Firm Size Distribution in *Fortune Global 500*

Qinghua Chen, Liujun Chen, and Kai Liu

Department of Systems Science,
School of Management, Beijing Normal University,
Beijing, 100875, People's Republic of China
qinghuachen@bnu.edu.cn

Abstract. By analyzing the data of Fortune Global 500 firms from 1996 to 2008, we found that their ranks and revenues always obey the same distribution, which implies that worldwide firm structure has been stable for a long time. The fitting results show that simple Zipf distribution is not an ideal model for global firms, while SCL, FSS have better fitting goodness, and lognormal fitting is the best. And then, we proposed a simple explanation.

Keywords: worldwide firm size distribution, Fortune Global 500, lognormal distribution.

1 Introduction

Firms or enterprises play key roles in world economy by providing various products and services. Many indexes including revenue, gross asset, numbers employed, and profit are usually used to present their economic status and strength. Those indexes are regarded as *firm size* generally, and they vary enormously, from sole traders up to huge corporations employing tens of thousands of workers. However, there are many evidences show that firm size distributions in different economies are likely the same – they are all highly skew, which means that small firms are very common whereas huge ones are extremely rare [1]. What character do they have? Does this collective character derive from independent individual firms? These issues have attracted lots of attentions in the recent hundred years. Many attempts have been done, among which statistical analysis plays a very important role [2]. Although the economic evolution exhibits complexity, by displaying the statistical characteristics of microeconomic individuals, some stable macroscopic distributions are exhibited, which make clear that there are certain basic rules in complex economy.

In 1931, R. Gibrat, who's the pioneer studying firm size distribution, found that firm size could be described by lognormal distribution [3]. The Gibrat's Law (or law of proportionate effect) was proposed simultaneously to depict such lognormal character. Later in 1949, G. K. Zipf found that distribution of the U.S. corporations' assets was approximately subjected to a power-law, or Pareto-Zipf law [4]. Since then, lots of empirical researches have been conducted on firm size distribution.

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1774–1782, 2009.

There are many empirical works continuously coming forth for a long time. By studying British firms, Hart and Prais (1956) found evidence which shows that lognormal fits fairly well the firm size distribution, while the growth rate of certain firm was independent with its size [5]. And H. A. Simon and C. P. Bonini in 1958 concluded that firm size distribution basically obeyed power law [6]. B. H. Hall studied publicly traded firms in the U.S. manufacturing sector from 1972 to 1983, and he found the number of employees in those firms obeyed Gibrat's Law [7]. M. H. R. Stanley and his companions found the truth that "the upper tail of the size distribution of firms is too thin relative to the log normal rather than too fat, as had previously been believed" [8]. For Japanese companies, K. Okuyama (1999) et al. reported a power-law income distribution over more than three decades [9]. J. J. Ramsden and Gy. Kiss-Haypál (2000) analyzed firm size distribution in different countries using annual net revenue as substitute for firm size, and they found that so-called Simple Canonical Law (SCL, a sort of modified Zipf's law) fitted the data well [10]. R. L. Axtell (2001) obtained Zipf's law when firm size was defined as the number of employees or the amount of receipts over the U.S. firms [11]. J. Voit found lognormal firm size distribution in Germany from 1987 to 1997 [12]. More recently, R. Hernández-Pérez and his colleagues adopted three models including SCL, finite size-scaling (FSS), and lognormal to analyze developing as well as developed countries and found that in different cases different models fit the data better [13]. P. Angelini and A. Generale [14], and J. Growiec [15] also have done lots of meaningful researches on firm size distribution in different countries.

It is natural for people to consider one country as a unit, whose citizens as well as their activities are linked by the same cultural and economic elements – language, laws, currency, history, etc. And the works mentioned above basically focus on firm size distribution within the scope of one country.

However, as we know that almost every country is made up of several heterogeneous regions, and each of them has different degree of industrialization and probably unique cultural background. Commercial relationships between different nations are much the same comparing with that within a large nation. Meanwhile, under economic globalization – the integration of economies and markets worldwide through transnational trade, direct foreign investment, migration, and spread of technology, the biggest corporation wherever it is faces a worldwide economy nowadays. So, thinking about firm size distribution throughout the world is meaningful.

In this paper, we will focus on the *Fortune global 500* and try to find which model from above literature review fits these data better. In Section 2 we would introduce the source of the data and explain it briefly. Then we will carry out data fitting in section 3 in detail and a simple explanation in section 4, and finally drive our conclusion in section 5.

