Volume Title: Business Concentration and Price Policy

Volume Author/Editor: Universities-National Bureau Committee for Economic Research

Volume Publisher: Princeton University Press

Volume ISBN: 0-87014-196-1

Volume URL: http://www.nber.org/books/univ55-1

Publication Date: 1955

Chapter Title: Survey of the Empirical Evidence on Economies of Scale

Chapter Author: Caleb A. Smith

Chapter URL: http://www.nber.org/chapters/c0961

Chapter pages in book: (p. 213 - 238)

SURVEY OF THE EMPIRICAL EVIDENCE ON ECONOMIES OF SCALE

CALEB A. SMITH
BROWN UNIVERSITY

AT THE outset the reader should be warned that only after what may seem an over-long introduction have I attempted to carry out the commission assigned me: to survey the available empirical information on the variation of cost with size of plant and company, to appraise the validity of the literature on economies of large-scale production, and to indicate what generalizations it will support.

The time available proved inadequate to anything approaching a comprehensive review of the scattered and uneven material. Only a sample of the literature on economies of large-scale production has been consulted and summarized and a very limited appraisal of individual studies undertaken. The validity of the available empirical information on the variation of cost with size of plant has been weighed in general on the total information, rather than in detail.

Based on the survey, the conclusion is that the generalization that available empirical information supports are indefinite and disappointing. In part, this is because much information in government and trade association files has never been studied[1] but in part I suspect a more fundamental difficulty.

Whenever the best answer to any question adduced by empirical investigators tells them little, the question should be re-examined. Because I was disappointed in the answers to the question, "What generalization will the empirical evidence on economies of scale support?" I undertook the introductory analysis comprising so much of this paper. This re-examination may be divided into two parts.

[1] So far as I have been able to discover, very little information obtained by the national war agencies which might throw substantial light on this subject has been compiled and published in a usable form. The only exception of any importance is the Office of Price Administration Economic Data Series published by the Office of Temporary Controls. The material here presented, while it seems worthy of more study than it has received, is certainly less significant than the data we had reason to hope might emerge from the OPA files. The most needed research job today in the field of the relation of cost to size of plant and firm would be one done by a government agency with full access to the data accumulated during the war. If one of the small business committees of Congress, for instance, would put a modest research staff to work on this task, a monograph of very great value probably could be produced.

First, what sort of problems arise when we seek empirical evidence of a relationship developed in deductive theory?

Second, what related questions might have more significant answers?[2] These two questions will be examined in some detail in the following Introduction to this discussion.

1. *Introduction*

THE effort to obtain empirical evidence of relationships which have been developed in deductive economic theory encounters two sets of problems. The first centers around the widespread but fundamental misconception as to the nature of emiprical facts. Certain empirical facts—the length of a table (under specified conditions), for instance—can be defined operationally. (That is, the process of measurement against a specified standard of length may be described.) Such facts have validity for any use where the operational process of measurement and the standard used are acceptable.

But there is another sort of empirical fact which has meaning only in terms of a complicated conceptual framework. That a particular empirical study shows or does not show economies of scale is an example of this. Only in terms of a carefully delineated concept can we say that particular evidence shows or does not show economies of scale. The Introduction to this paper first considers the concept of economies of scale and distinguishes this from concepts with which it easily may be confused in empirical work.

The second set of problems centers around the elimination of variations in cost which do not result from differences in size of plant or firm. Simplifying assumptions are essential to the development of theoretical concepts. Inevitably, however, each simplifying assumption blocks the path toward an empirical investigation of the relationship which the theory states. In empirical investigation the complexities cannot be removed by a simple declarative sentence; nor can the empirical economist—like his empirical brethren in the laboratory sciences—remove complicating factors by carefully regulated experiments in which the factors are held constant. The second subsection of the Introduction to this paper discusses some of the more serious problems posed by factors other than scale which influence costs in the empirical studies of economies of scale.

The final subsection of the Introduction considers some questions

[2] The author claims no originality in questioning the question asked. Such questions have been posed by the authors of some of the studies surveyed in the preparation of this paper.

closely related to economies of scale, in the hope of finding a more promising line of inquiry.

To give some precision to this discussion, economies of scale may be defined as equivalent to a falling long-run average cost function.[3] These economies can be considered either with respect to size of plants or of firms. The long-run average cost function of economic theory shows the long-run relationship between average cost and the output of one homogeneous product.

Questions of definition arise with respect to the terms "average cost" and "size." Average cost includes (in the economist's definition) the imputed cost of capital or any other services supplied by the owners. Empirical data generally follow accounting practice and exclude dividend payments and other payments of "profits" to owners from costs. Unfortunately, there is no reason to assume that costs covered by dividends are the same for all plants or firms making the same product. In fact there are reasons to think that there may be systematic variations in these costs with size of firm. Salaries of owners who are officers of corporations which show little relation to value of services rendered pose a similar problem of a possible systematic variation between costs recorded by accountants and the economist's costs for firms of different size. When using empirical material to test economic theory, we must not forget these different concepts of cost.

To measure size in empirical studies of economies of scale is even more difficult. The measurement of output is unequivocal only if the output is homogeneous. In practice we do not find either plants or firms which, during a period of growth from small-scale to large-scale, produced one homogeneous product, nor do we find a group of plants or firms of widely different size which produce a single homogeneous product.[4] Output of plants and firms is in fact heterogeneous to a very substantial extent. Since the definition of economic theory has little relation to reality, let us explore the implicit definition of common usage.

When we talk about the size of a plant or firm we ordinarily mean its capacity to turn out its entire product mix, not its capacity to

[3] The long-run average cost function is the envelope curve to the short-run average cost functions. Thus the capacity or size of a plant or of a firm is that output for which the short-run cost function and the long-run cost function coincide.

[4] The reader should not conclude that data from even such plants or firms would, without surmounting further problems, yield an empirical counterpart of the theoretical long-run cost function.

turn out one specific product. When we talk about cost in relation to size we ordinarily mean the cost of a specific product. We thus pose two problems: (1) What do we mean by cost for one of a group of products? (2) What are we doing when we relate the cost of one product to capacity to produce a multiplicity of products?

