WORKING PAPERS

THE DIVERSITY OF FIRM SIZE DISTRIBUTIONS IN MANUFACTURING INDUSTRIES

John E. Kwoka, Jr.

WORKING PAPER NO. 12

February 1978

BUREAU OF ECONOMICS
FEDERAL TRADE COMMISSION
WASHINGTON, DC 20580

The size distribution of firms in manufacturing industries is a matter of considerable importance for economics and public policy. Interest in it, however, has been stymied by the lack of sufficiently detailed data and by the scarcity of insights offered by conventional economic theory. In the face of these problems, explanations for the typical skewed distribution of firm sizes have been defaulted to stochastic growth theories, but they in turn are subject to a number of reservations which limit their appeal. In this paper we shall introduce a new data source on firm sizes and describe the distributions there found. The detailed information on market shares in this data set suggests a richness to firm size distributions not easily summarized, and rather clearly not captured by the log-normal or Pareto distributions.

We begin with a critical review of economic theory and empirical work on firm size distribution. Then the construction of the new data set is described. The remainder of the paper is devoted to an intensive examination of the actual distributions for the industries in the data set, utilizing a variety of tools to demonstrate their diverse nature.

I

Static economic theory has few clear implications for market share distribution. For example, if long-run costs are constant, it would appear that all firm sizes are equally likely. At the other extreme, with a U-shaped cost curve in a competitive industry, only one size of firm can persist. But if the shares are substantial enough, market power may permit above-competitive price to be set and to protect non-optimal plant or firm sizes. There is little indication, however, as to the resulting distribution. Similar ambiguity exists if the long-run cost configuration is L-shaped, i.e., with a minimum efficient size but no persistent scale economies past that point. Finally, in consumer goods industries, early entry may make a leading firm's position difficult to overtake and may produce a pattern of declining market shares.

In contrast to these slender theoretical insights, available data have repeatedly shown a size distribution of firms that is highly skewed [1]. Indeed, the resemblence to the log-normal distribution is quite striking, with a few large firms and a long, thin tail of small and nearly-equal-size firms. The prevalence of this pattern has given rise to the search for a general explanation.

The vacuum left by conventional theory has been at least partially filled by theories of stochastic growth processes [10; 11]. What most have in common is the so-called Gibrat's Law, that the percentage change in a firm's size in a given period is independent of its initial size. With a given population of firms, it is readily shown that over time such a growth process produces a log-normal size distribution of firms. If, on the other hand, a constant rate of birth (entry) of new firms into the smallest size class is assumed, a Yule distribution results. The upper tail (largest size classes) of the latter can be shown to resemble a third distribution, the Pareto, which is often used to describe that portion of actual firm size distributions [9].

Yet stochastic growth processes do not command wide support as explanations for the size distribution of firms. Inconsistencies exist between data and theory, and attempts to reconcile them often appear ad hoc and unconvincing [2; 3; 4]. Inadequate attention has been paid to the distinction between firm size distribution within an industry and for the entire economy. Moreover, tests of goodness of fit of the data to such distributions are not straightforward, and those that have been performed demonstrate that no single distribution is universally applicable [6; 8]. Apart from such specific problems, stochastic growth processes often seem to explain too much, leaving inadequate scope for other widely accepted determinants of market structure.

II

This investigation of the size distribution of firms is based on a new data source which we shall first describe. Market share data by four-digit SIC industry are derived from a tabulation by Economic Information Systems, Inc., of 120,000 manufacturing establishments employing at least twenty persons in 1972. From County Business Patterns data on employment by SIC for each country, EIS allocates employment to plants which have been identified by a mailing to 300,000 establishments.

Shipments for each plant are estimated by multiplying employment data by Census of Manufactures ratios of shipments per employee for various size classes of plants in each four-digit SIC industry. The result is fairly comprehensive listing of plants and associated data which, despite some inevitable errors and biases, appears quite reliable. 1/

The Bureau of Economics of the Federal Trade Commission has constructed an alternative version of this basic EIS data set, one more suited to industry studies. This version organizes the data to show total shipments and market share of each company operating in each four-digit industry in which it has at least one plant with 20 or more employees. The present author has created another data set linking 1972 Census of Manufactures data to this industry - share version of the EIS compilation. Several factors reduce the number of industries in the final data set to 314, out of 451 original Census four-digit industries. Seventy of the original number were not industries at all, but catch-all categories labelled "miscellaneous" or "not elsewhere classified." They were eliminated, as were 43 others which could not be matched due to the major revision of SIC definitions used in the 1972 Census but not at that time by EIS. Twenty-four additional Census industries were condensed into 11 based on old definitions for which share data were available.

