THE SIZE DISTRIBUTION OF BUSINESS FIRMS

By HERBERT A. SIMON AND CHARLES P. BONINI*

The distribution of business firms by size has received considerable attention from economists interested in the phenomena of competition and oligopoly and in the issues of government regulation to which these phenomena are relevant. That the size distribution of firms (whether within a single industry or in a whole economy) is almost always highly skewed, and that its upper tail resembles the Pareto distribution has often been observed, but has not been related very much to economic theory. Attempts at economic explanation of the observed facts about concentration of industry have almost always assumed that the basic causal mechanism was the shape of the long-run average cost curve; but there has been little discussion of why this mechanism should produce, even occasionally, the particular highly skewed distributions that are observed.

In Part I we shall discuss the adequacy of explanations of the size distribution based on the static cost curve. In Part II we shall propose an alternative theory based on a stochastic model of the growth process. In Part III we shall examine the empirical data in the light of the model. In Part IV we shall examine the implications of our analysis for public policy. In Part V we shall comment on some of the needs for empirical and theoretical research in this area.

I. *Economic Theory of the Size of Firms*

Economic theory has little to say about the distribution of firm sizes. In general, we are led to expect a U-shaped long-run cost curve or planning curve for a firm. But the scale corresponding to minimum costs need not be the same for different firms, even in the same industry. If we employ the concept of economic rent, we can say that firms will have the same minimum cost, but varying outputs at this cost [8, pp. 123-127]. If this is the case, the cost curve yields no prediction about the distribution of firms by size and no explanation as to why the observed distributions approximate the Pareto distribution.

* The authors are professor of administration and graduate fellow, respectively, in the Graduate School of Industrial Administration, Carnegie Institute of Technology. This analysis has been aided by a generous grant from the Ford Foundation for research on organizations. We are grateful to B. Wynne, who performed some of the exploratory analysis, and to F. Modigliani, with whom we have had several enlightening discussions about the theory.

Some theorizing has been concerned with long-run increasing, decreasing and constant cost curves for firms [10, esp. pp. 210-17]. But the theorizers have hesitated to draw conclusions about the observed size distributions. In some cases, the theory is indeterminate about the distribution, as in the case of constant costs [10, p. 211]. In others, the theorists point out that "industry" is such a vague and arbitrary term that comparing the sizes of different firms is like comparing oranges and apples. Differences in the size of markets for firms and the idea that firms are moving towards the equilibrium of the cost curve but haven't reached it are also mentioned as reasons why firms widely varying in size can survive in the same industry.

All these factors make static cost theory both irrelevant for understanding the size distributions of firms in the real world and empirically vacuous. And yet these distributions show such a regular and docile conformity to the Pareto distribution that we would expect some mechanism to be at work to account for the observed regularity.

In the previous discussion (as in much of the literature on this topic) our comments about the long-run cost curve have been completely a priori. Recently, J. S. Bain [2] has made a careful analysis of all the available information on the cost curves of firms and plants in a substantial number of industries, using both published and original data that he obtained by questionnaire. His data show that plant cost curves (ignoring the problem of intra-industry specialization) generally are J-shaped. Below some critical scale unit costs rise rapidly. Above the critical scale, costs vary only slightly with size of firm. Moreover, in only a very few industries (the typewriter industry is perhaps the most striking example) does the critical scale represent a substantial percentage of the total market. These facts correspond well with beliefs about these matters that are widely held by businessmen.

We can say, then, that the characteristic cost curve for the firm shows virtually constant returns to scale for sizes above some critical minimum—S_m. Under these circumstances, the static analysis may predict the minimum size of firm in an industry with a known value of S_m, but it will not predict the size distribution of firms.

II. *Stochastic Models of Firm Size*

In the context of a different theoretical framework, our limited knowledge of the shape of the long-run cost curve derived from static analysis might lead to much stronger predictions. This is, in fact, the case.

