail and the firm would experience diseconomies of scale. Thus Marshall saw that a single firm could experience all three phases of economies over its life. The first stage of youth and maturity would conform to economies of scale. The second stage, where the firm's entrepreneurial capacity had been extended to its limit, would conform to constant returns to scale. And the third stage, that of decay, would correspond to diseconomies of scale. The fact that Marshall did understand and explain the nature of the firm, within the context of the industry, set the stage for the development of the theories of economies of scale,
diseconomies of scale and constant returns to scale.

Marshall's theory of the firm, like much of his work, was not explained in precise terms. This impression in his theory of the firm led to his theory being criticised for the following reasons. First, Marshall assumed that the "entrepreneurial faculties" would decay after a time thus causing the death of the firm. In fact what he was doing was holding the management of the firm constant—in other words holding one input fixed in the long run which by definition has no fixed factor. Through the introduction of fixed "entrepreneurial faculties" he was able to show a firm which experienced economies of scale only to a certain point, this point being governed by the talent and energies of the entrepreneur. After this point was reached diseconomies of scale would result from further increases in output. In so doing he put the analysis of the long run on the same footing as the analysis of the short run and thereby forced his point, that being that the firm after a given point will experience diseconomies of scale. This "decaying entrepreneurial faculties" has also been criticised on the ground that it is an unrealistic assumption. John Blair explained the process of entrepreneurial decay is a far slower process than Marshall had
had any idea; "if, in fact, it takes place at all." 7 Second, his assumption or prediction that small firms with great entrepreneurial energies will grow and replace the decaying larger establishment also seemed unrealistic. In the words of Staeindi "it seems to be a more realistic assumption that most small businesses die before they have time to grow." 8 Third, it has been put forth that Marshall overlooked the nature of obtaining adequate venture capital and long term credit, in that it is extremely difficult for small firms to obtain this type of capital and credit. 9 Thus the basic problem, it was held, had not really been solved. There was still not way to explain how a competitive equilibrium and increasing return could be present in the same industry.

Marshall did, however, set the stage for an explanation of the nature and causes of economies and diseconomies of scale.

The question of how a competitive equilibrium and increasing returns could be present in the same industry has never really been solved, as far as this writer knows. A. C. Pigou, in Economics of Welfare, did attempt an explanation,

---

7Blair, "Does Large-Scale Enterprise Result in Lower Cost?" p. 124.
8Blair, "Does Large-Scale Enterprise Result in Lower Cost?" p. 124.
9Blair, "Does Large-Scale Enterprise Result in Lower Cost?" p. 125.
but based it on external economies of the industry in question. This solution is only applicable to industries which enjoy large enough external economies to allow the whole industry to grow. It is felt that external economies, in reality, are not sufficiently large to warrant the statement that the problem has been solved. This question, however, will not affect the contents of this paper. Whether equilibrium can be established will be governed by the shape of the LRAC curve but the shape of the LRAC curve will not be dependent on the establishment of an equilibrium. Thus, since the question of competitive equilibrium and increasing returns cannot be justified as far as this writer knows, it will not be considered in any further detail.

An article appeared in Zeitschrift für Nationalökonomie, in the year 1931, entitled "cost curves and supply curves." The author of this article was Jacob Viner and he had managed to assimilate the ideas of Alfred Marshall and others into a workable body of knowledge. It is important at this point to follow his article in detail because it became the focal point of the arguments (still raging today) over the different shapes that the LRAC curve can take and the causes for these shapes. This article, it should be pointed out, was based not only on the work of Marshall, whom had laid the foundation for this type of analysis, but also on the work of Pigou, Sraffa, Shove,
Harrod and Robertson. The procedure which he followed was to begin in each case with "the mode of adjustment of a particular concern to the given market situation when the industry as a whole is supposed to be in stable equilibrium." He assumed the firm under study to be a typical one, at least in respect to the qualitative behavior of its cost when its own output is varied or when the output in the industry, of which the firm is a part, varies. He assumed that there are no long-run differences in efficiency as between concerns in the same industry, thus every firm experiences the identical total cost per unit at a given level of output. He assumes further that under long-run equilibrium conditions in the industry, every firm in the industry will experience the same relationship between "average cost, its marginal cost and market price" as the firm under investigation.

Viner first attacks the problem of the individual firm in the short run. He explained the short run much like it has been explained earlier in this paper, and through the use of a chart similar to the one used earlier.


CHART VIII

NOTE: The labels of these curves have been changed to comply with the labels used in this paper.


Under the assumptions of competition it is assumed that no firm is large enough to affect the supply price of its factors of production. Thus, as explained earlier, the laws of increasing and diminishing return can be applied to the cost curves with no trouble. The short run according to Viner's definition is the same as the generally accepted definition of short run given earlier, simply, a period so short that the scale of plant cannot be changed but within which, changes can be made in plant outputs. One of the key implicit assumptions that he makes, that will be of great interest later, is that "fixed cost" are fixed only in their aggregate amounts and vary with output in their amount per unit," thus allowing for
The assumption that automatic competition prevails which allows the firm a normal profit, but no more, and a normal profit is contained within the SMC of the firm. Thus the producer can produce at a higher or lower point on the SMC curve but it is the firm's normal profit. Thus at this point SMAC (SMM) and SMC with price. Thus at this point SMAC equal SMM and SMC at price. Short run average cost is equal to price by the equal price. In the short run it will be to his benefit to equate product is perfectly elastic, and equal to the horizontal price line. The marginal cost curve cuts the SMP curve from below and at the lowest point on the SMP curve. Under the assumption that automatic competition prevails we can assume that a change in the producer's output will have no effect on the price at which he sells. Thus the demand curve for his product is a rectangular hyperbola because by definition the average fixed cost curve must be a rectangular hyperbola because by definition the average fixed cost curve rises as depicted in Chart VII because the marginal cost curve cuts the SMP curve at a point with the price of inputs held constant, the average variable cost must rise with any increase in output. The SMP curve, or average total cost curve, is necessarily upward sloped. The SMP curve, or average total cost curve, is necessarily upward sloped.
to his advantage to produce where SRMC equals SRAC.\textsuperscript{12}

Viner defines the long run as "a period long enough to permit each producer to make such technologically possible changes in scale of plant as he desires, and thus to vary his output either by a more or less intensive utilization of existing plant, or by varying the scale of his plant or by some combination of these methods."\textsuperscript{13} Thus he explains that no cost will be "technologically fixed" within the definition of the long run. In order to change the output of the industry as a whole, there are three possibilities open: "change in intensity of use of existing plants, change in scale of plants and change in the number of plants."\textsuperscript{14} In order to achieve long run equilibrium, which Viner describes as only lasting for a moment and as being the unusual rather than the usual economic situation, marginal cost must equal price which in turn must equal average cost for the existing plant. Under any other circumstances abnormal profits or losses would exist which would attract more firms into the industry, thus increase output, or would cause existing capital to be withdrawn from the industry.

It is also necessary for long-run equilibrium that each producer


\textsuperscript{13}Viner, "Cost Curves and Supply Curves," p. 28.