The cost of a number of products produced by a firm can be determined only on the basis of arbitrary allocations. In practice, empirical studies of economies of scale have accepted the cost allocations made by the firms studied. Unfortunately, the cost accounting techniques used by business are not nearly so highly standardized in most industries as are income accounting techniques. These have been subjected to at least two powerful standardizing influences which have not affected cost accounting techniques: the rules of the Bureau of Internal Revenue, and the pronouncements of the American Institute of Accountants, which shape accounting practices through the institution of outside auditors. In spite of the standardizing influences to which income accounting techniques are subject, the lack of comparability of balance sheet and income statement data of different firms lead the author of a well-known text on financial statement analysis to issue this warning: "The figures of one enterprise may be compared with those compiled for another only with great care. The combination of the financial statement data of different enterprises for statistical studies is usually unsatisfactory."[5]

It is extremely unlikely that the cost figures obtained from the accounting records of a group of firms are really comparable. The student of empirical data on economies of scale is seldom in a position to make the figures comparable; he can only hope that there will be no incomparability systematically related to size. The dangers in using cost accounting figures in empirical studies of economies of scale are, I believe, greater than most economists realize, because their lack of familiarity with accounting practice leads them to underestimate the uncertainties of cost data.

The second problem posed by the comparison of cost for one product with size measured in terms of a multiplicity of products is a conceptual one. What relation, if any, is there between the concept of economies of scale as defined in economic theory (a relation between cost and size in plants or firms producing homogeneous out-

[5] John N. Meyer, *Financial Statement Analysis* (2nd ed., Prentice-Hall, 1952), p. 44.

put) and either (1) the concept of a relation between the cost of producing one of many products and the size of plants or firms measured in terms of composite output, or (2) the concept of a residual relation between cost and size after allowing for the effect of variations in other dimensions of output?

If the product mix of the composite output for the different observations of cost and size is highly similar, then we are seeking a relationship which might be regarded as similar to the concept of economies of scale as defined in economic theory. The problem is to distinguish sufficiently similar product mixes from those which are diverse. Here we must rely upon our own judgment and that of the investigator.[6]

Even if we accept similar product mixes[7] as the output in a study of the long-run cost output relation, the problem of measuring output remains. If all the elements of the product mix occurred in the same proportion at all different sizes of plant or firm any element might be used as the measure of output, and we might as well regard the output as homogeneous. The real problem of measurement of output arises because the elements of the product mix occur in different proportions for the different observations of cost and size. The use of a number of different dimensions of size, which might appear to be a way around this problem, is explored later and found to be generally impractical. Perhaps the most practical method of measuring the amount of somewhat heterogeneous product mix is to use the familiar though arbitrary common denominator of economics—money value. We may at times wish to question price as a measure of output but at least it has the merit of being a market evaluation.[8]

There are at least two ways to get around the two problems posed

[6] Unfortunately, once an economist discovers data which might be used for an empirical study he is strongly inclined to use them, if it is at all possible, and to present the study "for whatever it is worth." Others, less familiar with the problems posed by the basic data, are likely to overrate the significance of the study.

[7] What should be meant by product mix similarity? To require similar percentages of identical products would be too restrictive. Product mixes might better be regarded as highly similar even though no product in one is homogeneous with any part of the other if all the products are highly similar and are produced in highly similar percentages. Thus the output of Fords and of Chevrolets might be regarded as similar product mixes.

[8] The use of this common denominator has the added practical advantage that general price-level changes will tend to move an observation of cost and size more or less along the function rather than almost at right angles to it as would occur if a physical measure of output were used with a money measure of costs.

by cost allocation and the relation between the cost of one part of output and size in terms of a composite measure of output. The first is to study the relation between cost per composite unit of output and scale in terms of the composite unit, i.e. if money value is used as the composite unit, the cost of a dollar of output at various dollar scales of output. The second is to regard output, and hence size of plant, as having many dimensions. Each of these routes around the problems will be explored in some detail.

The difficulties with studies of cost per dollar of output in relation to capacity measured in dollars of output are obvious. Different firms may charge different prices for the same product. More serious questions arise when the price differences between similar products do not fairly represent differences in the "quantity" of output. In spite of these obvious difficulties, studies relating cost per dollar of sales to dollar capacity for plants or firms which have similar product mixes may be preferable to those seeking to relate cost for a particular product to either a physical or a dollar measure of capacity. Further, the data needed to determine costs per dollar of output in relation to scale measured in dollars of sales are more generally available than are data needed to determine the relation between costs per unit of a particular physical output and size, though empirical economists have been afraid to use them. This reluctance may rest on no more secure basis than a greater familiarity with variations in price for the same product than with variations in costs allocated to the same product by different cost accounting systems or with the variability of product mixes within an industry.

The second way around the problems posed by cost allocation and by relating cost of a part of output to capacity to produce many different products is to regard output as having many dimensions. In a study of costs for airlines, Allen R. Ferguson explores this method.[9] Dr. Ferguson, in summarizing his study, says that it "avoids the assumption of a homogeneous product and deals explicitly with the problems of varying the quality and the product mix of output."[10] Explicitly, the product is not conceived of as simply ton-miles or passenger-miles; speed and length of flight are introduced as measures of output characteristics. The introduction of additional dimensions of output results in complications which are more than computational.

[9] Allen Richmond Ferguson, "A Technical Synthesis of Airline Costs" (Ph.D. dissertation, Harvard University, 1949).

[10] *Ibid.*, Summary of thesis, p. 1.

On the surface, the problem of measuring somewhat heterogeneous output appears to be more manageable conceptually when the idea of several dimensions of output is introduced, but here new problems arise. First, what is the significance of one of several dimensions of output? Second, can the data support statistical procedures for separating the effects of variations in several dimensions?

We must realize that not all dimensions of output measure the size of plant or firm. It is easy to conceive of a scale relationship between costs per ton-mile and "capacity" in ton-miles in which all other dimensions of output are held constant; but to conceive of the relation between costs per mile per hour and capacity in miles per hour *as a scale relationship* is bizarre. (Incidentally, capacity in miles per hour is a dimension for which cost increases rapidly with large "scale.") The whole problem of what is meant by "scale" is cast in a different light when viewed as the dimensions of a single output. The usual notion of a larger plant or firm involves the production of a greater number of identical units of output. It is better to regard the other dimensions of output as elements in the heterogeneity of output. Interesting relations between cost and any of these dimensions of output may be discovered empirically, but they should not be regarded as cost-size relations.

