A Letters Journal Exploring the Frontiers of Physics

November 2009

EPL, **88** (2009) 38001 www.epljournal.org
doi: 10.1209/0295-5075/88/38001

Deformed Zipf's law in personal donation

Q. Chen, C. Wang and Y. Wang[(a)]

Department of Systems Science, School of Management, Beijing Normal University - Beijing 100875, PRC

received 30 September 2009; accepted in final form 21 October 2009
published online 4 November 2009

PACS 89.65.-s – Social and economic systems
PACS 89.75.Da – Social obeying scaling laws
PACS 89.65.Gh – Economics; econophysics, financial markets, business and management

Abstract – The power law or Zipf's law phenomena in human behaviors have been widely observed and attracted intensive attention. In this letter, a new evidence on personal donation is presented and analyzed. A sample of donation to the victims of Sichuan earthquake in 2008 demonstrates that donation distribution has a particular pattern. The upper part is governed by Zipf's law and the lower part exhibits a uniform distribution. We propose a theoretical model in which people's wealth distribution follows a power law, they are willing to donate a random part of their wealth and have preferences on some specific numbers. This model provides us not only a reasonable explanation on the empirical donation pattern but also an effective method to get access to large-scale personal-wealth distributions.

Introduction. – The power law distribution or Zipf distribution has attracted much attention because it is found in an extraordinarily diverse range of natural and social phenomena [1–7]. In economic and social systems, many power law phenomena come from the competition among individual participants for a constraint resource, such as personal income [8], firm size [9], city size [10], and country wealth [11]. The interaction among individuals plays an essential role in the formation of these collective patterns. Meanwhile, heavy-tail phenomena can also be derived from independent individual behaviors or activities, for instance, papers' citation [12], human travel distance [13], surface mail communication [14], web browsing [15], movie watching [16], e-mail communications [17], and library loans [18]. When exploring the formation of the collective pattern in these cases, the interactive impacts of individual choices can be neglected.

The individual activities are mostly determined by his or her local external situation and inner psychology. It is very interesting to ask how the large-scale patterns emerge from almost independent behaviors of human individuals. Recently a new research branch, namely *Human Dynamics*, was formed focusing on the formation of universal heavy-tail pattern which deviates from the Poisson one [19]. All these facts reflect the timing character of human activities. Actually, some non-timing characters also come to the similar collective pattern, such as papers' citation [12], human travel distance [13], but they have been less concerned.

In this paper, we provide a new evidence on the non-timing character of individual activities. As a human behavior, charity or donation is determined by individual wealth and psychology motivation. Since people differ in personal wealth and generosity, the donation amounts of individuals should be different from each other. As each donator decides his or her giving independently, it seems that no particular pattern of distribution is expected. Recently, F. Schweitzer and R. Mach have found that the distribution of donations in some cases follows a power law [20]. Then we have some questions in hands. Is this collective pattern universal? If so, how does it emerge from the activities of every single individual?

In this work, we show that collective donations present a particular pattern. To explain this pattern, we propose a primal-donation model in which personal-wealth distribution and donation motivation are taken into account. To further reproduce the empirical facts, we take number preference [21] of donators into account.

Besides the fact that this new evidence can enrich the research field of collective human behavior, we think it can also help us to figure out the statistical character of personal wealth in the economic system under consideration. Lists of the richest in some economies can be easily found in some magazines and webs, for instance *New Fortune* in China, *Forbes* in USA, but information on the wealth of most people is not available there. In

(a) E-mail: ygwang@bnu.edu.cn

Table 1: The number of recording for different donation amounts.

x	$(0,1)$	1	$(1,2)$	2	$(2,5)$	5
n	187	556	23	121	105	506
x	$(5,10)$	10	$(10,20)$	20	$(20,50)$	50
n	162	2566	346	3355	2321	17078
x	$(50,10^2)$	10^2	$(10^2, 2\times10^2)$	2×10^2	$(2\times10^2, 5\times10^2)$	5×10^2
n	2020	66146	3648	44554	19937	26359
x	$(5\times10^2, 10^3)$	10^3	$(10^3, 2\times10^3)$	2×10^3	$(2\times10^3, 5\times10^3)$	5×10^3
n	7244	21697	3444	4800	3167	1431
x	$(5\times10^3, 10^4)$	10^3	$(10^4, 2\times10^4)$	2×10^4	$(2\times10^4, 5\times10^4)$	5×10^4
n	661	1169	276	159	179	46
x	$(5\times10^4, 10^5)$	10^5	$(10^5, 2.8\times10^6)$			
n	32	18	39			

such a case, it is hard to figure out the actual statistical character of personal wealth for the whole economy; while our work may provide a potential way to approach it.