\textsuperscript{14}Viner, "Cost Curves and Supply Curves," p. 29.
produce at the optimum for his given scale of plant and that no other producer is able to supply an equal amount of output at a lower average cost. Thus long-run competitive equilibrium would be established.

The condition of constant cost as described by Viner was wholly inconceivable in the short run for firms that have any fixed cost, simply because of the law of diminishing returns. This also applied to constant marginal cost in the short run. In the long run, however, constant costs were considered feasible owing to two kinds of conditions. In the first case, increases in the scale of plant would show no effect on the long-run average cost curve. Thus he described the situation as depicted in Chart IX.

**Chart IX**

![Chart IX](chart.png)

**NOTE:** The labels for the curves have been changed in order to conform to the labels used in this paper.

**Source:** Jacob Viner, "Cost Curves and Supply Curves," Zeitschrift für Nationalökonomie, III (September, 1931), p. 33.
produce any amount, and at a price above any the firm would be to produce at all, or price at which the firm would be unmotivated to act. Any price below any price at which the firm would be unmotivated to act horizontally the x-axis, the supply curve, between the long run supply curve and the market supply curve. Consequently, if perfect competition, it would be impossible to introduce the x-axis horizontally the intersection point that is, the point at which supply a perfectly elastic supply. Then, for the horizontal line at this rate, the price would be equal to price, the horizontal line, and the price line for the firm, the price line, the long line for the firm would be constant along the line. Consequently, the horizontal line, the horizontal line representing time, it would follow that time must have point on each of the supply curves at tangent to the best point of production of each output in the long term given that the point on the supply, which would also be the output rate of point on the supply, which would pay the producer to produce each output at the lowest economies or diseconomies of scale, since in the long run at the point at which the shape of a horizontal line—without elasticity so on. Therefore, the long can average cost function would be scale, if output increases, output can go to 0, and average cost decreases, if output increases, output can go to 0, so that there is no gain or loss in short run average cost curve. Outlining it at the lowest point, as can be seen in a graph, each supply curve is a different scale of plant, which the firm
The second case of constant cost would be to produce at long run maximum average cost, in which the least cost of production would tend to monopolize the industry. If all cost were not uniform in the industry as a whole, the lowest cost producer would tend to monopolize the industry.
assumption of perfect competition an equilibrium would be impossible to establish under either type of constant returns.\textsuperscript{16}

In handling the new internal economies of scale, Viner made the mistake of assuming that the long run average cost curve passed through the minimum point of each SNAC curve. His analysis was incorrect to that extent. However the cause of these internal economies and the meaning of this particular shape of SNAC curve are well worth stating. Internal and external economies of scale are concepts which were founded by Alfred Marshall. Viner defined:

\textit{"net internal economies of large scale production"} to mean net reduction in cost to a particular concern resulting from a long-run expansion of output when each output is produced from a plant of the optimum scale for that output. The word \textit{"net"} is introduced to make it clear that increase in output may result at the same time in economies and in diseconomies and that it is only the excess of the former over the latter to which reference is made here.\textsuperscript{17}

Viner goes on to explain that internal economies should not be confused with the \textit{"spreading of overhead,"} a short run phenomenon, but rather should be looked on as depending upon the optimum adjustment of plant scale to output levels. Internal economies of scale are independent of the size of output of

\begin{itemize}
\item \textsuperscript{16}Viner, \textit{"Cost Curves and Supply Curves,"} pp. 33-34.
\item \textsuperscript{17}Viner, \textit{"Cost Curves and Supply Curves,"} pp. 34-35.
\end{itemize}
the whole industry. This goes along with Marshall's definition in that internal was used to distinguish between those economies directly associated with the firm and those economies which accrue to the firm because of changes in the output of the whole industry. These internal economies, as explained by Viner, take the form of either technological efficiencies or pecuniary economies. Technological efficiencies would lower the average cost of the firm through changes in the firm's technical efficiency. Pecuniary economies would accrue to the firm because as it expanded its scale it would experience reductions in the prices paid for the input factors of production as a result of large scale buying, discounts and the like. Technological internal economies would be associated with improved methods of coordinating labor, materials, and/or equipment, or could be associated with improved methods of production that could be realized in large scale production. Viner realized, as did all other economists before and after him, that these net internal economies of large scale were not consistent with a stable equilibrium under competitive conditions. Under conditions of continuous internal economies of scale it will pay the producer to enlarge his scale of operation until he has reached the point at which his ERAC curve cuts the demand curve for the industry—thus giving him a pure monopoly. Then for any concern or for any industry as
a whole working under these conditions the long-run supply
curve would be indefinite. 18

Economies which accrue to the particular firm as a
result of the industry's expansion of output, and not as a
result of the single firm's expansion of output, are what
Marshall termed external economies of large production. Net
external economies of large production, as defined by Viner,
are economies which accrue to the member of an industry, the
scale of each member's plant being held constant and its out-
put at that scale being held constant, in which the output of
the industry is increased—presumably by an increase in the
number of plants—thereby decreasing the average cost per unit
of each member firm in that industry. These economies are
also derived from technological or pecuniary sources. Tech-
nological factors which could cause external economies in an
industry are such things as "better organization of labor and
raw materials market. . . .," or just improved communication of
ideas between the different producers. Sources of external
economies of the pecuniary nature, given by Viner, are price
reduction of factor inputs brought about by the quantity of
such inputs purchased by the industry as a whole, or a supply
input price decrease due to internal economies in output
increases in the industry from which factor inputs are

purchased. Thus one firm's internal economies may prove to be another firm's external pecuniary economies. A graphical presentation of what happens to the LRAC curve when external economies are assumed significance is given by Viner, but it has already been stated that these external economies are not significant enough to be of any measurable effect in many industries. And since they accrue equally to each member of the industry, they could have the effect of dropping the LRAC curve down but could not affect the shape of the LRAC curve.

Viner also gives consideration to what he terms "net external diseconomies of large production." The external diseconomies are caused by both technological and pecuniary considerations. Pecuniary diseconomies of this kind may result from the expansion of an industry's output. As the industry expands its output, its demand for primary input factors which it uses "must tend" to drive the price of these inputs up. If these diseconomies are not to accrue, then either the increase in demand for the primary input factor must be offset by a similar decrease in the demand of other industries for the given primary input factor, or because of net internal

---

or external economies accruing to that industry which is producing the primary input factor, its supply curve will be negatively shaped. Viner also points out that the pecuniary external economies could possibly be offset by technological external economies, resulting in no net external diseconomies. The external technological diseconomies experiences as industrial output increased could be illustrated by higher transportation cost due to increased road use; which, in turn, would be due to the greater number of units to be shipped thus causing traffic congestion.\textsuperscript{23} It should be pointed out, which Viner did not do, that these external technological diseconomies may be offset or more than offset by external technological economies, i.e., the construction of new and faster roads.

Viner's description of diseconomies of scale or what he termed "net internal diseconomies"\textsuperscript{24} leaves a lot to be desired. He explains that net internal diseconomies are "clearly conceivable," and he gives the example that the common feeling in the United States is that the one family farm is the most efficient scale of operation in the agriculture industry. This, Viner feels, would imply that the

\textsuperscript{23}Viner, "Cost Curves and Supply Curves," p. 41.