It may be suggested that empirical study should seek an over-all relationship between cost and all the various dimensions of output and size. One phase of this complex would be the cost-size relation. Output has too many dimensions to make this approach seem promising as an empirical method of deriving a cost-size relation. Ordinarily, observations are too few to indicate the shape of a cost-size relation after allowing for the effect of many other variables.

The multiplicity of dimensions of product is fantastic. Simply in terms of its physical characteristics, a product always has more than one dimension, and the dimensions of a product cannot be so limited. Among its other dimensions are consumer and trade acceptance of the brand name and the characteristics of its distributive system. Thus, although it may be possible to describe the outputs of a steel rolling mill by many fewer dimensions than the number of products, even the minimum number of dimensions probably would leave the problem of deriving a cost-output function statistically unmanageable.

Competing products may have quite different dimensions in these respects. The possibility of having a larger firm or plant may depend upon the existence of different dimensions for the product of the

large-scale firm. A good example of the different nonphysical dimensions of competitive products is found in a study of costs incurred by fourteen manufacturers of rubber tires. The data are presented in the *Survey of Rubber Tire and Tube Manufacturers.*[11]

In August 1943 for the 6.00-16 4-ply synthetic rubber passenger car tire the selling, general, and administrative expense for the four largest manufacturers was 65 cents more per tire than for the ten other manufacturers studied. This 69 per cent higher cost indicates a different dimension of the product. Whether or not we enjoy listening to the "Firestone Hour" we must recognize that it helps give the Firestone tire a different dimension; that is, it makes it a different product. The higher cost of selling and administration is not just the result of the larger size of the firm, but of expenses undertaken to differentiate the product. These expenses may be necessary in order to have the larger firm or they may be profitable only for the larger firms. If the former, then the data may be regarded as revealing the cost-size relation for firms producing this product. On the other hand, if they are not necessary to the maintenance of the larger firms but are profitable for the larger firms though they wou'd not be profitable for smaller firms, the data pertain to two products too dissimilar to reveal a cost-size relation.

2. *Methods Used to Handle Cost Variations from Causes Other than Economies of Scale*

IDEALLY, empirical evidence on economies of scale should be obtained by observing the variations in cost associated with different scales of plant or firm with all other cost influences constant. Since it is obviously impossible to find any such situation, the relation between average cost and the scale of plant or firm must be sought by other means. Two methods of study have been used which may be characterized as the statistical approach and the engineering approach.

In the statistical approach, the costs of the plant or firm as a unit or the costs allocated to some type of output are related to size. Other influences on cost either are ignored or allowed for by such techniques as deflation or multiple correlation. Though the details will not be discussed here, some inherent weaknesses of this approach will be considered.

[11] Office of Temporary Controls, OPA Economic Data Series 10, 1947. A portion of this data is presented in an article by John M. Blair, "Does Large Scale Enterprise Result in Lower Costs?" *American Economic Review*, May 1948, p. 149.

First, statistical data show costs in relation to scale for the many different technologies actually used by plants or firms for which sufficient empirical evidence is available to make it possible to include them in a statistical study. If the technologies used by the different plants varied only because different sizes of plant require different technology, the data would be appropriate. But technology varies from plant to plant for other reasons. For example, some plants are old, others newly built while technological horizons have changed from year to year. Further, technologies were selected at various times because of different relative factor prices, or because of different demand expectations, etc. Statistical studies have limited significance because they regarded all or most of the plants or firms currently classified as part of the "industry" as sufficiently homogeneous in technology to warrant grouping them together to derive a long-run cost function.

Another difficulty with the statistical approach is that it assumes that each cost-size observation used represents a point on the long-run cost function; that is, that every output studied is the optimal output for that plant or firm. This is an heroic assumption, but most studies make no attempt to eliminate any observations because they represent obviously nonoptimal outputs. There is, perhaps, a tacit assumption that at any one time all plants or firms are operating in the same relation to optimal output and that their nonoptimal function is similar to the optimal cost function. Simply to state these assumptions reveals their inherent dangers.

In the engineering approach, each element of the production process is studied to discover the relation between inputs and outputs at different scales for that process. The input-output relations of the processes are then combined to give the over-all input-output relations. The introduction of prices for the inputs transforms these relations into cost-output relations. Since this method of study is less familiar, two studies of economies of scale made from the engineering point of view will be discussed as examples of the method. These studies are presented in unpublished doctoral theses written independently but at about the same time at Harvard.

The first of these studies, already mentioned, was submitted in April 1949 by Allen Richmond Ferguson and is entitled "A Technical Synthesis of Airline Costs." The second, "Engineering Bases of Economic Analysis," was submitted in August 1949 by Hollis B. Chenery. Both studies apply parts of the great mass of engineering observations of the relations between inputs and outputs to an

analysis of over-all cost functions of both the short- and long-run variety. The technique of using engineering laws in discussions of economies of scale is not new[12] except in the thoroughness and precision with which it is applied, but this exact use opens exciting new vistas for the further study of economies of scale which can be made on the basis of the empirical relations developed by engineers.

Chenery, in the summary of his thesis, says, "The purpose of this study is to determine the usefulness of physical laws for the economic analysis of production. It seeks to develop a method by which the type of calculation made by engineers in designing plants and equipment may be used to derive the general relations among productive factors expressed in the production function of economic theory."[13]

Ferguson's conception of the subject for investigation, though it is essentially the same, is broader. He is concerned not only with the "technical" but also with the "institutional determinants of the amount of each type of input required and [with] ascertaining the quantitative input-output relationship so determined."[14] The author recognizes that the institutionally determined input-output relations are subject both to arbitrary changes and to changes which may be induced by a large or sudden change in this input or in other inputs. However, the inclusion of what have sometimes been called human engineering relations in the purview of engineering studies considerably broadens the usefulness of the technique.