We postulate that size has no effect upon the expected percentage growth of a firm. We shall formalize this into the assumption (Gibrat's law of proportionate effect) that the distribution of percentage changes in size, over a year, of the firms in a given size class is the same for all size classes. That is to say, we assume that a firm randomly selected from those with a billion dollars in assets has the same probability of growing, say, 20 per cent, as a firm randomly selected from those with a million dollars in assets.

There are two reasons why this is a plausible assumption on economic grounds. First, it agrees with the empirical findings, as we shall discuss more fully at a later point. Secondly, if, as we have postulated, there exists approximately constant returns to scale (above a critical minimum size of firm) it is natural to expect the firms in each size-class to have the same chance on the average of increasing or decreasing in size in proportion to their present size.

Before discussing the model in detail, we should like to comment on the numerous related models that have been proposed in recent years for explaining various skewed distributions of economic variables—including income [3], wealth [11], sizes of firms [7], and sizes of labor unions [6]. It has often been noted that many economic variates—and not only firm size—have frequency distributions with highly skewed upper tails. In the past, these distributions have been most often approximated by the log-normal distribution or the Pareto curve—sometimes with quite good fit.

Now many of the simple and commonly used statistical distributions can be generated from simple stochastic models—the normal distribution, the Poisson, the exponential, and so on. A stochastic process (e.g., the simplest random walk [4, pp. 279-307]) that will generate the normal distribution of a variate will, of course, when applied to the logarithm of the variate, generate the log-normal. But in applying the assumptions to the logarithm of the variate, we have, in effect, assumed the law of proportionate effect.

We can state the same point in a different way. If we incorporate the law of proportionate effect in the transition matrix of a stochastic process, then, for any reasonable range of assumptions, the resulting steady-state distribution of the process will be a highly skewed distribution, much like the skewed distributions that have been so often observed for economic variates. In fact, by introducing some simple variations into the assumptions of the stochastic model—but retaining the law of proportionate effect as a central feature of it—we can generate the log-normal distribution, the Pareto distribution, the Yule distribution, Fisher's log distribution, and others [9, pp. 425-27]—all bearing a family resemblance through their skewness. Contrariwise, we generally get quite different steady-state distributions from stochastic processes that do not embody the law of proportionate effect, or some approximation to it.

For the moment, we prefer to emphasize the generic similarities rather than the specific differences among the various stochastic processes that incorporate the law of proportionate effect. The log-normal and the Pareto distribution have been most often discussed in the literature; our own investigations, and Champernowne's, have led more often to the class of distributions we have called the Yule distribution.

Let us assume that there is a minimum size, S_m, of firm in an industry. Let us assume that for firms above this size, unit costs are constant. Individual firms in the industry will grow (or shrink) at varying rates, depending on such factors as (a) profit, (b) dividend policy, (c) new investment, and (d) mergers. These factors, in turn, may depend on the efficiency of the individual firm, exclusive access to particular factors of production, consumer brand preference, the growth or decline of the particular industry products in which it specializes, and numerous other conditions. The operation of all these forces will generate a probability distribution for the changes in size of firms of a given size. Our first basic assumption (the law of proportionate effect) is that this probability distribution is the same for all size classes of firms that are well above S_m. Our second basic assumption is that new firms are being "born" in the smallest-size class at a relatively constant rate.

It has been shown elsewhere that under these assumptions the Yule distribution will be the steady-state distribution of the process [9, pp. 427-30]. Let $f(s)ds$ be the probability density of firms of size s. Then the Yule distribution is given by:

$$f(s) = K\mathrm{B}(s, \rho + 1), \tag{1}$$

where $\mathrm{B}(s, \rho + 1)$ is the Beta function of s and $(\rho + 1)$, K is a normalizing constant, and ρ is a parameter. It is easy to show that as $s \to \infty$,

$$f(s) \to Ms^{-(\rho+1)}, \tag{2}$$

which is the Pareto distribution. Hence the Pareto distribution approximates the Yule distribution in the upper tail.