In this paper,we first analyze a sample of donation data collected in the event of Sichuan earthquake in 2008, then offer a theoretical explanation on the empirical pattern. In the following section, we briefly introduce the data source and show some main characters. In the third section, a primal-personal-donation model is presented. In the fourth section, we rectify the model by considering personal number preference.

The data source. – The data we study in this paper were drawn from *Chinese Red Cross Foundation*. On May 12th, 2008, an earthquake with a magnitude of 8.0 struck Sichuan province, southwest China. It has been the most devastating earthquake happened in this country in the past three decades. Chinese people and corporations donated to the victims voluntarily through some non-profit foundations. The primary data was collected from *Chinese Red Cross Foundation* by 21 June 2008. There are 250477 original records among which 234352 records are personal donators.

The obvious features of these data are summarized as follows. a) The donation amount varies enormously from 0.01 RMB to 2.79 million RMB. b) Records of quite small or quite large donation amount are very few. There are 187 donators with a donation amount of less than 1 RMB and 39 with an amount of more than 100 thousand RMB. c) Most donations have a medium amount. There are 189585 donators whose givings range from 100 RMB to 1000 RMB. d) The donations congregate at some specific numbers. For example, there are 66146 records with donation amount of 100, 44554 records with 200, 26359 records with 500, 21697 records with 1000. In contrast, only 31 with 99, 3 with 199, 1 with 499, and 17 with 999. These characters can be seen more clearly in table 1, where x presents the range of donation amount and n denotes the corresponding quantity of records.

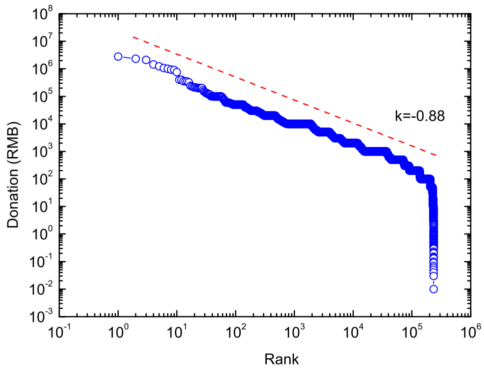

Fig. 1: (Colour on-line) The volume of donation *vs.* the corresponding rank in log-log scale. The dashed straight line is for comparison.

There are 205006 records whose amounts are not less than 100. They cover the main part (87.5%) of the total sample. In this part, we found that the top minority of the donators contribute to the majority of the aggregate donation amount. On the contrary, the contribution of other donators is very limited. The relations between donation amounts and their corresponding ranks are shown in fig. 1. The donation amount goes downward in double-log scales approximately as a line with an abrupt bending at the tail. To make a comparison, we draw a dashed straight line with a slope of -0.88 above the flatter part of the curve, indicating a Zipf character.

In order to check whether the upper part of donation data obeys Zipf's law, we should proceed to make a parameter estimation and hypothesis test to confirm it. First, Zipf's exponent can be obtained by ordinary least squares technology (OLS). Then we use Kolmogorov-Smirnov (KS) test [22] to check the fitting goodness of the empirical distribution by calculating the maximum distance between the cumulative distribution function (CDF) of the real data and the hypothetical Zipf distribution.

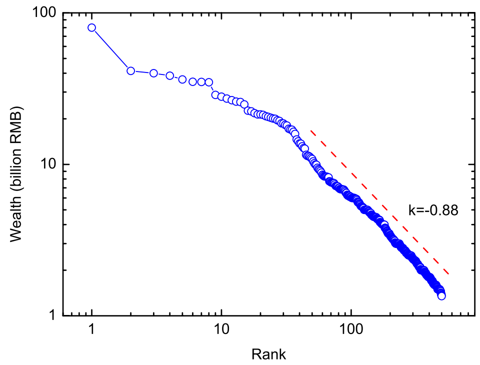

Fig. 2: (Colour on-line) Zipf plot of the top 500 richest Chinese.