Studies made by economists on the basis of observations by engineers avoid the problem which arises from the fact that the existing plant was built when different technological horizons existed, but they have similar limitations. They have been over-oriented toward the study of input-output relations in the elements of the presently utilized productive processes because these relations are the only ones thoroughly investigated by the engineers. The authors have recognized explicitly the fact that this narrowed the economic significance of their studies. The importance of this limitation is greatest if a change in scale or factor prices makes it more economical

[12] Economic discussions frequently have pointed out engineering laws which may lead to either economies or diseconomies of scale, e.g. that heat loss is proportional to the square of any one dimension while volume is proportional to the cube.

[13] Hollis B. Chenery, "Engineering Bases of Economic Analysis" (Ph.D. dissertation, Harvard University, 1949), Summary, p. 1.

[14] Ferguson, *op. cit.*, p. 2.

to adopt techniques, the input-output relations of which have not been studied. Furthermore, engineering studies show ideal rather than actual relations of size and cost. This is good or bad, depending on what we want the cost function to reveal. Finally, the relation of factor cost to scale, if not explicitly studied, further limits the usefulness of the engineering studies.

The problem of cost and size of plant is more susceptible to study through the engineering approach than is the problem of cost and size of firm because the relations between the plants which make up a firm do not lend themselves to engineering study. These relations probably are dominated by the more or less unique considerations for each aggregation of plants into a firm.

3. *The Question to Which We Seek an Answer*

WHEN we ask the supposedly precise scientific question, what is the relation of cost to size of plant or firm, we are usually concerned in fact with finding answers to questions which appear less precise: Are giant firms more efficient or do they prosper because of "unfair" advantages? How much saving does the public get from giant firms? How?

Examination of the data that are available and that conceivably might be available shows that we cannot hope to make very satisfactory empirical studies of the long-run cost function. I believe that if we asked the student of the data to tell us in detail just what cost differences exist between different types and sizes of plant and firm and what causes, if any, he could discover for those differences, the information would go farther to clarify the practical questions to which we seek answers than would studies of the relation of cost to size.

4. *Comments on a Sample of the Empirical Studies*

BEFORE venturing on a statement of the generalizations which the empirical evidence warrants, some comments on a sampling of the material which has become available since 1940 will be presented. Earlier material on the subject was sampled in *Cost Behavior and Price Policy*,[15] published ten years ago by the National Bureau of Economic Research. This discussion provides, in my opinion, an adequate treatment of the material prior to 1940. The material published since 1940 enables us to fill in some portions of the picture

[15] Committee on Price Determination, Conference on Price Research, National Bureau of Economic Research, New York, 1943.

drawn there. We still lack sufficient data for confident and precise generalization.

In 1941 the Temporary National Economic Committee Monograph 13, *Relative Efficiency of Large, Medium-Sized, and Small Business*, a study prepared by the Federal Trade Commission was published. This presents substantial material each part of which, however, is rather briefly and superficially analyzed. The results show that, in general, medium-sized business[16] is most efficient in the industries and instances studied. The material on costs of individual plants and companies in a number of industries shows that very seldom (1 out of 59 cases for companies, and 2 out of 53 cases for plants) did the largest plant or company show the lowest costs. These results, unfortunately, prove little since the much greater number of medium-sized, small and very small plants and companies included biases the results. This may be, in part, the result of more frequent accounting abnormalities among small and medium-sized businesses. Furthermore, we should expect to find that some small or medium-sized plants or companies had especially favorable cost conditions. These facts—coupled with a lack of detailed discussion and analysis of the material presented— (much of it rather sketchy) prejudice the scholarly mind against accepting the conclusions which seem to follow from the data in the monograph.[17]

On the other hand, a second look at the data reveals somewhat better evidence in support of the idea that small or medium-sized business is more efficient. (1) "On the average, over one-third of the companies in every array" and "over one-third of the plants in each cost array had . . . costs lower than that of the largest company" or "plant."[18] (2) When data for companies or plants grouped according to size were used, in ten out of eleven cases when companies were studied, and in all five cases when plants were studied, the medium-sized group showed lowest costs.[19] (3) A few quick calculations on some of the cost data presented in arrays but not included in the grouped data mentioned above, show instances—although

[16] The classification "medium-sized" is sometimes strained, e.g. Chrysler is called medium-sized when clearly the significant difference between Chrysler and Ford is in degree of vertical integration.

[17] For a reasoned and highly critical review of this monograph see John M. Blair, "The Relation between Size and Efficiency of Business," *Review of Economic Statistics*, August 1942, pp. 125-135.

[18] *Relative Efficiency of Large, Medium-Sized and Small Business*, Temporary National Economic Committee, Monograph 13, 1941, p. 12.

[19] *Ibid.*, p. 12.

not an overwhelming predominance of cases—in which grouping the plants or companies would result in the medium-sized plants showing the lowest costs. The mass of data is so great and the conclusion of lower costs for medium-sized plants or companies so general that in spite of the fact that almost every individual study is subject to serious criticism it is necessary to give considerable weight to the findings.

There may be a bias in the data on cost and size of firm which prejudices our conclusions. It is entirely possible that, although almost always we find costs for the largest firms higher than for a group of medium-sized firms, it is not general. In those industries where cost continues to decrease with increasing size it is probable that all the medium-sized firms either have become giants and swallowed the other medium-sized firms or they have failed or shrunk into small firms. If this is the case, the sound generalization is not that medium-sized firms have, in general, lower costs than do large firms, but that in industries where both medium-sized and large-sized firms are found, the costs of the medium-sized firms are probably lower than those of the large firms. (But, let us not forget that "medium-sized" here includes the Chrysler Corporation.)

A study by Steindl presents some interesting ideas and evidence on the general subject of size of plants and firms although it offers little bearing directly on our question. He shows, by the use of data from the *Statistics of Income*, that capital intensification accompanies increasing size of firm for all manufacturing industry as a group, for mining, for trade, and for most of the major subgroups of manufacturing. He rejects as unlikely the possibility that this showing of greater capital intensity for larger firms is caused only by differences in the products produced by small and large firms or by greater vertical integration of larger firms.[20] Steindl devotes a considerable part of his study to demonstrating that with capital intensification, the profit rate will fall beyond a certain point even if cost per unit of output continues to fall. He thus offers a possible explanation of the declining profit rates frequently found by Crum[21] for the largest corporations in any industry which is consistent with the idea that unit costs are lower for larger companies. He also explores suggestively certain difficulties the large firm may face be-

[20] Joseph Steindl, *Small and Big Business* (Monograph 1, Oxford University Institute of Statistics, 1945), pp. 23-25.