\textsuperscript{24}Viner, "Cost Curves and Supply Curves," p. 38.
optimum scale for agriculture, at least in the United States, would be achieved at a very small scale of operation, thus implying that net internal diseconomies are reached very early in this industry. Then Viner appears to retract his statement by making the following statement:

But when increases of output by means of the increase of scale of existing plants involve a substantial increase in unit cost, it will always be possible for the industry as a whole to avoid the net internal diseconomies of large-scale production by increasing its output through increases in the number of plants without increases in their scale.25

This is a very interesting statement, because it implies that there are no net internal diseconomies of scale. If it is possible for the industry as a whole to avoid these diseconomies of scale by simply repeating the given optimum scale of plant, it must follow that an entrepreneur operating one plant can either achieve decreased per unit cost or, at a minimum, constant per unit cost if he decides to open another plant of the same optimum size, given that the optimum scale is very small. Thus it would be possible for the firm, as opposed to the concept of the plant, to repeat this process again and again, thereby increasing its scale of operation, and achieve either economies of scale or at least constant returns to scale. This process could be continued and would only be limited, theoretically, by the entrepreneurial capabilities

of the firm's management. It seems highly probable that after some scale of firm is reached, the complications that would result from the management being spread over the enlarged number of plants would cause a loss in efficiency and thus an increase in per unit cost. But this case of net internal diseconomies of scale was not given by Viner. He concluded this section of his paper with the statement: "This case (net internal diseconomies) has no practical importance, therefore, except as it represents an economic barrier against increase in scale of plants, and it is not worthwhile to illustrate it graphically."

Thus Viner disposes of the problem of net internal diseconomies of scale without giving even one possible cause for these diseconomies, and in fact not even admitting that they exist unless the producer is tied down to owning a single plant as opposed to a multiplant operation. Viner cannot really be criticized on this point because the only reasons or causes for these internal diseconomies of scale that have been put forth are based on the assumption of either a constant state of technology, arguing that the LRAC curve is a planning curve, or that entrepreneurial abilities are limited factor input which must reach an optimum and diminish thereafter. Both of these methods seem to be contrived solutions to the problem of finding a theoretical

cause for diseconomies of scale.

In exploring the literature in this field, one cannot help but be impressed with the lack of theoretical discussion of internal and external diseconomies of scale as well as constant returns to scale. It seems that everyone can think of very good and realistic reasons for internal and external economies of scale. Viner's article is a perfect example of the literature in this field. He describes almost all the major types of economies of scale, but then fails miserably in his description of the reasons for constant returns to scale and diseconomies of scale.

The appendix to Joan Robinson's book, *Economies of Imperfect Competition*, is another good example of the over-abundance of explanations one can find on economies of scale and complete lack of reasonable explanation for the phenomenon of diseconomies and constant return to scale. Mrs. Robinson evidently believes that economies of scale are the result of the factors of production being indivisible. She states that "men" and "money capital" are finely divisible but for technical reasons must be used in a definite size in the productive process. Money capital must be used in the purchase


of machinery, which when put into productive process is indivisible. Men provide both the labor and managerial factors of production and to the extent that men are not finely divisible, they must be indivisible. Thus Mrs. Robinson makes a very good argument for economies of scale. If there is but one factor which is indivisible, such as entrepreneurship, then it follows that by increasing the other factor, such as an increase in the scale of operation, this fixed factor will be spread over more and more total output. The effect of this will be that average cost of that indivisible factor must decrease as output expands. Thus any factors which are used in a productive process that are indivisible, and not being used to their fullest extent, could be used more efficiently at a large scale of operation. Thus great internal economies of scale could be realized simply because of these indivisible factors. Mrs. Robinson goes on and gives many reasons how economies could be realized by factor indivisibility. This concept seems to be little more than applying a short run concept, that of holding at least one factor fixed below its most efficient rate of use and adding equal increments of other variable factors in order to receive increase returns, to the long run. Thus it seems to follow that if one adds variable inputs, including the scale of plant in the long run, to an indivisible input factor, for example
entrepreneurship, that is not being used at its most efficient rate, it will first give increasing returns to scale, and then will experience decreasing returns to scale after the point is reached where the indivisible factor is used at its optimum rate. Thus this concept of economies of scale does little more than change the definition of the long run, in order to use the law of diminishing returns to explain economies and diseconomies of scale.

Mrs. Robinson also notes that if all factors were finely divisible there would be constant returns to scale:

If all the factors of production were finely divisible, like sand, it would be possible to reproduce the smallest output of any commodity with all the advantages of large-scale industry.\(^{29}\)

But because factors of production are not finely divisible in the real world, according to Mrs. Robinson, it must follow that some other reason must cause constant returns to scale. Mrs. Robinson was able to show constant returns to scale but only through a highly contrived example. She assumes that all factors can be measured in "efficiency units."\(^{30}\) If the quantity of each factor is increased by the same proportion of efficiency units, then output will be increased in a like proportion and the marginal physical product of each factor,

\(^{29}\)Robinson, *The Economics of Imperfect Competition*, p. 334.

\(^{30}\)Robinson, *The Economics of Imperfect Competition*, p. 335.
also measured in efficiency units, will not change; thus, constant returns to scale. By measuring input factors in efficiency units she has in effect defined away any other possible outcome.

Frank H. Knight was probably the first to relate the argument of divisibility to the problem of economies of scale. He states:

If the amount of all elements in a combination were freely variable without limit and the product also continuously divisible, it is evident that size of combination would be precisely similar in its working to any other similar composed.\(^{31}\)

Knight goes on to state that this would have the tendency to foster the growth of monopoly. In order for competition to work, under these conditions, Knight believes that "an establishment of relatively small size in proportion to the industry as a whole is more efficient than a larger one."\(^{32}\) The reasoning behind this latter statement is nonexistent, as far as the writer is concerned. If there is perfect divisibility, and the "size of combination" is "precisely similar in its working to any other similar composed," how could one scale be more efficient than a second scale?


\(^{32}\)Knight, *Risk, Uncertainty and Profit*, p. 98.
Knight's argument that divisibility of factors would mean constant returns to scale is, according to Edward H. Chamberlin, "evident' only if the effect upon efficiency of dividing factors is ignored, . . . ." Thus, taking Chamberlin's view, Knight has assumed away economies of scale in order to show constant returns to scale. Chamberlin also argues that there is nothing in the "mathematics' of divisibility which washes out the economies." A mathematician before he could solve a problem involving perfect divisibility would have to ask the economist "what divisibility meant concretely in the problem at hand and how it would affect efficiency." Thus economies of scale are not purely a mathematical concept.