The details of the derivation of the Yule distribution need not be repeated here [9, pp. 427-35]. What distinguishes the Yule distribution from the log-normal is not the first assumption—the law of proportionate effect—but the second—the assumption of a constant "birth rate" for new firms.[1] If we assume a random walk of the firms already in the system at the beginning of the time interval under consideration, with zero mean change in size, we obtain the log-normal. If we assume

[1] The otherwise excellent study by Aitchison and Brown [1, p. 109] is in error in supposing that Champernowne's model of income distribution would have yielded the log-normal instead of the Yule distribution if Champernowne had taken a continuous rather than a discrete model. The real difference between the models lies in the assumptions about boundary conditions.

a random walk, but with a steady introduction of new firms[2] from below, we obtain the Yule distribution.

The parameter, ρ, of the Yule distribution has a simple interpretation. Let G be the net growth of assets of all firms in an industry during some period, and let g be that part of the net growth attributable to new firms—firms that have reached the minimum size during the period. Then, it can be shown that:

$$\rho = \frac{1}{1 - g/G} = \frac{1}{1 - \alpha}, \quad \text{where} \quad \alpha = g/G. \tag{3}$$

Thus, if $g/G = .10$—new firms account for 10 per cent of the growth in assets in the industry—we will have $\rho = 1/(1 - .1) = 1.11$. In the limit, as the contribution of new firms to total growth approaches zero, ρ approaches 1. Although it is assumed in the derivation that α be a constant, a slow change in α can be expected to modify the steady-state distribution only slightly.

III. *The Empirical Data*

Since published empirical data on the distributions of firms by size are numerous and monotonously similar, we will limit ourselves to some illustrative figures. Whether sales, assets, numbers of employees, value added, or profits are used as the size measure, the observed distributions always belong to the class of highly skewed distributions that include the log-normal and the Yule. This is true of the data for individual industries and for all industries taken together. It holds for sizes of plants as well as of firms.[3]

The log-normal function has most often been fitted to the data, and generally fits quite well. It has usually been noticed, however, that the observed frequencies exceed the theoretical in the upper tail, and that the Pareto distribution fits better than the log-normal in that region. This observation suggests that the stochastic mechanism proposed in the previous section is the appropriate one, and that the data should be fitted with the Yule distribution.

We have fitted straight lines to the logarithms of the cumulative distributions for the British data of Hart and Prais, and for the data on large American firms in 1955 published in *Fortune* [5], obtaining

[2] They need not be new-born, merely small. That is, we may assume some arbitrary lower size limit and regard any firm that reaches this size as "new-born." In this case the equilibrium distribution will hold only for firms above the minimum.

[3] It is the ubiquitousness of these functions in size distributions of firms, as well as in distributions of wealth, incomes, city populations and a host of other, more or less unrelated, phenomena that argues most persuasively for their common base in some kind of weak probabilistic hypothesis.

good fits in both cases.[4] In the British case, we get $\rho = 1.11$, in the American $\rho = 1.23$. On the basis of these parameters, we would infer that a little less than one-fifth (18.7 per cent) of the growth in assets of the American firms was accounted for by new firms, and about one-tenth (9.9 per cent) in the British case.

It is not necessary, of course, to make indirect inferences of this sort from the steady-state distributions. Data are now available, both in Britain and the United States, that allow us to follow the changes in size of individual firms, and to construct the transition matrices from one time period to another. Hart and Prais have published such transition matrices for British business units for the periods 1885-96, 1896-1907, 1907-24, 1924-39, and 1939-50 [7, Tables 3, 4, 5, 6, 7]. From the matrices, they have been able to test directly the first assumption underlying the stochastic processes we are considering—the law of proportionate effect. They found that the frequency distributions of percentage changes in size of small, medium, and large firms, respectively, were quite similar—approximating to normal distributions with the same means and standard deviations. We found the same to be the case with the transition matrix for the 500 largest U. S. industrial corporations from 1954 to 1955 and 1954 to 1956.