[21] William Leonard Crum, *Corporate Size and Earning Power* (Harvard University Press, 1939).

cause of imperfect competition or oligopoly. The high selling and administrative costs of the four largest tire manufacturers, discussed earlier, is an example of this problem. He re-analyzes the material on the growth of concentration in manufacturing and shows that although the percentage of wage earners in manufacturing establishments with over 1,000 wage earners hardly changed from 1919 to 1937, establishments with 250 to 1,000 wage earners gained substantially relative to establishments with fewer than 50 wage earners.[22] This loss by small manufacturing business he regards as a significant continuance of the concentration pattern which was so marked before and during World War I.

Steindl, in his discussion of capital intensity, highlights the association of size with capital intensity. He asserts that "large-scale economies are in reality *technically* inseparable from capital intensification, so that the greater plant, if it is to make use of large-scale economies, has also to use a greater proportion of capital to labor."[23] While we may admit readily that many large-scale economies require much capital there seems no a priori reason why a small-scale plant using the best available technology and facing the same relative factor prices should use relatively less capital.

Greater capital intensity in large-scale plants could result if small-scale plants have inadequate capital resources, or if their managers believe they can find more profitable uses for capital in horizontal expansion rather than in capital intensification. If few small-scale enterprises want to put capital into intensive investment, the capital-intensive technology for small-scale plants will not be as adequately developed and the appropriate capital-intensive machines will not be readily available to them. There are, also, other reasons why small-scale technology may be less well developed in general than large-scale technology. On the basis of all these factors, we may conclude that, if capital intensification, as a rule, leads to lower unit costs, a systematic difference in the capital intensity of the technology between plants of different sizes may prejudice seriously the unit-cost size relations which we discover empirically.

Many numbers of the Office of Price Administration Economic Data Series published in 1947 by the Office of Temporary Controls contain material on cost and size of firm.[24]

[22] Steindl, *op. cit.*, Figure 1, p. 49. [23] *Ibid.*, p. 22.

[24] Study 10, on rubber tire and tube manufacturers, cited earlier, shows costs divided into several categories for nine products for two size groups of firms. Study 7 on retail furniture stores shows operating expenses as a percentage of

In a paper presented at the December 1947 meetings of the American Economic Association, John M. Blair analyzed data for several industries which show the lowest cost for groups of plants or companies smaller than the largest group.[25] On the basis of this evidence he seems ready to make the generalization that both plants and firms which are larger than the lowest cost size are found in practice in a substantial number of industries. His conclusion is especially interesting in view of his attack, referred to above, on the similar conclusions drawn in the TNEC Monograph 13. It should be noted that he attaches great significance to the growth of new "decentralizing" techniques which have improved the position of plants and firms of less than maximum size.

A study by Florence[26] presents a great deal of information on subjects related to that here under survey. He shows that in many industries the predominant size of plant is not large. More specifically, he shows that highly localized industry generally has medium-sized plants. If predominance of a plant size less than that of the largest firms may be taken as an indication of a certain sort of efficiency (even if not of lowest average cost) for medium-sized firms, his data clearly establish substantial areas in which medium-sized firms are "efficient."

There also is scattered evidence, some of it of very high quality, on the relation of cost and size. Unpublished Harvard doctoral theses by Ferguson and by Chenery have already been mentioned. There are undoubtedly similar theses at other universities. Other Harvard theses on this subject include:

John B. Lansing, "An Investigation into the Long-Run Cost Curves for Steam Central Stations," 1948, in which he concludes, "For a station under [theoretically ideal] conditions the answer is clear, 'No, the long-run cost curve does not turn up.' The important

---

sales for three size groups in metropolitan and nonmetropolitan areas. Study 12, on fresh fruit and vegetable wholesalers, gives cost figures for four cost categories for five size groups of wholesalers for each of three types of wholesaler. Study 13, on women's underwear and nightwear, gives cost figures for two different years by five cost categories for six size groups of manufacturers. Study 26 on grocers retail chains and wholesale, gives expense as a percentage of sales by size groups for a considerable number of time periods. Other studies in this series contain similar data. Detailed appraisal of this mass of material would be a major task which, so far as I know, has not been undertaken.

[25] Blair, "Does Large Scale Enterprise Result in Lower Costs?" as cited.

[26] P. Sargent Florence, *Investment, Location, and Size of Plant* (Cambridge, 1948).

question is different: what factors tend to make long-run cost curves turn up?"[27]

Morris A. Adelman, "The Dominant Firm," 1948, a study of the Great Atlantic & Pacific Tea Co. A chart from a revision of this thesis supplied by the author shows a regular decline with increasing size in practically all elements of costs for the company's supermarkets.

Raymond G. Bressler, Jr., "City Milk Distribution," 1946. Also a study, *Economies of Scale in the Operation of Country Milk Plants*, published in 1942 by the New England Research Council on Marketing and Food Supply in cooperation with the New England Agricultural Experiment Stations and the Department of Agriculture, and an article "Research Determination of Economies of Scale," *Journal of Farm Economics*, August 1945. The emphasis in Bressler's thesis is on the analysis of the elements of the costs of urban milk distribution with regressions fitted to scatters between short-run variations in output and cost. Observations on the long-run cost function are derived from these short-run considerations.

Finally, mention should be made of three largely unexplored sources of empirical evidence. First, the publications of the Agricultural Experiment Stations contain a substantial volume of evidence on this subject pertaining not only to agriculture but to first stage assembling and processing. Second, the engineering journals contain occasional articles giving empirical relations. The data often are not described carefully but a thorough search would reveal much interesting information. An example is an article by J. G. Berger, "Does a Laundry Cut Costs by Buying or by Generating Its Electricity?"[28] The article presents "Curves A and B [which] indicate cost of purchased and generated power based on actual average laundry experience." The curves show falling cost per kw. hr. with increasing size for both with generated power cost falling below purchased power at 3,700 kw. hr./month and—what is more significant—still falling appreciably at 15,000 kw. hr./month, the largest size shown.

