Size-Dependency of Income Distributions and Its Implications

Jiang Zhang[*] and You-Gui Wang[†]

Department of Systems Science, School of Management, Beijing Normal University

(Dated: December 13, 2010)

This paper highlights the size-dependency of income distributions, i.e. the income distribution curves change with the population of a country systematically. By using the generalized Lotka-Volterra model to fit the empirical income data in the United States during 1996-2007, we found an important parameter λ can scale with a β power of the size (population) of U.S. in that year. We pointed out that the size-dependency of the income distributions, which is a very important property but seldom addressed by previous studies, has two non-trivial implications: (1) the allometric growth pattern, i.e. the power law relationship between population and GDP in different years, which can be mathematically derived from the size-dependent income distributions and also supported by the empirical data; (2) the connection with the anomalous scaling for the probability density function in critical phenomena since the re-scaled form of the income distributions has the exactly same mathematical expression for the limit distribution of the sum of many correlated random variables asymptotically.

I. INTRODUCTION

The power law distribution of incomes in a nation is one of the most important universal patterns found in economic systems due to the seminal work of Pareto[1]. It is suitable for not only incomes in different countries and different years [2–5] but also other complex systems e.g. languages[6] and complex networks[7]). Although this statistical law is supported by many empirical data[8] and theoretical works[9, 10], it can only describe the distribution in high incomes. Some recent studies have shown that the distribution for the great majority of population can be described by an exponential function which is very different from the power law in the high incomes[5, 11]. Silva and Yakovenko [5] defined these different income intervals as thermal and super-thermal regions whose dynamics may follow very different rules.

Some early studies in family names have pointed out that the distributions can change with the size of the system[12–14]. This size-dependency of distributions is also found in languages[15]. In this paper, we try to propose that the income distributions also have this size-dependency property which means that the distribution curves change with the system size (the population) systematically.

We also pointed out that the size-dependency of income distributions actually implies that the power law relationship between population and GDP, i.e. the allometric growth(scaling) phenomenon which is also found in various complex systems such as ecological systems[16–19], cities[20–22] and countries[23, 24]. Some studies in family names[12] and languages[25, 26] have linked the patterns of power law distributions and power law rela-

tionships of two variables. However, in this study, we argued that the exponent of the power law relationship between population and GDP doesn't depend on the Pareto exponent of income distribution but the size-dependency exponent.

Furthermore, the re-scaled form of income distribution curves can be re-expressed as a generalized mathematical form. This formula actually has been found to describe the anomalous scaling of probability density function in critical phenomena, e.g. spin systems[27, 28], where the re-scaled distribution form can be treated as a limit distribution of the sum of a large number of correlated random variables[27, 28]. Therefore, the size-dependency of income distribution also implies the connection with the central limit theorem of correlated variables.

This paper is organized as follows: In Section II, we used the revised form of the generalized Lotka Volterra model to fit the empirical income data of U.S. during 1996-2007. In this formula, we inserted a scale factor λ which changes with the population as a power law with exponent β in different years. Then, the income distribution can be re-scaled by the population of the United States in that year. In Section III we presented how one can lead to the allometric growth pattern from the size-dependency property and tested this pattern with empirical data of the population and the GDP in U.S. during the same years. And we stressed that the allometric exponent mainly relates to the re-scale exponent of the income distributions but not the Pareto exponent. Finally, in Section IV, we summarized a generalized formula of the size-dependent distribution which is identical to the anomalous scaling probability density function for the limit distribution of the sum of many correlated random variables.

---

[*] zhangjiang@bnu.edu.cn; http://www.swarmagents.cn/jake
[†] ygwang@bnu.edu.cn

II. SIZE-DEPENDENT INCOME DISTRIBUTIONS

The personal-income distribution data in the United States during 1996-2007 is available. This data is compiled by the Internal Revenue Service (IRS) from the tax returns in the USA for the period 1996-2007(presently the latest available year [29]). The original data gives the percentage in given income intervals. We can plot the cumulative distributions in Figure 1.

As pointed out by [5], the distribution curves exhibit exponential form in low incomes and power law distribution in high incomes. However, we use the generalized Lotka-Volterra model[9, 30] instead of the method in [5] to fit these data since the generalized LV model only needs two free parameters. We assume the density curves of income distributions in different years follow the equation,

$$f(x) = \lambda \frac{(\alpha - 1)^\alpha}{\Gamma(\alpha)} \frac{\exp(-\frac{\alpha-1}{\lambda x})}{(\lambda x)^{1+\alpha}}, \tag{1}$$

where, α and λ are parameters needed to be estimated. $\Gamma()$ is the Euler gamma function. Note that, in the original form of generalized-LV model[30], there is no factor λ since the main purpose of that paper is to give an explanation of the shape of the income distribution. However, we must insert this factor in Equation 1 because we care not only the shape of the income distribution curves but also its dependency on size (the population of a country) in different years. And this size-dependent property can only be reflected by λ. In addition, α is the Pareto's exponent in high incomes regime because Equation 1 has a truncated power law form. Nevertheless, we will use the cumulative distribution function instead of Equation 1 directly to reduce the estimated errors.