A simple, direct way to test the law of proportionate effect is to construct on a logarithmic scale the scatter diagram of firm sizes for the beginning and end of the time interval in question. If the regression line has a slope of 45 degrees and if the plot is homoscedastic, the law of proportionate effect holds and the first assumption underlying the stochastic models holds. A plot of the U. S. data shows these conditions to be well satisfied for the 1955-56 period.

In addition, as an independent check of our parameter ρ, we calculated for the American firms for the years 1954-56 the quantities G (net growth in assets—for all firms above the $200 million category) and g (the part of this growth due to new firms—those entering the $200 million group). The figure obtained for g/G was 21.2 per cent which yields a ρ of 1.27. These may be compared with the respective indirect estimates, 18.7 per cent and $\rho = 1.23$, above. Thus we have obtained a close correspondence between the parameter obtained by fitting a steady-state distribution and that obtained by studying the growth of firms over time.

Thus far our data have encompassed an entire economy rather than a single industry. We justify applying the process to the whole economy on several grounds. First, the stochastic growth model we have de-

[4] In the absence of better developed theories about goodness of fit of these skew distributions than we now have, we prefer not to make definite statements about "how good" the fits are.

scribed makes no reference to any feature of the cost curve, other than that costs are constant above some minimum point. Nothing in the model requires the firms in the sample to have the same cost curves. Second, if firms in various industries are distributed according to the Pareto curve with slopes close to 1 in each case, the composite curve for all industries will be a Pareto curve with slope close to 1. For these reasons, the arbitrariness of industry classification, and the heterogeneity of firms within industries do not create the same difficulties in applying the present theory as in applying classical cost theory to explain size distributions.

As an example of a distribution for a single industry, and because it represents an intrinsically interesting case in view of recent discussions

TABLE 1.—INGOT CAPACITIES OF TEN LEADING STEEL PRODUCERS
(Millions of Net Tons per Year, based on Capacity as of January 1, 1954)

Producer	Capacity	
	Actual[a]	Estimated
U. S. Steel	38.7	34.3
Bethlehem	18.5	17.1
Republic	10.3	11.3
Jones & Laughlin	6.2	8.5
National	6.0	6.8
Youngstown	5.5	5.2
Armco	4.9	4.8
Inland	4.7	4.2
Colorado Fuel & Iron	2.5	3.8
Wheeling	2.1	3.4
Total, 10 Companies	99.4	99.4

[a] *Source:* Actual from *Iron Age,* January 5, 1956, p. 289.

of mergers, we present in Table 1 a comparison of the actual ingot capacities of the ten leading steel producers with the theoretical capacities computed from the Yule distribution, with ρ taken at its limiting value, 1.

Perhaps the most interesting question for single industries is whether we can find any evidence of the minimum economic scale, S_m, from the size distributions. However, it is much more difficult to establish the minimum feasible size of firm than to establish the minimum feasible size of plant. The latter can often be estimated reasonably from engineering design considerations, and Bain found in most industries some consensus about this minimum scale for an efficient plant [2, Appendix B]. There is much less basis for estimating, and much less consensus about, the minimum scale for an efficient firm.

TABLE 2.—ESTIMATE OF MINIMUM FEASIBLE PLANT SIZE[a]

Industry	Bain Estimate as Per Cent of National Market	Estimate from Census Data by Yule Distribution as Per Cent of Total Value Added by Manufacture
Flour and milling	0.05 to 0.25	0.07 to 0.19
Footwear	no minimum	0.03 to 0.07
Canned fruits and vegetables	no minimum	0.06 to 0.11
Cement	0.4 to 0.7	0.14 to 0.54
Distilled liquors (except Brandy)	0.2 to 0.3	0.03 to 0.11
Petroleum refining	0.4 to 0.9	0.12 to 0.34
Meat packing	no minimum	0.3 to 0.7
Rubber tires and tubes	0.35 to 0.7	1.6 to 5.5
Rayon	1.0 to 3.0	0.14 to 0.37
Soap and glycerin	0.2 to 0.3	0.03 to 0.11
Cigarettes	1.0 or less	0.08 to 2.0
Fountain pens and mechanical pencils	1.3 to 2.5	0.06 to 0.16
Typewriters	5.0	5.7 to 14.1

[a] Minimum feasible plant size is that below which costs per unit rise substantially. The industries listed are those used by Bain, with seven omitted because of the inadequacy or incomparability of the data.