2 The Data

The data we used in this paper come from *Fortune Global 500*, which is a ranking of the top 500 corporations worldwide as measured by revenue. This list which includes the top 500 industrial corporations, financial corporations, and service providers is compiled and published annually by Fortune magazine from 1995. Its latest issue is published on July 21, 2008, and we have collected all the thirteen lists from 1996 to 2008.

From these data, we could find easily the truth that the list has been fluctuating all the time. Without considering change of the firm's name, there were about 400 firms in one list retaining in the list of the next years', and only 111 firms showed up in all the thirteen lists. Moreover, if a firm showed up in both of two adjacent lists, most likely the two ranks would be different. Even the list of countries and districts showed up is varying – the number of countries or districts having at least one firm in the list trends up, from 25 in 1996 to 33 in 2008. Particularly, Poland, Singapore, Thailand, Saudi Arabia, Portugal, Ireland, and Austria became a new member of the list in more recent years.

The number of firms within certain country or district varies too. As showed by Fig. 1, the number of Chinese firms in the list have increased stably from 2 to 29, while of Japanese firms dropped rapidly from 141 to 64 in the same period. The number of U.S. firms increased firstly and then decreased, yet France presents strong stability, holding about 40 firms all the time. There is some stability character for all the years, U.S. holds more than 30% and five developed countries, including U.S, Japan, France, Germany and U.K., take the absolutely big share.

Then we concentrate on each list separately. In each list, there are big gaps between the numbers of different countries or districts. For example, in the newest list the U.S. holds 153 seats, which is more than two times over Japan which holds 64 seats. And in the same list, Luxembourg, Malaysia, Poland, Portugal, Saudi Arabia, Singapore, Thailand, and Turkey only have one firm each. This is a skew character, too.

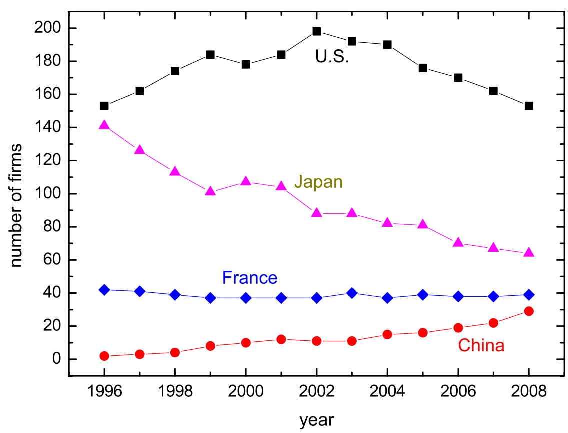

Fig. 1. The number of firms of some countries in the *Fotune Global 500*. The curves from top down are for the U.S., Japan, France and China respectively.

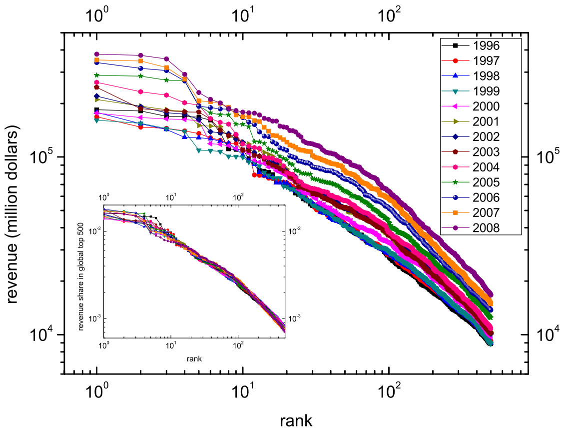

Fig. 2. The size distribution of *Fortune Global 500* from 1996 to 2008. The curves from lower to upper are approximately for earlier to latter. The subplot at left-bottom is distribution of the normalized revenue.

Zipf plot, which is a plot of the log of the rank vs. the log of the variable being analyzed, is very proper for discussion about upper tail distribution. Fig. 2 shows the log of firm revenue vs. the log of corresponding rank in *Fortune Global 500* for each of all the 13 lists. The sub-figure at left-bottom of this Figure shows the log of revenue share, which is the proportion of a firm's revenue among the *Fortune Global 500*'s total revenue in the year, vs. the log of firm rank in that year.

