Size Distribution of Industrial Firms: Connection between Zipf's Law and Gibrat's Law

I. Introduction

Physicists and economists have held contradictory beliefs regarding the size (sales) distribution of industrial firms. Physicists have tended to utilize Zipf's Law, which predicts a power-law function when the magnitude of items of a set are plotted against the rank order of those items (for example, the sales of corporations). Gell-Mann in *The Quark and the Jaguar* clearly enunciates the physicists view when he states, Is there an approximate rule that describes how the sales figures of the firms vary with their rank? Yes and it is the same rule as for populations – Zipf's Law. However, economists base their beliefs regarding the size distribution of industrial firms upon Gibrats Law, which states that a histogram giving the number of companies with a given volume of sales fits a log-normal distribution. We shall show in Section IV that these two views contradict one another.

We resolve this contradiction by showing tht neither law is correct. Specifically, we find that while Gibrat's Law holds for relatively small corporations (and therefore the majority of corporations), it fails to explain the size distribution of the largest corporations. Conversely, Zipf's Law appears to accurately describe the largest corporations, but fails to predict the size distribution of smaller corporations. In the process of resolving this contradiction, we find that there appears to be a threshold which marks the boundary between the larger corporations which seem to exist under different economic laws than the corporations on the other side of this threshold.

II. Zipf's Law

Zipf's Law, developed by George Kingsley Zipf, states that a plot of the number of occurrences of a given word versus the rank order of the words results in a power-law decay. This law has been adapted to occurrence of certain words formed by strings of base-pairs in non-coding DNA, as well as the population distribution of cities.

To implement of Zipf's La, we require a list of corporations and their sales. CompuStat CD-ROM database gives the sales of every publicly traded United States corporation for each year between 1975 and 1994. We list the corporations in rank order by sales and plot the result on log-log paper. This procedure results in the familiar power-law decay characteristic of Zipf's Law for approximately the 2 decades (Fig. 1). Further proof to the accuracy of Zipf's Law for the first 2 decades of the plot is shown in Fig. 2, which is a plot of the value of the average of the successive slopes of logarithmicly binned (by the square root of 2) corporations in rank order over the 20 year period under investigation (between 1975 and 1994) versus the logarithmic bining. We note the plateau at $\zeta \approx 0.82$between rank 5 and rank 105, which indicates that the slope of the log-log plot in Fig. 1 does not change over this interval. Furthermore, every year, the straightness of the power-law and the its slope remain approximately constant. as is evidenced in Fig. 3 and Table 1. Fig. 3 shows the successive slopes with the successive slopes measured for all 20 years in the interval from 1975 to 1994. Table 1 is a table gives the slope over rank 5 to rank 106 of the plot for each year between 1975 and 1994.

Figure 4 depicts the log-log plot for all the corporations in 1994. This figure clearly shows the rapid rate of decay after approximately rank 100. We are not aware of any other attempts by the physics or mathematics community to explain (or even claim knowledge of) the rapid decay existing in the third decade of the log-log plot.

III. Gibrat's Law

Gibrat's Law, formulated by the French mathematician R. Gibrat, states that the size distribution of business firms fits a log-normal distribution. Intuitively, one might guess this result given the fact that the increment in sales from one year to the next appears to be a multiplicative function. Using the CompuStat data, we confirmed Gibrats Law as the proper result for the majority of corporations. We begin by logarithmicly bining the sales of the corporations in a year in log bins of size square root of 2. Then, we counted the number of corporations in each bin and plotted the number of corporations in that bin as a function of the bin number. The result was the log-normal distribution seen in figure 5.

IV. Resolving the Contradiction

We next show that the Zipf plot for a system obeying Gibrat's law is a parabola, not a straight line. We define the parameters of Gibrat law

$$S_0 = \frac{1}{N}\sum_{i=1}^{N}\ln(S_i)$$
$$D = \frac{1}{N}\sum_{i=1}^{N}[\ln(S_i)]^2 - S_0^2,$$

where N is total number of companies and S_i is sales of each company. The Gibrat law states that the probability to find a company with $\ln(S_i)$ between x and $x + dx$ is

$$P(x)dx = dx \cdot \frac{1}{\sqrt{2\pi D}}\exp[-(x - S_0)^2/2D].$$

The rank of the company with sales S can be defined as

$$r(s) = N\int_{\ln S}^{\infty}\rho(x)dx = \frac{N}{\sqrt{\pi}}\int_{A}^{\infty} dx e^{-x^2} = N\ erfc(A),$$

where

$$A \equiv \frac{\ln S - S_0}{\sqrt{2D}}.$$

For large arguments,

$$erfc(y) \approx \frac{1}{2y\sqrt{\pi}} \cdot e^{-y^2}.$$

In order to find the slope of the Zipf law, one has to computer $\ln S$ as a function of $\ln r$. Thus

$$\ln r - \ln N - \frac{(\ln S - S_0)^2}{2D} - \ln\left(\frac{2\sqrt{\pi}}{\sqrt{2D}}(\ln S - S_0)\right).$$

Neglecting the last term, which is small compared to the first and second, we have

$$\ln r = \ln N - \frac{(\ln S - S_0)^2}{2D},$$

or

$$\ln S = S_0 + \sqrt{(\ln N - \ln r) \cdot 2D}.$$

This means that the Zipf graph for the Gibrat law is not straight, but can be approximated by a parabola.

Before concluding, we note that the slope of a Zipf graph is

$$\frac{d \ln S}{d \ln Z} = -\sqrt{\frac{D}{2(\ln N - \ln Z)}},$$

and the initial slope is

$$\sqrt{\frac{D}{2 \ln N}}.$$

A more precise way of deriving this formula is to differentiate $\ln r$ over $\ln S$ as an inverse function

$$\begin{aligned} \frac{d \ln r}{d \ln S} &= \frac{dr/d \ln S}{r} \\ &= -\frac{(1/\sqrt{2D})e^{-A^2}}{\int_A^\infty e^{-x^2} dx} \\ &\approx \frac{(\ln S - S_0)}{D}. \end{aligned}$$

Then

$$\frac{d \ln S}{d \ln r} = -\frac{D}{\ln S - S_0} \approx -\frac{D}{\sqrt{(\ln N - \ln r)2D}} = -\sqrt{\frac{D}{2(\ln N - \ln r)}}$$

as before.

A good estimation of the initial is thus

$$\frac{D}{\ln(S_{\max}) - S_0},$$

where $S_{\max}$ is the sales of the maximal company.

However, this contradiction can be resolved by the realization that Gibrat's Law does indeed hold for the relatively smaller and middle-sized corporations. However, it does not sufficiently explain the size of the largest corporations. Among these corporations, Zipf's Law provides the best accuracy for describing their size distribution. Therefore, we conclude that a different set of laws appears to hold for corporations below a certain threshold, approximately those with sales less than...

Corporations above this threshold appear to exhibit properties similar to those of cities or language in which there exists a high degree of correlations, order and interdependence. In contrast corporations below this threshold appear to exhibit properties more similar to XXX, in which size is relatively randomly placed and almost entirely free of correlations and inter-dependence.

V. Summary

In this paper, we have highlighted a contradiction between the views of the economics community and those of the physics community. Furthermore, we have resolved this contradiction and in so doing, established that the laws of economics appear to change drastically with the size of the corporations.

VI. Acknowledgments

We would like to thank Dr. G. Lowry, Dr. J. Miron and Dr. M. Schary for their assistance in the early stages of this work.