There seems to be a consensus among most of the leading economists, including Hicks, Samuelson, Chamberlin and Stigler, that perfect divisibility does not imply constant returns to scale. Chamberlin goes one step further and says that "indivisibilities play no part whatever in explaining

---


34 Chamberlin, The Theory of Monopolistic Competition, p. 238.


economies of Scale." Chamberlin argues that indivisibility implies holding at least one factor fixed while varying the other factors, thus yielding a "U" shaped plant curve showing economies in its negatively shaped portion. But the case of continuous quantitative variations implies a movement along the envelope or LRAC curve, thus implying a series of increases in the scale of plant. The negatively shaped portion of the LRAC curve is the result of the relative positions of the SRAC curves, not the shape of the SRAC curves composing it. Thus holding one input indivisible in the short run, giving the SRAC curve a "U" shape, has nothing to do with economies or diseconomies of scale or the shape of the LRAC curve. Chamberlin, seemingly, has missed the point that the factor is indivisible in both the short and long run, thus making the shape of the LRAC curve dependent on the law of diminishing returns. Other attempts have been made to theoretically explain constant returns to scale and diseconomies of scale. All the attempts to explain diseconomies of scale, as far as this writer knows, have been based on the proposition that something is held constant in the long run. This would have the effect of making both the long run average cost

37 Chamberlin, The Theory of Monopolistic Competition, p. 244.
fll 1 :fi.f1. ('li 1 }i I !1:.11 1
[144x457]ftr:~ri11.·
[585x457],"l~tl
[611x408]
[585x365]=
[75x302]• • I •
[256x302]ti
[283x302]& 0
[448x302]f f
[530x302]• .
[611x302]i
[648x302]· •
[200x251]=1
[254x251].• I!'- •• r . . .
[173x234]l
[200x234]! • ., I
[171x205]:,1!rf,..
[117x175]~':if
[255x175]l
[284x175]ff~,·
[365x175]J t;.'' l ¾I,--
[583x175]i.
[118x160]r r I r I f , r
[503x160]~
[558x160]~
[586x160]~
[613x160]~
[613x160]~
[173x146]r
[202x146]_ 
[228x146]~
[257x146]I a , • I . , I . ,
[613x114]i
[640x114]r
[668x114]~
[256x114]~
[287x114]. ; r I
[391x114]t I I r ; - I 2
[667x114]I
[75x84]Jill Jf I ' 5ri• ,,.;, i&f r
[669x84]I;.
[145x66]• 
[310x66]'i
[723x66]w

that the firm must make its decision in the short run. Thus even though history has been marked with rapidly changing technology, it will not be taken into account because there is no way to measure or forecast what technological changes will be forthcoming in the long run. The LEAC curve then appears to be a "U" shaped curve in the short run, but could take any shape in the long run. This would help in the understanding of why the LEAC curve is "U" shaped. But it would be of little use in trying to describe the actual optimum size of plant in the long run. Thus this concept of a planning curve offers little help in the understanding of why the LEAC curve turns up.

The second possibility is that in the long run managerial talents are the constant factor; in other words, there are limitations on the efficiency of management. As the firm expands it is necessary for top management to delegate responsibility and authority. The close contact the management of a small firm has with its operation is lost to a large firm's management. This tends to create complications, expanded paper work, time loss between the time the decision is made and when it is carried out, and in general a loss of managerial efficiency. This would at some point in the firm's growth tend to increase the cost of the managerial function and thus the per unit cost of production. If in fact it is true that
management will, at some scale of operation, reach a point at which it will become inefficient, then this would place a limit as to how large a firm could grow. After a certain scale of operation the firm would experience diseconomies of scale. This seems to be a plausible explanation as to why the LRAC curve should turn up. But there are still problems with this concept. A larger firm is able to pay higher salaries for its managerial talent than a smaller firm. Thus the larger firm could afford to pay to get the best management possible and thus expand its scale of operation beyond the smaller firm's optimum scale in the long run. Under Marshall's theory the firm could achieve the advantage in size, to begin with, by either good luck or an exceptionally energetic entrepreneur. After the firm once achieves this advantage in size it could afford the most modern machinery. It seems likely today that such a firm could extend its managerial efficiency, almost without limit, through the use of modern computers and communication techniques. Thus even if the statement, that entrepreneurship will sooner or later reach a point at which its efficiency will decline, is reasonable it still does not give a unique optimum scale of operation in the long run. And each firm may experience a different optimum. Therefore it is conceivable that a given firm may reach its optimum scale of plant, being limited by
management, at a level of output which is either as large or larger than the total industrial demand for that product, thus allowing the firm to become a monopoly. It can easily be seen that if entrepreneurship is considered to be the limiting factor on the size that a firm can achieve, the actual size that the firm can attain in the long run is still unpredictable. Therefore the "U" shaped LRAC curve would be of no help in supporting the theory of pure competition, nor would it be of help in predicting the behavior of an actual firm.

A summation of this chapter could be put as follows: Viner assimilated the ideas of Alfred Marshall and others into a congruent body of knowledge, which set the stage for discussions over the theoretical shape of the LRAC curve that still continues today. After a survey of the literature, it can be concluded that the theoretical causes of economies of scale lie mainly in the areas of division and specialization of labor, the theoretical causes for constant returns to scale are largely non-existent, and the theoretical causes for diseconomies of scale lie mainly in the area of limitations of entrepreneurship. It could be stated that in theory economies of scale most certainly exist, constant return to scale are probably non-existent and diseconomies of scale to be realized in theory must break the condition that in the long run all inputs are variable. The following chapter will take up the
question of whether diseconomies of scale and/or constant returns to scale are experienced by firms in the actual market place. Therefore, the following chapter will survey the empirical research that has been undertaken for the purpose of discovering the actual shape of a firm's LRAC curve.
experience are not feasible in the manufacturing sector (at
are caused by a wide range of influences. Thus, controlled
account periods, partitioning in cost between accounting periods
show the components as well as the total cost for successive
which cost data will be in the form of accounting records where
necessary to obtain the cost data from a firm.

In analyzing the relationship between efficiency and

by reference to a particular study--the domestic kerosene--
spectra. Long-run average cost, which will be accounted
conditions in light of the theoretical concept of cost or,
are accounted for in attempting to analyze actual long-run cost
these studies have proceeded by some of the major problems that
of scale. First, however, it is necessary that the reader of
existence of other constant returns to scale of diseconomies
investigation be to prove or disprove, it is possible, the actual
real firm operators in the market place. The purpose of the
area of market the actual cost condition under which a
various empirical studies that have been undertaken in the
purpose of the present chapter is to test the existence of
scale or diseconomies to scale or parities both. Thus the
possible for firms to deploy either constant rent to

92)
the present time) it will be necessary for the investigation to adjust this accounting data by:

1) elimination of the effects of those factors, other than variations in output, which impinge on cost, and
2) measurement of this residual relationship between cost and the rate of utilisation.²

These two steps are necessary to the ultimate purpose of the investigation, to relate size and efficiency in light of the cost concept given earlier.

Now, the first problem encountered in the rayon industry was the fact that this industry contains one company operating seven plants, two companies operating four plants, one company operating two plants, and one company that operates one plant and owns the controlling interest in two other plants, not operating domestically. Thus instead of having one "representative firm" in the industry, the rayon industry is found to have almost no two firms alike. This industry is, then, made up of three entities: "The plant, the firm and the interest group."³ Each of these entities, then, may be different from each other in its size and efficiency. Thus what is true for one company, as far as economies of scale are

³Markham, Competition in the Rayon Industry, p. 43.
concerned, need not be true for any other company. The economies which may exist for multiplant firms were omitted from consideration on the ground that they "are both conceptually and empirically more difficult if not impossible to treat." Thus the study was confined to the cost conditions of a single plant—or a plant which was controlled and operated by large multiplant firm thereby making it impossible to infer anything about the economies of scale in the industry as a whole.