These curves in Fig. 2 are approximately parallel, and the curves in the sub figure almost go all the way with them. It means these data follow the same distribution. Then we use Kolmogorov-Smirnov (KS) test as a goodness of fit test. KS statistic is simply the maximum distance between the cumulative distribution functions (CDF) of the data and the fitted model:

$$D = \max_{x \in R} |F(x) - P(x)| \,. \tag{1}$$

Here $F(x)$ denotes the first CDF and $P(x)$ the second CDF, which is usually a hypothesized CDF.

We found the KS statistics for every couple among the 13 lists is no more than 0.034, and there is no sufficient evidence and the null hypothesis of same distribution cannot reject at level 0.1. So we have the first conclusion, although there are dynamics in size distribution of Fortune Global 500, but we argue the distribution retain consistently for all years. Which character hold in the upper tail distribution? In next section, we will fit the average revenue share by some classic forms.

3 The Fitting

In this section, we will fit the average revenue share with four kinds of functions including simple power law (Zipf law), SCL, FSS, and lognormal distributions which were frequently used in literatures on firm size.

the Zipf's law has the form

$$s_r = \beta_1 r^{-\beta_2}, \tag{2}$$

where s_r is a company's size, and r is its rank of revenue in the list, beginning with the largest; β_1 is a normalizing coefficient, and β_2 is Zipf exponent [4]. The corresponding exponent in power law is $1 + 1/\beta_2$ [16].

The so-called SCL is defined as

$$s_r = \beta_1 (r + \beta_2)^{-\beta_3}, \tag{3}$$

where β_1, β_2 and β_3 are the parameters [4,10,13]. It is obvious that if we set $\beta_2 = 0$, equation 3 will return to the simple Zipf's law.

The fitting curve named Finite Size-Scaling (FSS) has two components: power law and exponential properties as follows [13,17],

$$s_r = \beta_1 r^{-\beta_2} e^{-\beta_3 r}. \tag{4}$$

And finally comes the last one–log-normal distribution.

$$s_r = \exp\left[\beta_1 \Phi^{-1}\left(1 - \frac{r}{\beta_2}\right) + \beta_3\right], \tag{5}$$

where Φ is a function of standard normal cumulative distribution, and β_3 and β_1 are the mean and standard deviation of $\ln(s_r)$; while the β_2 is firm number coefficient [8].

We use nonlinear least-squares regression to fit the average revenue share with these models. The fitting results are shown in figure 3 and table 1.

The estimated exponents' values are listed in column 2 to column 4 in table 1. The estimated mean squared error (MSE), which is a important index reflecting the goodness of fitting, are listed in column 5. The MSE values indicate the SCL, FSS and lognormal models have better fitting goodness than the Zipf's because

Table 1. Estimated exponents, corresponding mean squared error and KS statistics for Zipf's, SCL, FSS and lognormal models

fitting function	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	MSE	KS statistics
Zipf	0.022	0.477	–	1.919×10^{-7}	0.069
SCL	0.048	4.311	0.644	1.423×10^{-8}	0.017
FSS	0.019	0.381	0.002	5.402×10^{-8}	0.016
lognormal	1.105	1130	-7.440	3.207×10^{-8}	0.007

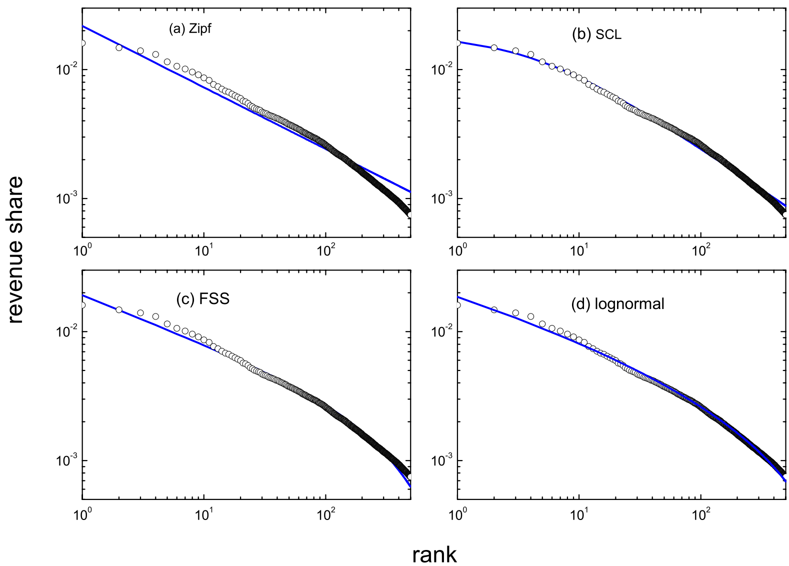

Fig. 3. The fitting results. the circles are the average revenue share. The solid lines are fitting curves from the four models as (a) Zipf's law, (b) SCL, (c) FSS and (d) the lognormal.