The resulting data set is nonetheless very large and detailed. It has already been employed to conduct an intensive examination of structural features which influence the magnitude of price-cost margins in such industries [5]. Our present focus on a few leading firms is consistent with the results of that study, but also is intrinsically the most interesting part of the distribution of firm sizes.

III

We begin the analysis by reporting some characteristics of the top ten market shares in the 314 industries in 1972. Table I shows the mean, maximum, and minimum values, along txth the number of industries (N) which have a firm of that mgje !a EIS count. Thus the average share of the largest firm (S1) in these industries was 17.5 percent, ranging from a low of 1.1 percent to a high of 68.7 percent. The mean values of successive shares S2,...S10 decline in a fairly regular fashion and with sharply diminishing ranges.

A related question of some interest is the degree of correlation among successive shares. As shown in Table II, simple correlation coefficients also exhibit a fairly regular, if somewhat complex, pattern. All coefficients among successive shares are positive and significant, 2/ rising among lower-ranked shares. For example, the correlation between S1 and S2 is .702; between S2 and S3, .780; and between S9 and S10, .944. Correlations between shares farther removed from

Table I

Share	Mean	Minimum	Maximum	N
S1	.175	.011	.687	314
S2	.100	.008	.334	314
S3	.070	.007	.214	314
S4	.053	.006	.169	314
S5	.042	.005	.121	314
S6	.034	.003	.095	314
S7	.028	.001	.060	314
S8	.024	.000	.054	312
S9	.020	.000	.046	311
S10	.018	.000	.043	309

Table II

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	1.000	.702	.597	.437	.330	.211	.076	-.022	-.108	-.166
S2		1.000	.780	.608	.463	.379	.161	.073	-.038	-.112
S3			1.000	.825	.658	.543	.328	.229	.122	.069
S4				1.000	.843	.707	.485	.358	.277	.251
S5					1.000	.851	.666	.558	.497	.437
S6						1.000	.830	.715	.618	.551
S7							1.000	.903	.801	.718
S8								1.000	.900	.820
S9									1.000	.944
S10										1.000

each other in the rank order decline rather sharply. In fact, in the extreme cases of S1 with S8, S9, and S10, and S2 with S9 and S10, they even become negative.

While these correlations reveal some associations, their economic significance is less clear. A correlation between S1 and S2 can result simply from the fact that S2 can take on larger values when S1 is larger. Hence a statistical correlation between the two may result even if the S2 is distributed randomly over its possible range, the latter generally between zero and S1. The discovery of such possibilities is indeed one purpose of this paper.

Table III casts some light on this issue. The observed range of S1 is broken into six groups, five deciles and a residual category where S1 exceeds 50 percent. Industries are sorted into these categories, and mean values for the largest three shares are given in each. In addition, the ratios of S2 to S1, and of S3 to S2 are calculated and denoted as P21 and P32 respectively. 3/ For descriptive purposes below, mean value added (VA), mean number of companies (NCO), and the total number of observations within each S1 category are also provided.

These data reveal some noteworthy patterns. As the largest share rises, so does the second share, but considerably more slowly. Thus while S1 increases from .063 to .435, S2 grows only from .048 to .167 and in fact is nearly constant

Table III

S1

	0-10	10-20	20-30	30-40	40-50	50+
S1	.063	.143	.243	.350	.435	.564
S2	.048	.092	.152	.166	.167	.172
S3	.040	.068	.096	.106	.121	.063
P21	.781	.651	.634	.479	.386	.405
P32	.834	.768	.663	.682	.752	.540
VA	825	827	761	628	460	3215
NCO	1273	412	166	95	91	155
N	109	93	62	32	13	5

from the 20%-30% decile for S1. As a fraction of its maximum possible value, the second share declines throughout, from .781 to .386. Clearly S2 is not randomly distributed, but instead the top two firms are more disparate in size when the leading share is greater.