Other difficulties, not discussed in the rayon study, which arise in attempting to measure economies of scale within the context of an industry are worth note. The first difficulty arises out of the definition of what constitutes an industry. Theoretically an industry is defined as a group of products which are considered to be class substitutes by buyers in a common market. Therefore, in theory, industries are distinguishable from one another in that within each industry the degree of substitutability of products is very high, but between each industry the degree of product substitutability is very low. Under the theory of pure competition all products within an industry must be perfect

---

4 Markham, *Competition in the Rayon Industry*, p. 43.

5 Markham, *Competition in the Rayon Industry*, p. 43.

substitutes; in fact, they must be identical products. In the market place the limits of a particular industry are not so well defined. Commodities which provide the same service are rarely identical. Just about all producers attempt to differentiate their products, either by brand names, trademarks or some other device. The result of this is that it becomes extremely difficult on an objective basis to distinguish among products which belong in a given industry rather than in some other industry. Also related to this problem is the problem created by the firm producing more than one product or a multiproduct firm. Theoretically and in actuality the firm could then be a member of two or more industries. The problem, as applied to scale economies, is how to allocate the cost of the factors of production which are used in common in the production of both per unit of output of each product basis, on a proportionate scale base on price per unit of output, or should we just divide these costs down the middle and allocate half to one product and half to the other. The industry is also, theoretically, defined in terms of supplying a common market. If every firm in an industry is producing for a national market then the theoretical definition is fulfilled. But if perfectly substitutable products are produced by firms which are all located in a different part of the country and each of these firms
sold their output locally, then do we still have an industry? In this case each firm in the industry would be the industry, monopoly, in its particular region. It is, therefore, very important to analyze each firm in an industry and its relation to the industry before one can take the results of a cost study of a particular firm and make broad generalizations about the cost conditions of an entire industry.

The second problem encountered in the rayon study was one that involved a change in technology. It seems that technology changes very rapidly in the rayon industry and plants that are built several years apart display such marked physical differences that precise cost comparisons are impossible. These cost comparison difficulties were compounded by the fact that identical plants were not constructed nor equipped at the same cost. Also, the plants in the rayon industry tend to be built with a great deal of upward flexibility—or built-in indivisibility—which tends to make higher cost than necessary for their identical output level. Taking these things into consideration, it was put forth that perhaps the small sample, of only one large plant, was "more homogeneous than data which might be collected from the entire industry."  

The third major problem is this study concerned "normal profits" which must be considered as part of the cost of

---

7Markham, *Competition in the Rayon Industry*, p. 44.
production. This created a rather large problem in this particular study because the industry was financed from the ground up through equity stocks and retained earnings. The damage was compounded because the rather large investment out of the firm's self-owned resources was incapable of being measured in a quantitative sense as either a cost of production or as an economy of scale accruing to the larger firm. It is usually possible to tell whether a larger firm has an advantage, in the capital market, over a smaller firm by comparing the interest rates they were charged for short-term loans or bonds. But in this study no such figures were available because of the extensive use by the firm of its own monetary resources.

The fourth major problem, in the rayon industry study, was created by the fact that raw materials were purchased by some firms and produced by others. This problem could have been of a greater magnitude than it was because it would have been very difficult to measure the cost of raw material inputs for the firms which produced their own natural resources. Also since the firms in this industry are widely located geographically it follows that for some firms the raw materials would have been cheaper than for other firms, due to the cost

---

8Markham, *Competition in the Rayon Industry*, p. 44.
of transportation, because of their location and not because of the firm's size. But the problem was solved, in this case, because the firms which produced their own raw materials stated that these intercompany transfers were priced at the prevailing market value at the time of transfer. It seemed to be the opinion of all producers in this industry that resource price differences were due entirely to location and not to scale advantages.  

These problems are by no means confined solely to the rayon industry, neither are they to be taken as a comprehensive general outline of all the problems which must be reckoned with before one can move from the theoretical firm to the firm in the actual market place. The theory of returns to scale is meant to be a general approximation and thus does not have to be right one hundred percent of the time. It is a well understood fact that some industries are natural monopolies, making it extremely inefficient from society's point of view to have a hundred firms in a given locality offer competing natural monopoly services, i.e., phone services. Thus the theory of returns to scale is meant to be applied in a general way. But the question of its usefulness will depend on whether or not firms in the real world, as

---

opposed to the theoretical models, actually display a tendency to become more efficient as scale increases, reach a maximum rate of efficiency at some unknown scale of operation, and then experience decreasing efficiency at a scale of operation beyond the maximum. The transition from what should happen in theory and what does happen in reality must be bridged in this case by empirical evidence. The problems created by having to use accounting data and not being able to work under perfect experimental conditions will invalidate the empirical research in this field to a certain degree. Therefore logic must be used in interpreting the empirical results of any economic study.

The empirical results of the rayon industry study disclosed that as the size of the plant expanded, unit cost continued to fall. Unit cost decreased as plant size increased up to the maximum size plant operated by the firm. The multi-plant firm that this study was conducted by disclosed that it had never attempted to build a larger plant than the largest one in use at the time the study was conducted. Therefore it could furnish no further information on the possible economies or diseconomies that would be realized with a larger scale of plant. The long run average cost curve that was derived is reproduced in Chart X.
CHART X

VARIATION IN UNIT COST (A) WITH SIZE OF PLANT (B)
WITH PERCENT OF INSTALLED CAPACITY OPERATED

Frequency Distribution of Rayon Plants by Size, January 1949

<table>
<thead>
<tr>
<th>Annual Capacity (Million of Pounds)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10 and under</td>
<td>7</td>
</tr>
<tr>
<td>11-20</td>
<td>6</td>
</tr>
<tr>
<td>21-30</td>
<td>6</td>
</tr>
<tr>
<td>31-40</td>
<td>4</td>
</tr>
<tr>
<td>41 and over</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>31</strong></td>
</tr>
</tbody>
</table>

It should be noted that the variation in unit cost flattens out between 90 and 100 percent of capacity for the largest plant. This indicates that average unit cost is decreasing but at a decreasing rate as it reaches full present capacity.\(^\text{10}\) Thus it cannot be predicted what shape the curve will take with a larger plant scale. The flatness that is characteristic of this LRAC curve would seem to suggest that with a continued increase in scale the plant may experience constant returns to scale.

This study attributed the shape of the LRAC curve to several factors. First, overhead cost for a single plant averaged one-third of the total cost of production. Some of these overhead costs vary but not proportionately with the size of plant. Some overhead cost, also, remains constant through a wide range of plant scales. Indivisibility of certain plant installations was offered as the prime reason for the lower fixed cost per unit of output as the capacity of the plant is increased. It was also found that some of the economies were not derived solely from the size of plant; the major part of these economies came from spreading the cost of research and development by either increasing the scale of plant or by building a second medium size plant. The curve

\(^{10}\text{Markham, } \textit{Competition in the Rayon Industry}, \text{ p. 50.}\)
is relevant to the industry for analytical purposes as is indicated in the frequency distribution. It is also interesting to note that the seven large producers of rayon between the years 1933 and 1941 received a rate of return estimated at 12.7 percent; this is compared to the less than 5.7 percent average annual return for the four small producers.\footnote{7. Johnston, \textit{Statistical Cost Analysis} (New York: McGraw Hill Book Company, Inc., 1960), p. 44.}

The conclusion to be reached is that the rayon plant under study did experience economies of scale up to the largest scale that it reached. However, the optimum scale has apparently never been reached; therefore it is impossible to tell at what output the optimum scale of plant would produce. This study only supports the theoretical evidence given in Chapter II. This plant enjoys increasing returns to scale and has experienced neither constant returns to scale nor dis-economies of scale, but neither has it reached the optimum scale of plant.