they are smaller. However, if KS statistics values are considered, the lognomal model is more appropriate. Although the SCL fitting has minimum MSE, the estimated values are lower than real share value in a long range from rank 41 to 232, so the fitting for the whole data is not good enough. By contraries, the lognormal is well fitted at the whole scope in spite that it is not well fitted for some single data.

4 Explanation

How to explain the worldwide firm size distribution is approximately consistent with lognormal model? The *Fortune Global 500* are firms from several countries and districts. If we know the distributions for firms of each country or district, we could know the distribution for the whole world. Actually, it is pointed out that the sum of multi lognormal variables have no determined distribution function, but it would be approximately lognormally distributed, especially when the number of original distributions is small and the independent distributions are similar [18,19]. So, if the firm size from each single country follows lognormal distribution, the worldwide distribution is approximately lognormal. Luckily, researchers have found that some country-wise firm size distribution follows lognormal distribution, so we can give a simple explanation for the worldwide phenomenon.

For convenience, we put all firms into three groups, the first group only includes the U.S., which is the biggest economy and the most powerful country in

Table 2. Estimated exponent values, mean squared error values

groups	fitting function	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	MSE	KS statistics
group 1	Zipf	4.374×10^5	0.573	–	1.123×10^8	0.054
	SCL	7.110×10^5	1.328	7.046×10^{-1}	6.709×10^7	0.022
	FSS	4.110×10^5	0.481	5.573×10^{-3}	6.721×10^7	0.007
	lognormal	1.408	639	8.742	6.404×10^7	0.007
group 2	Zipf	3.375×10^5	0.488	–	1.015×10^8	0.080
	SCL	1.034×10^6	5.283	7.54×10^{-1}	2.736×10^7	0.022
	FSS	2.923×10^5	0.333	7.357×10^{-3}	6.899×10^6	0.010
	lognormal	0.909	232	10.186	8.413×10^6	0.010
group 3	Zipf	3.480×10^5	0.541	–	4.172×10^7	0.046
	SCL	3.939×10^5	0.336	5.738×10^{-1}	3.994×10^7	0.037
	FSS	3.288×10^5	0.472	3.552×10^{-3}	2.474×10^7	0.011
	lognormal	1.532	2103	7.620	2.831×10^7	0.013

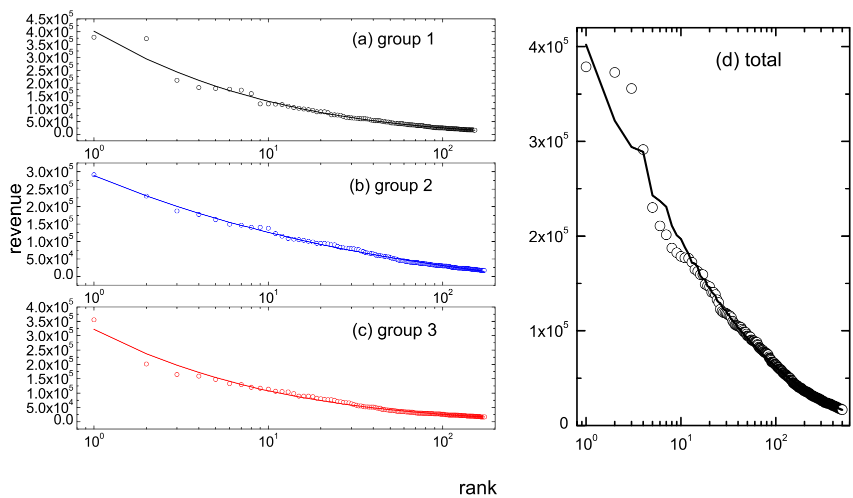

Fig. 4. The lognormal fittings for groups. (a) is for group 1 or the U.S. firms, (b) group 2 or the firms of Japan, France, Germany and the U.K., (c) other countries' firms or group 3, (d) is comparison of all the real data with integration of three estimated lognormal distributions.

the world and takes the biggest share more than 30% in each year. The second group includes Japan, France, Germany and U.K., which are developed countries and have some similar characteristics. We put all the other countries and districts into group 3.