With respect to the third leading share, a similar but weaker phenomenon occurs. S3 increases with S1, but as a fraction of its largest possible value (generally S2) it grows somewhat more slowly. Its relative decline is neither as steep nor as regular as that of S2. Nevertheless, we are likely to find relatively large third shares where S1 is small.

The characteristics of these industries by size classes are of some interest as well. Where S1 is larger, the average value added of the industry is smaller, as is the number of member firms. That is, large leading firms are generally found in small industries and are accompanied by fewer total firms. The exception to this pattern is the extreme category of "50 percent and up" for S1, which has large VADD and NCO due to two outlying industries (Motor Vehicles and Electrical Engines). It is also interesting to note that while a majority of industries have largest shares of less than 20 percent, 112 exceed that cut-off. Clearly a large number of industries are led by sizeable firms.

A more detailed examination of these phenomena is permitted by the data in Table IV. Here not only is S1 stratified by ten

Table IV

P21

S1	0-20	20-40	40-60	60-80	80-100	
0-10	0	1	14	39	55	109
10-20	1	13	25	25	29	93
20-30	1	12	17	13	19	62
30-40	2	10	12	5	3	32
40-50	1	6	5	1	0	13
50+	2	0	2	1	0	5
	7	42	75	84	106	314

percentage point intervals (plus the "50 percent and up" category), but P21 is broken into quintiles. It is evident that P21 varies quite systematically with S1. Industries with small leading firms have relatively large S2 most of the time. As S1 increases (lower rows), the entire distribution of P21 changes shape and position, becoming flatter and shifting toward smaller values. The decline in P21 confirms our previous observation that disparity increases as S1 grows, but the flattening of the distribution suggests a new and striking conclusion: There is an immense range to the values S2 can take on, even given S1. We shall now explore additional data to see if this diversity persists with respect to S3.

Tables V-A through V-F display the scatter of P21 and P32 by categories of S1. Thus Table V-A consists of those 109 industries for which S1 $\leq$ .10, and shows the frequency distribution of the ratios of S2 to S1 and S3 to S2. We note a rather high concentration of P32 toward large values in Table V-A, implying near equality of S3 to S2. Examination of the last columns of Table V-A through V-F shows changes in the shape and position of the distribution of P32, somewhat analogous to those of P21. That is, as S1 increases, S3 declines relative to S2 on average and has a wider range of observed outcomes. When S1 $\leq$ .10, about three-fourths of all P32 exceeded 80% regardless of the value of P21, whereas when S1 > .50, the distribution appears quite flat.

Table V

A

S1 from 0 to .10

P21

P32	0-20	20-40	40-60	60-80	80-100	
0-20	0	0	0	0	0	0
20-40	0	0	0	1	0	1
40-60	0	0	1	6	4	11
60-80	0	0	3	5	15	23
80-100	0	1	10	27	36	74
	0	1	14	39	55	109

B

S1 from .10 to .20

P21

P32	0-20	20-40	40-60	60-80	80-100	
0-20	0	0	0	0	0	0
20-40	0	0	1	1	5	7
40-60	0	0	1	4	7	12
60-80	1	2	5	8	5	21
80-100	0	11	18	12	12	53
	1	13	25	25	29	93

Table V (Continued)

C

S1 from .20 to .30

P32	P21 0-20	20-40	40-60	60-80	80-100	
0-20	0	0	0	0	0	0
20-40	0	0	1	4	2	7
40-60	0	2	1	3	9	15
60-80	0	7	8	4	6	25
80-100	1	3	7	2	2	15
	1	12	17	13	19	62

D

S1 from .30 to .40

P32	P21 0-20	20-40	40-60	60-80	80-100	
0-20	0	0	0	0	0	0
20-40	0	0	0	2	0	2
40-60	0	1	5	5	0	11
60-80	0	4	4	1	0	9
80-100	2	5	3	0	0	10
	2	10	12	8	0	32

Table V (Continued)

E

S1 from .40 to .50

P21

P32	0-20	20-40	40-60	60-80	80-100	
0-20	0	0	0	0	0	0
20-40	1	0	0	1	0	2
40-60	0	0	1	0	0	1
60-80	0	2	3	0	0	5
80-100	0	4	1	0	0	5
	1	6	5	1	0	13