This empirical study of the rayon industry has been presented in order to show the major difficulties that arise when one undertakes a study of this nature. Before moving on to a summary of the available empirical evidence on the actual shapes displayed by the long run average cost curves of different firms in the actual market place, it will prove helpful
at this point to make a couple of observations concerning the
gap that exists between the data used in the rayon industry
study and the theoretical concept of the long run. First of
all, the empirical study of the rayon industry was based on
the data supplied by a single firm. This data was taken in
historical sequence following the growth in the scale of plants
operated by this multiplant firm. The data used in this study
was adjusted to conform to a given state of technology, thus
the state of the arts were held constant at the 1948 level.
Consequently this study is conceptually wrong from the stand-
point of theory, because the concept of the long run has been
misused by holding constant the state of technology, and
statistically wrong in the inferences, concerning the industry
as a whole, drawn from data which was derived from a single
firm in the industry.

If one assumes that state of technology will remain
constant in the long run, then what is developed from the
given data is not a long run average cost curve but a planning
curve for that individual firm. The results of this study can
only be interpreted as applying to that firm and only to that
firm in the year 1948. To assume away the possibility of
technological change, in the long run, is to assume away the
long-run period as defined. Thus the growth in plant scale
is not put in terms of a historical sequence but rather in
terms of short run alternatives. The economies realized over time from both the growth of the firm and the change in technology are, I suspect, much greater than the economies to be realized merely by increasing the scale of the plant at a given level of technology. It is extremely doubtful that one could increase the scale of plant without increasing technology, given the definition of technology used in this paper.

Secondly, it does not seem reasonable to expect that simply because two firms manage to achieve a certain scale of operation that their average cost will be identical. This may appear to be a truism; a single scale of plant can have but one technological coefficient, by the definition of technology given earlier, and thus the short run average cost for two plants of the same scale must be identical. It would be a serious mistake to consider this proposition true under actual conditions in the market place, but it would be theoretically true using the concept of Marshall's representative firm. Under actual market conditions technological knowledge is not the same for each firm in the industry, the supply of input factors differs as between firms, and each firm could derive special benefits from the mere fact of its geographical location.\textsuperscript{12} The shape of the long run average cost curve is

important not because it relates the efficiency of existing large firms over the efficiency of existing small firms, which in fact it does not do, but because it relates the efficiency which an individual firm can achieve through expansion. The efficiency which an individual firm can achieve will bear no fixed relation to the level of efficiency which other large firms in the industry have achieved. But the efficiencies achieved by other larger firms may give a good general idea of the efficiencies which could be derived by an individual firm's expansion. The basic idea here is that the rate at which efficiency will grow with size has a tendency to differ somewhat as between different firms of the same industry.\textsuperscript{13} P. T. B. Wiles states: "There is in fact no such thing as the cost curve for an industry, only a cost curve for each firm."\textsuperscript{14}

The inferences concerning the cost conditions within the rayon industry as a whole are on a strained foundation in this study, because only the data of one firm was analyzed. If inferences must be drawn that will concern the cost conditions existing in the industry, it would be reasonable to use the representative firm concept—but how is one to know it is a representative firm unless more than one firm within

\textsuperscript{13}Wiles, \textit{Price, Cost and Output}, p. 211.

\textsuperscript{14}Wiles, \textit{Price, Cost and Output}, p. 211.
the given industry is analyzed. Even Marshall said that the representative firm was taken to be an average firm within a given industry. The question must then be asked, how do you find the average firm if your sample consists of only one firm? The point made here is that one must analyze each firm in the industry in order to predict the rate of growth the firm will experience in efficiency as the scale of its operation is expanded. To use a phrase of Joe S. Bain, "Not all firms can be painted with the same brush."\(^{15}\)

Knowing the limits to which a study of a particular firm can be applied to the industry as a whole will allow for a greater understanding of the empirical evidence concerning the shape that the firm's LRAC curve takes and to what extent generalizations (concerning the cost conditions of other firms in the industry) can be made. The empirical work undertaken in this area of long run average cost conditions of the firm is obviously too lengthy, and this paper too short, for a full and detailed account of all the pertinent information concerning each study. All that will be attempted here is to give in summary form a few of the major studies and the conclusions they reach on the existence of economies of scale, diseconomies of scale, and/or constant returns to scale. No attempt will be made to classify and categorize these studies because every

---

\(^{15}\) Bain, *Industrial Organization*, p. 159.
study undertaken differs in the number of firms analyzed, the method of analysis used, and purposes for which the study was undertaken. The first study that will be taken up here is one conducted by Joe S. Bain.\textsuperscript{16}

In 1956, Joe S. Bain's book, \textit{Barriers to New Competition},\textsuperscript{17} containing a study of twenty manufacturing industries was published. Bains adopted a questionnaire plus interview technique for the purpose of obtaining an estimate of

the minimum physical production capacity of plant required for lowest unit costs, and the percentage by which total unit costs would be higher at various smaller plant capacities.\textsuperscript{18}

Bain was able to estimate the shape taken by the plant scale curve for seven of the twenty industries studies. Table 2 contains the results of these seven estimates of the LRAC curves. (See next page.) This chart indicates that Bain considered the size of one plant of minimum optimum scale, expressed in terms of a percentage of the total plant capacity necessary to supply the common market, as a partial indicator of the importance of plant economies of scale.

\textsuperscript{16}Bain, \textit{Industrial Organization}, p. 347.

\textsuperscript{17}Joe S. Bain, \textit{Barriers to New Competition} (Cambridge: Harvard University Press, 1956), pp. 182-204.

TABLE 2

RELATIONSHIP OF RELATIVE UNIT COST TO PLANT SCALE
IN SEVEN MANUFACTURING INDUSTRIES

Relative unit cost of production at plant scales corresponding to specified market percentages of either the unsegmented national market or the largest recognized submarket (optimum cost = 100)

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3.5</th>
<th>3</th>
<th>2</th>
<th>1.5</th>
<th>1</th>
<th>0.5</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cement</td>
<td>100</td>
<td>105</td>
<td>115</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rayon</td>
<td>100</td>
<td></td>
<td>108</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soap</td>
<td>100</td>
<td></td>
<td>103</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petroleum refinery</td>
<td>100</td>
<td>105</td>
<td>101</td>
<td>102</td>
<td>103</td>
<td>105</td>
<td>103</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Cigarettes</td>
<td>100</td>
<td></td>
<td>101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tires and tubes</td>
<td>100</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liquor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>100.5</td>
<td></td>
</tr>
</tbody>
</table>


within a given industry. A plant of minimum optimum scale would be the smallest possible scale that would allow LRAC to equal LMC, or would be the first optimum scale of plant that would be reached when LRAC flattens out and becomes constant. For an example of this measure, the cigarette industry will be used. The firms in the cigarette industry face a national
In the industry, the concentration of resources on a single plant would allow the plant to achieve economies of scale more efficiently than if the plant were spread among several plants.