Sources: The Bain estimates were computed by multiplying his estimates of minimum efficient plant size [2, Table III, p. 72] by the fraction of their size that was encountered before costs rose substantially [2, Appendix B].

The estimates from Census Data were computed by plotting the cumulative number of firms from the 1947 *Census of Manfuacturers* against size in number of employees for the industries listed. Sharp breaks in the cumulative plot from a slope of -1 were taken as estimation points for the minimum feasible plant size and were converted to a percentage of total value added by manufacture from the same Census tables.

Taking Bain's estimates of the minimum efficient plant size, on the one hand, and Census of Manufacturers data on the size distribution of plants, on the other, we have made some preliminary attempts to compare for several industries the minimum efficient scales suggested by these two sets of data. The results are listed in Table 2. Our procedure was this: If there is a sharp increase in unit costs below some critical size, P_m, the number of plants in the industry below that size should be less than the number predicted from the Yule process. We plot cumulative numbers of plants against size on log paper, and look for sharp bends from a slope approximating -1 to a lower slope.

We have used census data for numbers of employees and converted these to per cent of total value added by manufacture. Our measure is thus comparable to Bain's which is based upon the percentage that a plant represents of total national market.

The reader can draw his own conclusions as to how far the two estimating procedures lead to similar results. Since we have made no more than preliminary explorations, we do not wish to push the point too hard. It is clear, however, that the stochastic model provides some novel ways of interpreting the data on size distributions that may cast considerable light on the question of economies of scale. The argument runs as follows: If we take the stochastic model seriously, then any substantial deviation of the results from those predicted from the model is a reflection of some departure from the law of proportionate effect or from one of the other assumptions of the model. Having observed such a departure, we can then try to provide for it a reasonable economic interpretation.

In concluding this discussion of the data, we should like to emphasize a point made earlier—that the transition matrices may provide an even more valuable source of data about the process determining the sizes of firms and plants than the size distributions themselves. Since most of the empirical work to date has focused on the latter rather than the former, the reversal of emphasis initiated by the work of Champernowne, Hart and Prais, and others, is a very promising one.

IV. *Implications for Economic Policy*

In discussions of the degree of competition in individual industries, various measures of degree of concentration have been used. Few of these have other than an empirical basis, and the values that are obtained depend, in ways that are only partly understood, on methods of classification, cut-off points, and the like. Among the frequently used measures are Lorenz's and Gini's coefficients of concentration.

As Aitchison and Brown [1, pp. 111-16] argue, if we fit a distribution function to the observed data on the basis of a theoretical model, it is reasonable to base our measures of concentration on the parameters of the distribution function. Thus, they propose the standard deviation of the log-normal as an appropriate measure of dispersion, and show that the Lorenz and Gini coefficients can be expressed as functions of that statistic.

Similarly, if we use the Yule process to account for the distribution of firm sizes, our interpretation of the observed phenomena should be based on the estimated values of the parameters of the distribution. In the simplest case, the only one we have considered here, there is a single parameter, ρ. We have already provided an economic interpretation for this parameter in the previous section—it measures, in a certain sense, the rate of new entry into the industry. Hence, in this particular model, the concentration in an industry is not independently determined, but is a function of rate of new entry.

We may put the matter more generally. If firm sizes are determined by a stochastic process, then the appropriate way to think about public policy in this area is to consider the means by which the stochastic process can be altered, and the consequences of employing these means. As a very simple example, if the rate of entry into the industry can be increased, this will automatically reduce the degree of concentration, as measured by the usual indices. Similarly, if, through tax policies or other means, a situation of sharply increasing costs is created in an industry, this situation should cause a departure of the equilibrium distribution from the Yule distribution in the direction of lower concentration.