The third neglected source is the data compiled and published by the agencies regulating railroads and public utilities. A careful study of a part of this data has been made by my colleague, George H. Borts, in a thesis at the University of Chicago. Borts develops

[27] P. 56.

[28] *Power*, July 1946, p. 457.

long-run production functions in railroading, showing declining costs with increased size up to the maximum size.[29]

5. *Conclusion*

IT IS both difficult and dangerous to generalize on the basis of the scattered and heterogeneous empirical material on economies of scale. The following generalizations, however, appear to be warranted by the evidence I have examined in the preparation of this paper and in my connection with the preparation of Chapter X of the National Bureau of Economic Research's study "Cost Behavior and Price Policy."

1. With increasing size of plant, at least from small to medium size, average cost of production declines as size increases if factor costs are held constant.

2. There is no substantial evidence that the decline in unit costs stops before the maximum size of plant available for study if factor costs are held constant and the product is the same. On the other hand, the little evidence available does not refute the idea that the long-run cost curve even with factor prices constant turns up at some attainable size. We can hardly hope to find an answer to this question as to whether there is in practice a plant so large that the costs of producing a specific product increase even if factor prices are held constant because:

a. Factor costs, especially labor costs, seem to vary with size of plant. Generalization is hazardous and must be based primarily upon studies not concerned with determining the long-run cost function because most of the studies of the long-run cost function have not given this problem careful consideration. Generally, wage rates are found to be higher in the larger plants while probably raw materials (exclusive of assembly costs) and almost certainly capital, are cheaper.

b. Assembly costs and distribution costs per unit decline for a time with increasing size of plant but usually start to increase within the range of size of plant available for study.[30]

These increases in factor prices and in assembly and distribution

[29] See also his articles "Production Relations in the Railway Industry," *Econometrica*, January 1952, and "Increasing Returns in the Railway Industry," *Journal of Political Economy*, August 1954.

[30] This conclusion is based primarily on studies of plants engaged in first stage agricultural processing, but there is no evidence to indicate that it is not generally applicable.

costs result in cost increases which make it impractical to build plants which might be large enough to have higher average cost. Therefore, we have no opportunity to study the costs of such giant plants.

The hypothesis that the long-run cost function for the production of a product typically turns up at some very large size cannot be subjected to empirical verification. Even if one or two products should be found for which the cost elements which are supposed to be impounded in *ceteris paribus* do not in practice increase so as to prevent practical businessmen from expanding and the giant plants which they then built showed higher unit costs (*ceteris paribus*), it would be foolhardy to generalize on these instances. Furthermore, if the range in size of plants in practice stops (because factor prices and assembly and distribution costs increase) within the range in which the cost of producing a specific product is still decreasing when factor prices are held constant, the hypothesis that the long-run cost function eventually turns up is not very meaningful even if it could be proved. The fact that there is no substantial evidence that the long-run cost function for plants does not continue to decline up to the largest sizes found in practice if factor prices are held constant (a generalization apparently justified by the evidence) seems to have much greater significance for economists.

3. With increasing size of firm, the best documented generalization about average cost is that when factor costs are not held constant and outputs which are not the same but are sold competitively are considered as similar, costs decline with increasing size of firm up to a rather high point but that frequently and perhaps generally beyond a certain size of firm, costs again increase. But there is no satisfactory basis for distinguishing types of product for which the long-run cost function of the firm rises within the range of firm size actually found in practice.

COMMENT

MILTON FRIEDMAN, University of Chicago

I HAVE great sympathy with Caleb Smith's conclusion that the right questions have not been asked of the data on the costs of firms of different sizes. My quarrel with him is that he does not go far enough. I believe that cross-section contemporaneous accounting data for different firms or plants give little if any information on so-called economies of scale. Smith implies that difficulty arises because the

observed phenomena do not correspond directly with the theoretical constructs; because there is no single, homogeneous product, and so on. I believe that the basic difficulty is both simpler and more fundamental; that the pure theory itself gives no reason to expect that cross-section data will yield the relevant cost curves. Some of the bases for this view are suggested by Smith in his discussion, but he stops short of carrying them to their logical conclusion.

NO SPECIALIZED FACTORS OF PRODUCTION

LET US consider first the simplest theoretical case, when all factors of production are unspecialized so there are numerous possible firms all potentially alike. This is the model that implicitly or explicitly underlies most textbook discussions of cost curves. For present purposes, we may beg the really troublesome point about this case—why there is any limit to the size of the firm—and simply assume that there is some resource ("entrepreneurial ability") of which each firm can have only one unit, that these units are all identical, and that the number in existence (though not the number in use) is indefinitely large, so all receive a return of zero.

In this case, the (minimum) average cost at which a particular firm can produce each alternative hypothetical output is clearly defined, independently of the price of the product, since it depends entirely on the prices that the resources can command in alternative uses. The average cost curve is the same for all firms and independent of the output of the industry, so the long-run supply curve is horizontal, and hence determines the price of the product.[1] In the absence of mistakes or changes in conditions, all firms would be identical in size, and would operate at the same output and the same average cost. The number of firms would be determined by conditions of demand. In this model, the "optimum" size firm has an unambiguous meaning.

Suppose this model is regarded as applying to a particular industry. Differences among firms in size (however measured) are then to be interpreted as the result of either mistakes or changes in circumstances that have altered the appropriate size of firm. If "mistakes" are about as likely to be on one side as the other of the

[1] This neglects some minor qualifications, of which two may deserve explicit mention: first, the irrelevance of the output of the industry depends somewhat on the precise assumptions about the source of any increased demand; second, strictly speaking, the supply curve may have tiny waves in it attributable to the finite number of firms. On the first point, see Richard Brumberg, "*Ceteris Paribus* for Supply Curves," *Economic Journal*, June 1953, pp. 462-463.

"optimum" size, the mean or modal size firm in the industry can be regarded as the "optimum"; but there is no necessity for mistakes to be symmetrically distributed, and in any event this approach assumes the answer that cross-section studies seek.