However, as the number of plants increases, the cost of transportation and distribution between plants increases, leading to higher production costs. This is because the economies of scale in the industry come from the concentration of resources, not from the distribution of resources.

The optimal number of plants in the industry depends on the size of the market and the cost of transportation between plants. If the market is large enough, the economies of scale in production can be achieved by a single plant. However, if the market is small, the transportation costs between plants become significant, and it may be more efficient to distribute resources among several plants.

This is a generalization of the concept of economies of scale in production, which states that larger plants can produce goods at a lower cost per unit than smaller plants. However, the optimal number of plants depends on the specific conditions of the industry, including the size of the market, the cost of transportation, and the concentration of resources.
supply less than 5% of the national market and in three of these nine industries, the minimum optimum scale would supply only 2% of the market demand. This data seems to suggest that the degree of increasing returns to scale are due to many more variables than simply the size of plant. It is assumed in theory that "at some point" increasing returns to scale will halt and decreasing returns to scale will begin. "At some point" seems to be a correct generalization and that point could be dependent not only on absolute size, but also on such factors as the demand conditions, type and quality of product, type and quality of inputs, and the divisibility of the method of producing. All these factors and many other variables will probably determine the minimum optimum scale of plant as between industries and plants in the same industry.

Bain also found that six of the industries in his study showed increasing returns to scale, beyond a single optimum scale of plant, for the firm to be "either negligible or totally absent." These firms, then, experience no multi-plant economies of scale. However, in another six industries multiplant economies of scale were evident. In the eight other industries, five of which contain firms each large

---


20 Johnston, Statistical Cost Analysis, p. 156.
enough to replace seven to twenty-five plants of optimum scale, no estimate could be obtained for possible multiplant economies.  

Bain is considered by many to be an expert on industrial organization. In his book entitled *Industrial Organization* Bain points out three different theoretical long run average cost situations. The first situation is the so-called "U" shape cost curve which denotes a situation in which economies of scale would increase, reach an optimum, and then decrease. This would mean only one point would be the most efficient scale of firm in the long run. Of this type of situation Bain has this to say:

Evidence suggests that the possibility illustrated in Figure 1 (the "U" shaped curve) is not commonly found in fact, and is perhaps realized very infrequently, if ever.  

The second situation is one in which the long run average cost curve is saucer shaped. This would denote a situation in which diseconomies of scale are reached but only at a significantly larger scale than the minimum optimal scale. This type of LRAC curve would display increasing returns to scale as the firm grows from medium to very large, and diminishing returns to scale thereafter. Thus the shape

---

of the LRAC curve in this situation would appear saucer shaped and the firm could increase or decrease its output along the horizontal line while holding marginal cost constant.

The third situation that Bain gives is one in which the firm's long run average cost curve takes on the shape of an "L." Therefore the firm after reaching its minimum optimum scale, could expand indefinitely experiencing only constant returns to scale. Consequently, any scale of firm between the minimum optimum scale and the scale large enough for one firm to supply the entire industry's demand (monopoly) would display the same cost per unit of output. The following two statements by Bain give more or less a summary review of what the empirical studies in the area of long run average cost have found:

"... We will seldom if ever find U-shaped scale curves for firms with a single unique optimal size. That is to say that the diseconomies of very large scale are typically encountered, if at all, only at scale substantially greater than minimum optimum scale of firm."

Bain then goes on to say that:

Most or all real cases of probably fit the pattern of either figure 2 (saucer shaped LRAC curves) or figure 3 ("L" shaped LRAC curves).

---


24 Bain, Industrial Organization, p. 159.
Even though Bain's statements use terms that tend to hedge, on the shapes that these cost curves normally take, the statements are based on much empirical work. Other economists, after exploring the empirical research undertaken on this question make statements to the effect that increasing average cost with size are almost never observed in reality, in fact even slight increases in average cost are very rare. Mr. J. Johnston admits that he is impressed by "the preponderance of the L-shaped pattern" of the long run average cost curve that arises "so frequently from various long-run analysis."26

Perhaps the best known work in the field of long run average cost conditions was put forth by P. W. D. Wiles in his book, Price, Cost and Output. Chapter 12 of Wiles' book contains 44 sets of data on the long run cost and output relationship which have been found to exist in a wide variety of industries. These 44 sets of data that Wiles puts forth in his book make up a total of 136 examples; of which 262 or 61.1% are "L" shaped correlations showing no tendency toward rising marginal costs; 107 or 31.9% are "U" shaped correlations, but only 12 of the 107 examples are cases in which increasing "costs in the largest size-class or classes is

25 Wiles, Price, Cost and Output, p. 213.

outside a reasonable margin of statistical error. . . . 27
and 27 examples or 8.6% showing "other correlations." 28
This is quite an impressive array of empirical findings.
Wiles makes a statement to the effect that in 50% of the
cases, that displayed "U" shapes, the major reason for the
upward shape of the average cost curve was simply caused by
the statistician's prejudice, who performed the study, and
would not have showed this upturn if an "unprejudiced"
statistician had performed the study. If this statement can
be considered true, then only about 16% of the empirical
findings display the characteristic of diminishing returns
to scale.

J. Johnston put forth a collection of studies in his
book, Statistical Cost Analysis, in which only 13% of the
studies showed definite characteristics of the "U" shaped
LRAC curve. 29 The fact seems to be implicit that the "Z"
shaped average cost curve is predominate the one which
actually firms are working under.

Another very interesting aspect of the empirical work
undertaken in the field of long run average cost analysis is
the behavior of the average administration or managerial cost

27Wiles, Price, Cost and Output, pp. 261-262.
28Wiles, Price, Cost and Output, p. 261.
29Johnston, Statistical Cost Analysis, pp. 149-168.
as the scale of operation changes. One of the most logical and most common theoretical reasons put forth for diseconomies of scale is that managerial efficiency is at some point outstripped by the enlarged scale of the operation. The empirical findings of P. T. D. Wiles, specifically in Section C of the appendix to Chapter 12 of *Price, Cost and Output*, seem to refute the increased managerial inefficiency with scale argument. Table 3 shows the effect on administrative cost as a percent of sales and as a percent of production cost as scale is increased.

**TABLE 3**

<table>
<thead>
<tr>
<th>Value of Assets (in Million)</th>
<th>Number of Firms in size class</th>
<th>Administrative Cost or Percent of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 1</td>
<td>78</td>
<td>Production Cost</td>
</tr>
<tr>
<td>1 - 3</td>
<td>201</td>
<td>24</td>
</tr>
<tr>
<td>3 - 5</td>
<td>119</td>
<td>19.5</td>
</tr>
<tr>
<td>5 - 10</td>
<td>199</td>
<td>22.5</td>
</tr>
<tr>
<td>10 - 20</td>
<td>114</td>
<td>17.5</td>
</tr>
<tr>
<td>20 - 50</td>
<td>116</td>
<td>18</td>
</tr>
<tr>
<td>50 - 100</td>
<td>45</td>
<td>17.5</td>
</tr>
<tr>
<td>100 - 200</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>200 - 500</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>500 - plus</td>
<td>10</td>
<td>16.5</td>
</tr>
</tbody>
</table>

"These correlations, too, are L-shaped. And since we know that, in general, total average cost decreases with size, a fortiori the table shows administrative cost to do so."