A third, and more complicated, example is this: the amount of "mixing" that takes place—reordering of the ranks of firms in an industry—depends on the dispersion of the columns of the transition matrix. The same equilibrium distribution may be produced with various degrees of mixing, since the latter can vary independently of the law of proportionate effect. Public policy might be concerned with the amount of mobility rather than with the resulting degree of concentration. As a matter of fact, a measure of mobility (for firms or individuals) would appear to provide a better index of what we mean by "equality of opportunity" than do the usual measures of concentration.

The net effect of approaching the subject of industrial concentration in this way will be to make the classical theory of the firm much less relevant to the subject, but theories of economic development and growth much more relevant. When we have a collection of adaptive organisms placed in a relatively stable environment, we can often make strong predictions about the resulting state of affairs by assuming that the system will come into a position of stable, adaptive equilibrium. When, however, the environment itself is changing at a rate that is large compared with the adaptive speeds of the organisms, we can never expect to observe the system in the neighborhood of equilibrium, and we must invoke some substitute for the static equilibrium if we wish to predict behavior. Our main objective in this paper is to suggest the need for, and the availability of such a substitute with which to analyze the size distribution of firms.

V. *Directions for Research*

We have emphasized the tentative character of our results, and should like to suggest in conclusion some directions of research that look exceedingly promising:

1. We need to accumulate a body of knowledge about skew distribution functions and the processes that generate them that is comparable to the rich knowledge we possess about the normal, Poisson, exponential, and related distributions. We need to know more about the relations between the distributions and the generating processes, about efficient methods for estimating parameters, about the distributions of these

estimates, and about efficient methods for choosing among alternative hypotheses.

2. We need to develop stochastic models of economic growth that embody as much knowledge as we have, or can acquire, about the underlying processes.

3. We need to re-examine the corpus of economic data to see what part of it can profitably be explained or reinterpreted in terms of such economic models.

4. We need to re-examine those principles of public policy that are based on static equilibrium analysis to see what part of them will remain and what part will be altered as stochastic processes begin to play a larger role in our explanation of economic phenomena.

In this paper we have tried to suggest some of the directions in which inquiry may lead if it is guided by questions such as these.

REFERENCES

1. J. AITCHISON and J. A. C. BROWN, *The Lognormal Distribution*, Cambridge 1957.
2. J. S. BAIN, *Barriers to New Competition*, Cambridge, Mass. 1956.
3. D. G. CHAMPERNOWNE, "A Model of Income Distribution," *Econ. Jour.*, June 1953, *63*, 318-51.
4. W. FELLER, *An Introduction to Probability Theory and Its Application.* New York 1950, Vol. 1.
5. "The Fortune Directory of the 500 Largest U. S. Industrial Corporations," *Fortune*, July 1956, *55*, Suppl.
6. P. E. HART and E. H. PHELPS BROWN, "The Sizes of Trade Unions: A Study in the Laws of Aggregation," *Econ. Jour.*, Mar. 1957, *67*, 1-15.
7. P. E. HART and S. J. PRAIS, "The Analysis of Business Concentration," *Jour. Royal Stat. Soc.*, Pt. 2, 1956, *A. 119*, 150-91.
8. JOAN ROBINSON, *The Economics of Imperfect Competition.* London 1942.
9. H. A. SIMON, "On A Class of Skew Distribution Functions," *Biometrika*, 1955, *52*, 425-40.
10. J. VINER, "Cost Curves and Supply Curves" from *Zeitschrift f. Nationalökon.*, 1931, reprinted in AEA *Readings in Price Theory*, pp. 198-232, Vol. VI, edited by G. J. Stigler and K. E. Boulding, Chicago 1952.
11. H. O. A. WOLD and P. WHITTLE, "A Model Explaining the Pareto Law of Wealth Distribution," *Econometrica*, Oct. 1957, *25*, 591-95.