$$C(x) = \int_x^{+\infty} f(x)dx = 1 - \frac{\Gamma(\alpha, \frac{\alpha-1}{\lambda x})}{\Gamma(\alpha)}, \tag{2}$$

where the function $C(x)$ is the probability that a person whose income is larger than x. Two steps fitting method is used in this paper. At first, we apply Equation 2 to the empirical data year by year and obtain the best estimations of parameters α and λ. Here, the "best" means the total distances between empirical data and theoretical curves on the log-log coordinate are minimized. The αs derived by the first step are (1.59043, 1.60112, 1.61717, 1.67562, 1.75147, 1.76187, 1.71051, 1.61152, 1.80999, 1.95683, 1.87374, 1.92885), they fluctuate around the mean value 1.74076. Second, we fix $\alpha = 1.74076$ and use the same method to obtain the best estimation of λ again. We will show that the size-dependency and implications actually are independent on α. The reason of using this two steps fitting method is to get a better estimation of λ which is more important than α. From Figure 1, we can see that the distributions change over time regularly. As time goes by, the distribution curves shift. This trend is
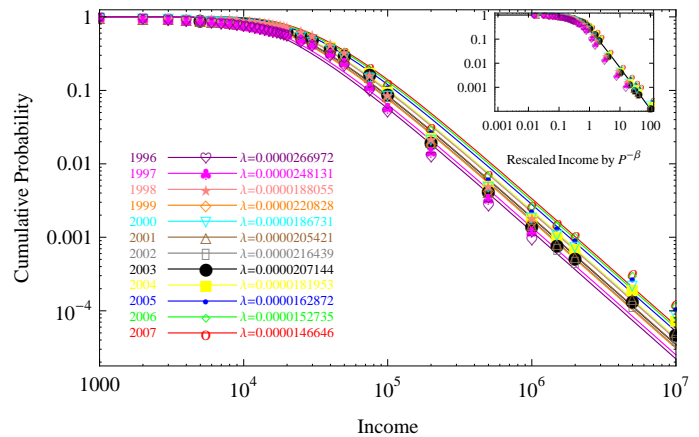

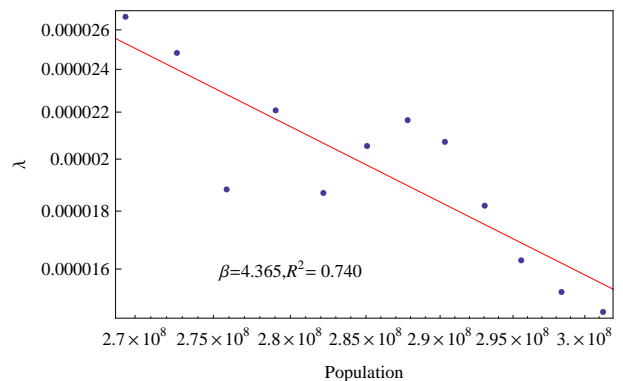FIG. 1. Income Distributions of U.S. in the period of 1996-2007.

FIG. 2. The Power Law Relationship between the Population and λ during 1996-2007

more obvious in the scaling regions (high income tails). The relationship between the best estimations of λ and years is shown in the legend of Figure 1.Furthermore, we know that the population of U.S. increases with time in these years. Therefore, λ actually is the function of population at a given year. This functional relationship can be presented by a power law relationship between population and λ as Figure 2 shows. From Figure 2, we can observe an apparent trend that λ decreases with population. This trend can be approximated by a power law relationship between λ and population.

$$\lambda \sim P^{-\beta}, \tag{3}$$

where β estimated as 4.365 is the slope of the line in Figure 2. Therefore, we conclude that the income distributions are size(population) dependent. This dependency is described by the power law relationship between the scale parameter λ and the population.

As a result, the income distributions in different years

can be re-scaled by $P^{-\beta}$,

$$f(x) \sim (P^{-\beta}) \frac{(\alpha-1)^\alpha}{\Gamma(\alpha)} \frac{\exp(-\frac{\alpha-1}{P^{-\beta}x})}{(P^{-\beta}x)^{1+\alpha}}. \qquad (4)$$

The re-scaled curve of income distribution is shown in the inset of Figure 2. We can conclude that the income distributions of the United States during 1996 to 2007 are size dependent. This size dependency is reflected by the power law relationship between the population and the scale factor λ in each year. Next, we will discuss the two important implications of this size-dependency.