Wiles believes that the only diseconomy involved in large-scale management is the fact that the absolute wage bill of management tends to grow with the increase in the size of the firm. However, as illustrated in Table 3, there seems to be no diseconomies of scale involved arising from the increased scale of operations if the management bill is taken in terms of a percent of the cost of production or as a percent of net sales. Therefore the absolute cost would have no effect on the LRAC curve if the average cost of management fell as the scale of operations are expanded. In the same way it could be argued that even if the per unit cost of management rises with increased scale of operation this is very seldom the cause of a rise in total unit cost. It appears useless to discuss in abstract whether diseconomies of scale outweigh economies of scale. The net result is all that matters, and the net result seems to be that diseconomies are very rare.\textsuperscript{30} Wiles does not argue that there are no diseconomies of scale. He cites the example that "two authors seldom write a good book." Thus, Wiles believes that diseconomies of scale are the result of the human factor being particularly important in the productive process. Wiles mentions that the reasons that these diseconomies would show up have nothing to do with profits, but rather "the love of bigness for its own sake."

\textsuperscript{30}Wiles, \textit{Price, Cost and Output}, p. 217.
"the irrational tendency to plough back profit" without adequate rewards and the "desire for power." In this way Wiles came to what he considered the inescapable conclusion that:

The doctrine of the optimum size of the firm must be abolished; it is quite wrong. We can only speak of the minimum tolerable size of the firm, and therefore of a long, flat stretch of low costs, with N. C. (T. A.) roughly equal to A. C.\textsuperscript{32}

The evidence of an empirical nature found in the literature on economies of scale "suggest" the possibility that in this modern age of continued technological progress net diseconomies of scale are a rarity. The evidence also implies that increasing returns to scale are followed by a long stretch of constant returns to scale. No factor, except the irrationality of individuals, has been found that would cause net diseconomies of scale. It has also been found that each plant or each firm in an industry must be considered separately in order to make specific and meaningful statements concerning the long run cost conditions which it faces. Broad general statements about the economies of scale for an industry are incapable of predicting the cost condition under which a given firm or plant in that industry must work. The individual firm is able to derive economies not only through its

\textsuperscript{31}Wiles, \textit{Price, Cost and Output}, p. 223.

\textsuperscript{32}Wiles, \textit{Price, Cost and Output}, p. 223.
growth in size but also through such factors as location, quantity and quality of input factor, and other variables which apply solely to the individual firm. Thus just as no two industries will experience the same types of cost conditions, no two individual firms within a given industry will experience identical cost conditions. Thus it is concluded that no broad general statement of cost conditions will apply to each and every industry, nor will it apply to every firm within a given industry.
CHAPTER IV

SUMMARY

The available theoretical and empirical evidence pertaining to the shape of the long run average cost curves may be summarised as follows:

The concept of constant returns to scale seems to be impossible to support theoretically without accepting the proposition that it is an accidental balance between the opposing forces of increasing returns to scale and decreasing returns to scale. But no theoretical or empirical cause and effect relationship has been found for the phenomenon. It does exist under actual market conditions but more theoretical and empirical work will be necessary before a cause and effect relationship can be determined.

The "U" shaped long run average cost curve cannot be supported by theoretical reasoning. In fact, no reason (in line with the concept of the long run) appears able to justify diseconomies of scale. Empirical studies have proven that net diseconomies of scale are very rarely observed in the market place. Therefore the "U" shaped long run average cost curve can be supported neither by theory nor actual empirical
evidence. The "U" shape LRAC curve is therefore considered to be the exception rather than the rule. The only usefulness of this "U" shaped curve will lie in the area of teaching principles of economics, or it may prove useful in analyzing actual cost conditions of firms in an underdeveloped area of the world. The shape of the long run average cost curve was based on the concept of Marshall's representative firm. The theory of returns to scale, then, was meant to be general in nature, but proves to be limited to exceptional conditions. Thus the usefulness of this theory in its present form has little or no use in analyzing actual long run cost conditions as they apply to the firm in the market place.
APPENDIX A

We may revert to the analogy between the supply price and the demand price of a commodity. Assuming for the moment that the efficiency of production depends solely upon the exertion of the workers, we saw that "the price required to call forth the exertion necessary for producing any given amount of a commodity may be called the supply price for that amount, with reference of course to a given unit of time." But now we have to take account of the fact that production of a commodity generally requires many different kinds of labour and the use of capital in many forms. The exertion of all different kinds of labour that are directly or indirectly involved in making it; together with the abstinences or rather the waitings required for saving the capital used in making it; all these efforts and sacrifices together will be called the real cost of production of the commodity. The sums of money that have to be paid for these efforts and sacrifices will be called either its money cost of production, or, for shortness, its expense of production; they are the prices which have to be paid in order to call forth an adequate supply of the efforts and waitings that are required for making it; or, in other words, they are its supply price.1

We saw how three letter letter economies are shaped

APPENDIX B

99
The large scale of their production would put great economies within their reach; and provided they competed to their utmost with one another, the public would derive the chief benefit of these economies, and the price of the commodity would fall very low.

But there we may read a lesson from the young trees of the forest as they struggle upwards through the benumbing shade of their older rivals. Many succumb on the way, and a few only survive; those few become stronger every year, they get a larger share of light and air with every increase of their height, and at last in their turn they tower above their neighbor, and seem as though they would grow on forever and ever becoming stronger as they grow. But they do not. One tree will last longer in full vigour and attain a greater size than another; but sooner or later age tells on them all. Though the taller ones have a better access to light and air than their rivals, they gradually lose vitality; and one after another they give place to others, which, though of less material strength, have on their side the vigour of youth.

And as with the growth of trees, so was it with the growth of business as a general rule before the great recent development of vast joint-stock companies, which often stagnate, but do not readily die. Now that rule is far from universal, but it still holds in many industries and trades. Nature still presses on the private business of limiting the length of the life of its original founders, part of their lives in which . . . their faculties retain full vigour. And so, after a while, the guidance of the business fall into hands of people with less energy and less creative genius, if not with less active interest in its prosperity. If it is turned into a joint-stock company, it may retain the advantage of division of labor, of specialized skill and machinery; it may even increase them by a further increase of its capital; and under favorable conditions it may secure a permanent and prominent place in the work of production. But it is likely to have lost so much of its elasticity and progressive
force, that the advantages are no longer exclusively on its side in its competition with younger and smaller rivals.¹

¹Appendix B of this paper is Marshall's description of the growth and decay of a firm within the context of an industry.
BIBLIOGRAPHY


Blair, John M. "Does Large-Scale Enterprise Result In Lower Costs?" American Economic Review, XXXVIII (May 1948), 121-52.


---


---


---