What more, if anything, can contemporaneous accounting data add? Can we use them to compute the average cost curve that was initially supposed to exist? Or even to determine the size of firm with minimum average cost? I think not. Consider a firm that made a "mistake" and is in consequence, let us say, too large. This means that the average cost per unit of output that would currently have to be incurred to produce the firm's present output by reproducing the firm would be higher than the price of the product. It does not mean that the current accounting cost is—even if there have been no changes in conditions since the firm was established, so that original cost corresponds to reproduction cost. If the firm has changed hands since it was established, the price paid for the "good will" of the firm will have taken full account of the mistake; the original investors will have taken a capital loss, and the new owners will have a level of cost equal to price. If the firm has not changed hands, accounting costs may well have been similarly affected by write-downs and the like. In any event, cost as computed by the statistician will clearly be affected if capital cost is computed by imputing a market return to the equity in the firm as valued by the capital market. In short, differences among contemporaneous recorded costs tell nothing about the *ex ante* costs of outputs of different size but only about the efficiency of the capital market in revaluing assets.

In the case just cited, data on historical costs would be relevant. However, their relevance depends critically on the possibility of neglecting both technological and monetary changes in conditions affecting costs since the firms were established. A more tempting possibility is to estimate reproduction costs. This involves essentially departing from contemporaneous accounting data and using engineering data instead, in which case there seems little reason to stick to the particular plants or firms that happen to exist as a result of historical accidents.

Under the assumed conditions, the unduly large firms would be converting themselves into smaller ones, the unduly small firms into larger ones, so that all would be converging on "the" single optimum size. Changes over time in the distribution of firms by size might in this way give some indication of the "optimum" size of firm.

SPECIALIZED FACTORS OF PRODUCTION

THE existence of specialized factors of production introduces an additional reason why firms should differ in size. Even if output is homogeneous, there is no longer, even in theory, a single "optimum" or "equilibrium" size. The appropriate size of firm to produce, say, copper, may be different for two different mines, and both can exist simultaneously because it is impossible to duplicate either one precisely—this is the economic meaning of "specialized" factors. Or, to take another example, Jones's special forte may be organization of production efficiently on a large scale; Robinson's, the maintenance of good personal relations with customers; the firm that gives appropriate scope to Jones's special ability may be larger than the firm that gives appropriate scope to Robinson's. It follows that in any "industry," however defined, in which the resources used cannot be regarded as unspecialized, there will tend to be firms of different size. One could speak of an "optimum distribution of firms by size," perhaps, but not of an "optimum" size of firm. The existing distribution reflects both "mistakes" and intended differences designed to take advantage of the particular specialized resources under the control of different firms.

The existence of specialized resources not only complicates the definition of "optimum" size; even more important, it makes it impossible to define the average cost of a particular firm for different hypothetical outputs independently of conditions of demand. The returns to the specialized factors are now "rents," at least in part, and, in consequence, do not determine the price, but are determined by it. To take the copper mine of the preceding paragraph, its cost curve cannot be computed without knowledge of the royalty or rent that must be paid to the owners of the mine, if the firm does not itself own it, or imputed as royalty or rent, if the firm does. But the royalty is clearly dependent on the price at which copper sells on the market and is determined in such a way as to make average cost tend to equal price.

The point at issue may perhaps be put in a different way. The long-run conditions of equilibrium for a competitive firm are stated in the textbooks as "price equals marginal cost equals average cost." But with specialized resources, "price equals marginal cost" has a fundamentally different meaning and significance from "price equals average cost." The first is a goal of the firm itself; the firm seeks to equate marginal cost to price, since this is equivalent to maximizing

its return. The second is not, in any meaningful sense, a goal of the firm; indeed, its avoidance could with more justification be said to be its goal, at least in the meaning it would be likely to attach to average cost. The equality of price to average cost is a result of equilibrium, not a determinant of it; it is forced on the firm by the operation of the capital market or the market determining rents for specialized resources.

Consider a situation in which a group of competitive firms are all appropriately adjusted to existing conditions, in which there is no tendency for firms to change their output, for new firms to enter, or for old firms to leave—in short, a situation of long-run equilibrium. For each firm separately, marginal cost (long-run and short-run) is equal to price—otherwise, the firms would be seeking to change their outputs. Suppose that, for one or more firms, total payments to hired factors of production fall short of total revenue—that average cost in this sense is less than price. If these firms could be reproduced by assembling similar collections of hired factors, there would be an incentive to do so. The fact that there is no tendency for new firms to enter means that they cannot be reproduced, implying that the firms own some specialized factors. For any one firm, the difference between total receipts and total payments to hired factors is the rent attributable to these specialized factors; the capitalized value of this rent is the amount that, in a perfect capital market, would be paid for the firm; if the firm were sold for this sum, the rent would show up on the books as "interest" or "dividends"; if it is not sold, a corresponding amount should be imputed as a return to the "good-will" or capital value of the firm. The equality between price and average cost, in any sense in which it is more than a truism, thus reflects competition on the capital market and has no relation to the state of competition in product or factor markets.

For simplicity, the preceding discussion is in terms of a competitive industry. Clearly, the same analysis applies to a monopolistic firm with only minor changes in wording. The firm seeks to equate marginal cost and marginal revenue. The capital market values the firm so as to make average cost tend to equal price. Indeed, one of the specialized factors that receives rent may be whatever gives the firm its monopolistic power, be it a patent or the personality of its owner.

It follows from this analysis that cross-section accounting data on costs tell nothing about "economies of scale" in any meaningful

sense. If firms differ in size because they use different specialized resources, their average costs will all tend to be equal, provided they are properly computed so as to include rents. Whether actually computed costs are or are not equal can only tell us something about the state of the capital market or of the accounting profession. If firms differ in size partly because of mistakes, the comments on the preceding simpler model apply; historical cost data might be relevant, but it is dubious that current accounting cost data are. And how do we know whether the differences in size are mistakes or not?

THE DEFINITION OF COST

THE preceding discussion shares with most such discussions the defect of evading a precise definition of the relation between total costs and total receipts. Looking forward, one can conceive of defining the total cost of producing various outputs as equal to the highest aggregate that the resources required could receive in alternative pursuits. Total cost so estimated need not be identical with anticipated total revenue; hence *ex ante* total cost, so defined, need not equal total revenue. But after the event, how is one to classify payments not regarded as cost? Does some part of receipts go to someone in a capacity other than as owner of a factor of production?