III. POWER LAW RELATIONSHIP BETWEEN POPULATION AND GDP

We will show that the size-dependency of income distributions implies the power law relationship between population and GDP. At first, we know that the GDP of a country is proportional to the total incomes of all people[31]. Second, the total incomes can be read from the income distribution curve. We can write down the equation,

$$X \sim P\langle I \rangle, \qquad (5)$$

where, X is the GDP, P is the population of a given year. I is the random variable income in a given year. And $\langle I \rangle$ stands for the ensemble mean value of incomes. Therefore, $P\langle I \rangle$ is just the total incomes of the whole country in the given year.

Then we can calculate the mean income from the cumulative probability function (Equation 2) as follow,

$$\langle I \rangle = \int_0^{+\infty} x f(x) dx = \int_0^{+\infty} C(x) dx = \frac{1}{\lambda}. \qquad (6)$$

Therefore,

$$X \sim P/\lambda, \qquad (7)$$

bring equation 3 into 7, we get:

$$X \sim P^{1+\beta}. \qquad (8)$$

Equation 8 is just the power law relationship (allometric growth) between population and GDP with an exponent $1+\beta$. We have estimated the exponent $\beta \sim 4.365$ from the income distributions. Therefore, we predict that the GDP is a 5.365 power of the population in the United States during 1996-2007.

On the other hand, we can obtain the real data of the population and the GDP of the United States during the given period. The two variables have a power law relationship which is shown in Figure 3. From the empirical data, we can estimate the power law exponent is about 5.065 which is closed to the exponent we have predicted from size-dependent income distribution data (the rela-
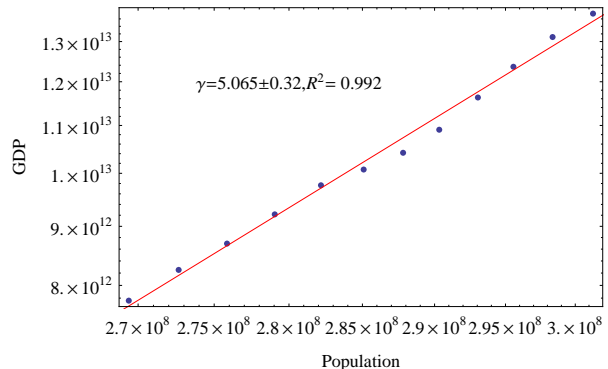

FIG. 3. Power Law Relationship between the Population and the GDP of U.S. during 1996-2007

tive error is $|5.365 - 5.065|/5.065 \approx 6\%$). However, there is still a little deviation between the empirical exponent and predicted one. The possible error sources may include (1) the income curves; (2) the estimation of β; (3) the deviation between GDP and total incomes.

IV. GENERALIZED SIZE-DEPENDENT INCOME DISTRIBUTION

One of an interesting fact which deserves more attention is the size-dependency of income distribution and its implication of the power law relationship between population and GDP are independent on the Pareto exponent α in the income distribution formula Equation 1. Therefore, we can further hypothesize that the size-dependency of distribution is a unique property independent on the concrete form of the density function.

Actually, from Equation 4, we can generalize an abstract form of the probability density function,

$$f(x) \sim P^{-\beta} g(P^{-\beta}x). \qquad (9)$$

Where, $g(y)$ is an arbitrary probability density function with size-independent argument y. We know that when we set $g(y)$ as the concrete form, $\frac{(\alpha-1)^\alpha}{\Gamma(\alpha)} \frac{\exp(-\frac{\alpha-1}{y})}{y^{1+\alpha}}$, we get the generalized LV model in Equation 4.

Actually, the power law relationship between population and GDP which is discussed in Section III can be derived from the abstract form (Equation 9) because,

$$\langle x \rangle = \int_0^{+\infty} x f(x) dx \sim \int_0^{+\infty} x P^{-\beta} g(P^{-\beta}x) dx, \quad (10)$$

replace the integral variable x with $y = P^{-\beta}x$, we obtain,

$$\langle x \rangle \sim \int_0^{+\infty} P^{-\beta} P^\beta g(y) P^\beta dy = P^\beta \int_0^{+\infty} g(y) dy \sim P^\beta, \qquad (11)$$

where, $\int_0^{+\infty} g(y)dy$ is a constant because $g(y)$ is size-independent. Finally, we can also obtain the power law relationship between population and GDP,

$$GDP = P\langle x \rangle \sim P^{1+\beta}. \tag{12}$$

So, we can conclude that the essence of size-dependency in income distribution is captured by Equation 9. Actually, this re-scaled form distribution is not first discovered by this paper. In [27, 28], the authors also gave a similar formula to describe the anomalous scaling probability density function in critical phenomena,

$$f(x) \sim n^{-D}g(n^{-D}x). \tag{13}$$

Where, n is the size of the system (the number of addends), D is also a re-scaled exponent. The same mathematical form must imply the ubiquitous natural laws, so the individual income can be viewed as a sum of many correlated random variables related to each person in the same country. However, we will not discuss the detail of this discovery and leave it to the future studies because of the size limitation of this paper.