F

S1 greater than .50

P21

P32	0-20	20-40	40-60	60-80	80-100	
0-20	0	0	0	1	0	1
20-40	0	0	1	0	0	1
40-60	0	0	0	0	0	0
60-80	1	0	1	0	0	2
80-100	1	0	0	0	0	1
	2	0	2	1	0	5

Quite apart from these general tendencies is the diversity revealed by these several tables. When S1 is greater, S2 and S3 fall in relative terms but often continue to span the range of possible values. The data display wide scatter, not tight fits. It is, for example, entirely possible when S1 = .15 for P21 = .70 and P32 = .30, or for P21 = .10 and P32 = .70. And combinations of P21 = .70 with P32 = .30 occur both when S1 = .15 and when S1 = .45. Such diversity is hard to reconcile with the rigid nature of the distributions implied by stochastic growth models. We can cast some light -- and perhaps sufficient doubt -- on their general applicability in the following manner. 4/

The upper tail of firm size distribution is thought to be best captured by the Pareto distribution:

$$SR^{B} = A \quad (1)$$

where S is market share, R denotes rank (1 being the largest firm), A a constant, and B a parameter unique to the industry. Definitions, substitutions, and manipulation yield the following condition:

$$P32 = (P21)^{.585} \quad (2a)$$

or

$$\log P32 = .585 \log P21 \quad (2b)$$

Note that this condition is independent of the particular S1, A, and B for the industry, and hence our data in this form are easily compared and interpreted.

Equation (2) is plotted in Figure 1. A comparison of the data in Table V with this relationship may be made by noting that P21 from zero to .20 is wholly contained in the first two quintiles of P32, that P21 from .20 to .40 is largely within P32 from .40 to .60, and so forth. The boxes drawn in Figure 1 represent the test categories. A simple count of those "IN" the boxes and those "OUT" of the boxes, both in total and by categories of S1, appears in Table VI. Only 38% of the total number of observations (industries) are correctly classified by this characterization of the Pareto distribution, and most of those occur in a single size class of S1 (the smallest, where $S1 \leq .10$). The preponderance of OUT observation in Table V lies above the line in Figure 1, suggesting not merely wide variation around the line, but an incorrect placement of the line itself. The data do not suggest the Pareto distribution as a general form to these distributions.

Further evidence to this effect is provided by regression analysis on Equation (2b). With all 314 industries as the data base, ordinary least squares yields the following results:

$$\log P32 = \underset{(.036)}{.347} \log P21 \qquad \bar{R}^2 = .23 \qquad (3)$$

The low $\bar{R}^2$ reflects the rather loose fit of these data to the Pareto specification, a finding consistent with our previous observation. In addition, the point estimate of the coefficient, .347, is very different from the hypothesized $\alpha = .585$. The standard error of .036 implies a 5% confidence interval of