All in all, the best procedure seems to me to be to define total cost as identical with total receipts—to make these the totals of two sides of a double entry account. One can then distinguish between different kinds of costs, the chief distinction in pure theory being between costs that depend on what the firm does but not on how its actions turn out (contractual costs), and the rest of its costs or receipts (noncontractual costs). The former represent the cost of factors of production viewed solely as "hired" resources capable of being rented out to other firms; the latter represent payment for whatever it is that makes identical collections of resources different when employed by different firms—a factor of production that we may formally designate "entrepreneurial capacity," recognizing that this term gives a name to our ignorance rather than dispelling it.

Actual noncontractual costs can obviously never be known in advance, since they will be affected by all sorts of accidents, mistakes, and the like. It is therefore important to distinguish further between expected and actual noncontractual costs. Expected noncontractual costs are a "rent" or "quasi-rent" for entrepreneurial capacity. They are to be regarded as the motivating force behind the firm's decisions, for it is this and this alone that the firm can

seek to maximize. The difference between expected and actual noncontractual costs is "profits" or "pure profits"—an unanticipated residual arising from uncertainty.

Definitions of total costs that do not require them to equal total receipts generally define them as equal either to contractual costs alone or to expected costs, contractual and noncontractual, and so regard all or some payments to the "entrepreneurial capacity" of the firm as noncost payments. The difficulty is, as I hope the preceding discussion makes clear, that there are no simple institutional lines or accounting categories that correspond to these distinctions.

Smith mentions the possibility of relating cost per dollar of output to size. Presumably one reason why this procedure has not been followed is that it brings the problems we have been discussing sharply to the surface and in consequence makes it clear that nothing is to be learned in this way. If costs *ex post* are defined to equal receipts *ex post*, cost per dollar of output is necessarily one dollar, regardless of size. Any other result must imply that some costs are disregarded, or some receipts regarded as noncost receipts. Generally, the costs disregarded are capital costs—frequently called "profits." The study then simply shows how capital costs vary with size, which may, as Smith points out, merely reflect systematic differences in factor combinations according to size. One could with equal validity study wage costs or electricity costs per unit of output as a function of size.

The use of physical units of output avoids so obvious an objection; clearly it does not avoid the basic difficulty and, as Smith points out, it introduces problems of its own. The heterogeneity of output means that any changes in average cost with scale may merely measure changes in the "quality" of what is taken to be a unit of output. Insofar as size itself is measured by actual output, or an index related to it, a much more serious bias is introduced tending toward an apparent decline of costs as size increases. This can most easily be brought out by an extreme example. Suppose a firm produces a product the demand for which has a known two-year cycle, so that it plans to produce 100 units in year one, 200 in year two, 100 in year three, etc. Suppose, also, that the best way to do this is by an arrangement that involves identical outlays for hired factors in each year (no "variable" costs). If outlays are regarded as total costs, as they would be in studies of the kind under discussion, average cost per unit will obviously be twice as large when output is 100 as when it is 200. If, instead of years one and two, we substitute firms one and

two, a cross-section study would show sharply declining average costs. When firms are classified by actual output, essentially this kind of bias arises. The firms with the largest output are unlikely to be producing at an unusually low level; on the average, they are clearly likely to be producing at an unusually high level, and conversely for those which have the lowest output.[2]

SIZE DISTRIBUTION OF FIRMS

IT MAY well be that a more promising source of information than cross-section accounting data would be the temporal behavior of the distribution of firms by size. If, over time, the distribution tends to be relatively stable, one might conclude that this is the "equilibrium" distribution and defines not the optimum scale of firm but the optimum distribution. If the distribution tends to become increasingly concentrated, one might conclude that the extremes represented mistakes, the point of concentration the "optimum" scale; and similarly with other changes. Whether, in fact, such deductions would be justified depends on how reasonable it is to suppose that the optimum scale or distribution has itself remained unchanged and that the emergence of new mistakes has been less important than the correction of old ones. None of this can be taken for granted; it would have to be established by study of the empirical circumstances of the particular industry, which is why the preceding statements are so liberally strewn with "mights."

THE RELEVANT QUESTION

I SHARE very strongly Smith's judgment that one of the main reasons why the evidence accumulated in numerous studies by able people is so disappointing is that insufficient attention has been paid to why we want information on so-called economies of scale; foolish questions deserve foolish answers. If we ask what size firm has minimum costs, and define "minimum costs" in a sense in which it is in a firm's own interest to achieve it, surely the obvious answer is: firms of existing size. We can hardly expect to get better answers to this question than a host of firms, each of which has much more intimate knowledge about its activities than we as outside observers can have and each of which has a much stronger and immediate incentive to find the right answer: much of the preceding discussion is really only a roundabout way of making this simple point.

[2] This is the general "regression fallacy" that is so widespread in the interpretation of economic data.

But surely studies of this kind are not really directed at determining whether existing firms make mistakes in pursuing their own interests. The purpose is quite different. It is, I believe, to predict the effect on the distribution of firms by size of one or another change in the circumstances determining their interests. The particular question may well suggest relevant criteria for distinguishing one kind of cost from another, and in this way enable cross-section accounting data to provide useful information. For example, Smith discusses studies supposedly showing that assembly and distribution costs rise with the size of plant whereas manufacturing costs decline. This finding might be decidedly relevant to predicting the effect of a decline in transportation costs on the distribution of firms by size. Or, again, the fact that some firms may use different combinations of factors from others may be due to identifiable differences, geographical or otherwise, in the prices of what in some sense are similar factors. The combinations of factors employed by different firms may then be relevant information in predicting the effect of changes in factor prices. This is the implicit rationale of some of the studies of production functions.

In many cases, the changes in circumstances that are in question are less specific. What would be the effect, for example, of repealing the Sherman antitrust laws on the distribution of firms by size? Of eliminating patents, or changing the patent laws? Of altering the tax laws? As Smith says, there must be much evidence available that is relevant to answering such questions. Unfortunately, as he recognizes, the generalizations assembled by him at the conclusion of his paper do not make much of a contribution; in the main, they simply confirm either the absence of obvious discrepancies between the existing size of firms and the size that is in their own interests or the effectiveness of the capital market in writing off mistakes.