V. CONCLUDING REMARKS

This paper discussed the size-dependency of income distribution which is a very important property and ignored more or less by previous studies. The size-dependency has two important implications: 1. The power law relationship between population and GDP (which is also known as allometric growth); 2. The re-scaled form of income distribution has the same mathematical form for the anomalous scaling probability density function of the sum of many correlated random variables.

Although the size-dependency of income distributions is only supported by the incomes data of U.S. during 1996-2007 in this paper, we believe that the results can be generalized to other countries and years. In addition, we have found the same size-dependency phenomena in human online behaviors [32], therefore, it is reasonable to accept the results in this paper as common ones.

[1] V. Pareto, *Cours dconomie Politique: Nouvelle dition par G.-H. Bousquet et G. Busino* (Librairie Droz, Geneva, Paris, 1964).

[2] V. M. Yakovenko and J. B. Rosser, Reviews of Modern Physics, **81**, 1703 (2009).

[3] F. Clementi and M. Gallegati, Physica A: Statistical Mechanics and its Applications, **350**, 427 (2005).

[4] N. Ding and Y. Wang, Chinese Physics Letters, **24**, 2434 (2007).

[5] A. C. Silva and V. M. Yakovenko, Europhysics Letters, **69**, 304 (2005).

[6] G. K. Zipf, *Selected Studies of the Principle of Relative Frequency in Language*, first ed. ed. (Harvard University Press, 1932).

[7] R. Albert and A. L. Barabasi, Reviews Modern Physics, **74**, 47 (2002).

[8] W. Souma, cond-mat/0011373 (2000).

[9] P. Richmond, S. Hutzler, R. Coelho, and P. Repetowicz, in *Econophysics and Sociophysics* (Wiley-VCH Verlag GmbH & Co. KGaA, 2006) pp. 131–159, ISBN 9783527406708.

[10] A. Chatterjee, B. K. Chakrabarti, and S. S. Manna, Physica A: Statistical Mechanics and its Applications, **335**, 155 (2004).

[11] A. Dragulescu and V. Yakovenko, Physica A: Statistical Mechanics and its Applications, **299**, 213 (2001).

[12] S. Miyazima, Y. Lee, T. Nagamine, and H. Miyajima, Physica A: Statistical Mechanics and its Applications, **278**, 282 (2000).

[13] B. J. Kim and S. M. Park, Physica A: Statistical Mechanics and its Applications, **347**, 683 (2005), ISSN 0378-4371.

[14] H. Anh, T. Kiet, S. K. Baek, B. J. Kim, and H. Jeong, Journal of the Korean Physical Society, **51**, 1812 (2007).

[15] S. Bernhardsson, L. E. C. da Rocha, and P. Minnhagen, New Journal of Physics, **11**, 123015 (2009).

[16] S. Nordbeck, Geografiska Annaler. Series B, Human Geography, **53**, 5

[17] M. Kleiber, Hilgardia, **6**, 315 (1932).

[18] J. Brown and G. West, *Scaling in Biology* (Oxford University Press, 2000).

[19] G. B. West and J. H. Brown, The Journal of Experimental Biology, **208**, 1575 (2005).

[20] A. Isalgue, H. Coch, and R. Serra, Physica A: Statistical Mechanics and its Applications, **382**, 643 (2007).

[21] L. M. Bettencourt, Research Policy, **36**, 107 (2007).

[22] L. M. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G. B. West, Proceedings of the National Academy of Sciences of. U. S. A., **104**, 7301 (2007).

[23] B. Roehner, International Journal of Systems Science, **15**, 917 (1984).

[24] J. Zhang and T. Yu, Physica A: Statistical Mechanics and its Applications, **389**, 4887 (2010).

[25] L. Lu, Z. Zhang, and T. Zhou, PloS ONE, **5**, e14139 (2010).

[26] D. C. van Leijenhorst and T. P. van der Weide, Information Sciences, **170**, 263 (2005).

[27] F. Baldovin and A. L. Stella, Physical Review E, **75**, 020101 (2007).

[28] A. L. Stella and F. Baldovin, Journal of Statistical Mechanics: Theory and Experiment, **2010**, P02018 (2010).

[29] "http://www.irs.gov/taxstats/indtaxstats/article/0,,id=134951,00.html,".

[30] O. Malcai, O. Biham, P. Richmond, and S. Solomon, Physical Review E, **66**, 031102 (2002).

[31] O. Blanchard, *Macroeconomics* (Prentice Hall, 2000).

[32] J. Zhang, "Allometric scaling and Size-Dependent distributions of massive human online behaviors," Unpublished.