For instance, we discuss the list in 2008. The number of group 1 to 3 are 153, 173, 174 perspectively.

The lognormal fittings for each group is shown in table 2. The fittings of the other three models for each group are also given in order to compare. We find the fittings of the lognormal and FSS models have higher goodness for each group than the Zipf and the SCL since the KS values are very small. So we can not reject the H_0 hypothesis that the true distributions follow the lognormal or FSS. Considering the better fitting results of logrmal for the whole data as shown in table 1 and the empirical evidences of the lognormal distribution provided in many works such as [3,5,7,8], we think that the lognormal is more appropriate. The fitting results of lognormal model for each group are good as shown in subfigure (a) to (c) in Fig. 4. We also compare all the real data with the integration of three estimated lognormal distributions, it is acceptable well.

5 Conclusions

In this paper, we focus on the worldwide firm size distribution which seems to be neglected by former researches studying firm size distribution. Through data of all the thirteen *Fortune Global 500* lists from 1996 to 2008, we found that although the firms, the firms' revenue, and relative countries and districts varies a lot, the revenue share distribution is fairly stable all the time. Using KS test, we confirm that distribution of each list is the same and there is no enough diversity to reject the null hypothesis. We fitted empirical average distribution of raw data by four kinds of classic functions used commonly in relative literatures. The KS statistics values support that lognormal one is the best model for *Fortune Global 500* firm size distribution while the simple Zipf's law is not an ideal one. Then, we argue that the collective lognomal-like character is derived from integration of single lognormal functions.

Acknowledgments

This work was supported by the National Science Foundation of China under grants of No. 60534080 and No. 70601002.

References

1. Ijiri, Y., Simon, H.A.: Skew Distributions and the Sizes of Business Firms. North-Holland, Amsterdam (1977)
2. Kleiber, C., Kotz, S.: Statistical Size Distributions in Economics and Actuarial Sciences. John Wiley and Sons Inc., New Jersey (2003)
3. Gibrat, R.: Les Inegalites economiques, Sirey, Paris (1931)
4. Zipf, G.: Human Behavior and the Principle of Last Effort. Addison-Wesley, Cambridge (1949)
5. Hart, P.E., Prais, S.J.: The analysis of business concentration: A statistical approach. Journal of the Royal Statistical Society, Series A 119, 150–181 (1956)
6. Simon, H.A., Bonini, C.P.: The Size Distribution of Business Firms. American Economic Review 48, 607–617 (1958)
7. Hall, B.H.: The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector. The Journal of Industrial Economics 35, 583–606 (1987)

8. Stanley, M.H.R., Buldyrev, S.V., Havlin, S., Mantegna, R.N., Salinger, A., Stanley, H.E.: Zipf plots and the size distribution of firms. Economics Letters 49, 453–457 (1995)
9. Okuyama, K., Takayasu, M., Takayasu, H.: Zipf's law in income distribution of companies. Physica A 269, 125–131 (1999)
10. Ramsden, J.J., Kiss-Haypál, G.: Company size distribution in different countries. Physica A 277, 220–227 (2000)
11. Axtell, R.L.: Zipf distribution of U.S. firm sizes. Science 293, 1818–1820 (2001)
12. Voit, J.: The Growth Dynamics of German Business Firms. Advances in Complex Systems 4, 149–162 (2001)
13. Hernández-Pérez, R., Angulo-Brown, F., Tun, D.: Company size distribution for developing countries. Physica A 359, 607–618 (2006)
14. Angelini, P., Generale, A.: On the Evolution of Firm Size Distributions. American Economic Review 98, 426–438 (2008)
15. Growiec, J., Pammolli, F., Riccaboni, M., Stanley, H.E.: On the Size Distribution of Business Firms. Economics Letters 98, 207–212 (2008)
16. Newman, M.E.J.: Power laws. Pareto distributions and Zipf's law, Contemporary Physics 46(29), 323–351 (2005)
17. Blok, H.J., Bergersen, B.: Effect of boundary conditions on scaling in the Game of Life. Physics Review E 55, 6249 (1997)
18. Beaulieu, N.C., Xie, Q.: An Optimal Lognormal Approximation to Lognormal Sum Distributions. IEEE Transactions on Vehicular Technology 53, 479–489 (2004)
19. Mitzenmacher, M.: A Brief History of Generative Models for Power Law and Lognormal Distributions. Internet Mathematics 1, 226–251 (2004)