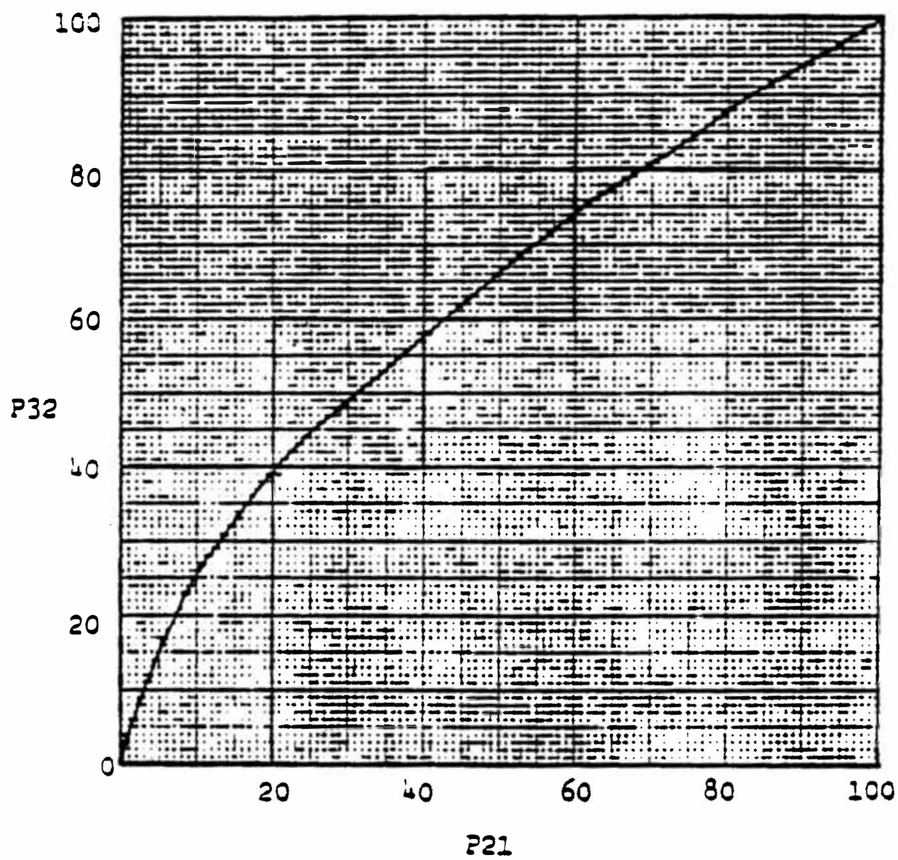

FIGURE 1

Table VI

SI	IN	OUT	TOTAL
0 - .10	66	43	109
.10 - .20	29	64	93
.20 - .30	14	48	62
.30 - .40	5	27	32
.40 - .50	4	9	13
50+	1	4	5
TOTAL	119	195	314

$.276 \leq \hat{a} \leq .417$, and we are led to reject the contention that the data fit the Pareto distribution. 5/

IV

The new data source described in this paper provides much more detailed information about firm sizes in various industries than heretofore available. Simple examination of these data reveals the enormous diversity of the size distributions of firms, and makes clear the reasons why attempts to impose a single distribution or summary statistic are unlikely to succeed. While doubts about the applicability of the log-normal and Pareto distributions appear well-founded, no clear alternative to stochastic growth theories has been advanced. This paper has affirmed the place of that issue on the agenda of industrial organization economics.

Footnotes

1/ For example, the elimination of plants with fewer than twenty employees and the estimation of shipments by Census productivity ratios introduce various kinds of errors. None have been found in careful tests to seriously damage the data. [5].

2/ With 314 observations, the simple correlation coefficient is significant at the .05 level at a value of .11.

3/ For those five industries where S1 exceeds 50 percent, the maximum possible S2 is (1-S1) rather than S1 itself. Hence P21 is smaller than S2 is as a fraction of its largest possible value. We present P21 (and similarly P32) for comparability and later regression work, but restrict our present discussion to the 309 industries for which S1 < .50.

4/ Precise statistical testing with (e.g.) chi-square is virtually precluded by the low frequencies in many cells and somewhat arbitrary classification. For further discussion, see [6, pp. 423-4] and [7,pp. 46-47].

5/ The use of linear regression across industry shares raises questions about the nature and properties of the error term, e.g., in Equation (3). Since we cannot hope to resolve these issues -- certainly not in this paper -- we raise this problem only as a caveat to the present results.

References

1. Hart, P. and S. Prais, "The Analysis of Business Concentration," Journal of the Royal Statistical Society, Part 2, 1956, pp. 150-91.

2. Ijiri, Y. and H. A. Simon, "Business Firm Growth and Size," AER, March 1964, pp. 77-89.

3. ____________, "Effects of Mergers and Acquisitions on Business Firm Concentration," JPE, March-April 1971, pp. 314-22.

4. ____________, "Interpretations of Departures from the Pareto Curve Firm-Size Distributions," JPE, pp. 317-331.

5. Kwoka, J., "The Effects of Market Shares and Share Distribution on Industry Performance," FTC Bureau of Economics Working Paper No. 2, March 1977.

6. Quandt, R., "On the Size Distribution of Firms," AER, June 1966, pp. 416-32.

7. Siegel, Sidney, Non-parametric Statistics, New York: McGraw-Hill, 1956.

8. Silberman, Irwin, "On Lognormalility as a Summary Measure of Concentration," AER.

9. Simon, H., "On a Class of Skew Distribution Functions," Biometrika, 1955, pp. 425-40.

10. Simon, H. and C. P. Bonini, "The Size Distribution of Business Firms," AER, Sept. 1958, pp. 607-17.

11. Steindl, J., Random Process and the Growth of Firms, New York: Hafner, 1965.