In our estimating and checking process, the rank of the sampling segment is from 1 to 135212, and the corresponding donation amount ranges from 200 RMB to 2.79 million RMB. The estimated exponent $\hat{\beta}=0.88$, while the corresponding determination coefficient of the regression R^2 is equal to 0.95. After data standardization, we compare our estimated distribution with the standard Zipf's distribution with exponent 0.88, and get Kolmogorov-Smirnov statistics $D=0.14$ which is so big that we have to reject the null hypothesis of Zipf's law.

The big value of KS statistic means a large deviation of the empirical data from the standard Zipf's distribution. From fig. 1 we can see that the upper part is composed of many sawtooth-like pieces, which lead to a large deviation. We believe that the sawtooth comes mainly from congregation of the donation on some specific numbers. If the sawtooth pattern were properly removed, then we would have a smooth trend of the whole data. The trend line can be divided into two parts. The upper one maintains the character of standard Zipf distribution, while the tail is almost a vertical straight line. We will firstly develop a theoretical explanation of the trend curve in the next section.

The primal-donation model. – The trend curve of donation is well regulated and follows Zipf's law in the upper part and then turns down abruptly at the tail. How is the global pattern formulated? The answer must lie on individual donation decisions. Common sense tells us that the kernel determinant of one's donation depends on his or her personal wealth. In fact, some empirical works have found that the majority of the wealth variables are statistically significant for participation in charity [23], and one's donation should be a part of his or her total wealth. The largest donation certainly comes from the richest. Sometimes the charity behavior is a signal of his or her absolute wealth [24].

Wealth inequality is common in every economy and has attracted some researchers' attention [25–27]. Particularly, Chinese wealth distribution from the year 2003 to 2005 has been found to follow Zipf's law with an exponent around 1 [28]. We also got the data from the same source in 2008 and show it in fig. 2. Zipf's law still holds in the wealth distribution from rank 30 to 500, and the exponent is estimated to be 0.88 by Gabaix's technology [29]. It is confirmed by Kolmogorov-Smirnov test for $D=0.01$.

Based on these facts, we propose a simple theoretical model to explain the main characters of donation. The postulations are listed as follows.

a) There are N agents who have the ability to donate in this system.

b) The personal wealth follows Zipf's law, which says the amount of wealth for an agent is given by

$$w_i = w_{\max} \times i^{-\beta}, \tag{1}$$

where i denotes the rank of the agent, and $w_{\max}$ is the wealth of the richest in the system. So the poorest has a wealth of $w_{\min} = w_{\max} \times N^{-\beta}$.

c) Only a portion of the agents have a desire to donate. In other words, there is a probability for each agent to donate. In addition, the sample is only a small part of all donators. Taking these two factors into account, we postulate the possibility which is the same for each agent to become a donator and be observed in a certain sampling. Given the possibility as p, the mean number of the observed donators is pN.

d) Donator i would give a share s_i of his or her total wealth w_i, where s_i reflects the people's donating desire. Generally, s_i is a random variable uniformly distributed in $(0,\lambda)$, where $\lambda \leqslant 1$. This postulation means that the donation amount z_i can be expressed as

$$z_i = s_i w_i. \tag{2}$$

Though this amount just represents the primal willingness of the donator, the actual giving may be modified due to some reasons, which will be discussed in the next section. So we call it primal donation.

We carried out the simulations of the model at a given set of parameters, and then made a series of comparisons of the results by varying the upper limit λ, the total number of agents N, the Zipf exponent β, the wealth of the richest $w_{\max}$, and the possibility p, respectively. The simulation results of the wealth values and primal donations for all cases are shown in fig. 3. It is obvious that the primal-donation distribution is robust for different parameters. In all subplots of fig. 3, the upper part of the distribution is very likely to follow Zipf's law and the Zipf exponent is equal to that of the corresponding wealth distribution. A turning point can be easily found in each case, and it will shift as any one of the parameters changes.

The characters of primal donation can also be obtained from mathematical derivation. For simplicity, we employ a continuous expression of wealth distribution to derive the donation distribution. Since the wealth distribution obeys a power law between the maximum and the minimum of wealth, the probability density function then can be normalized as

$$p\left(X=w\right) = \frac{\alpha-1}{\left(w_{\min}^{1-\alpha} - w_{\max}^{1-\alpha}\right)} w^{-\alpha}, \tag{3}$$

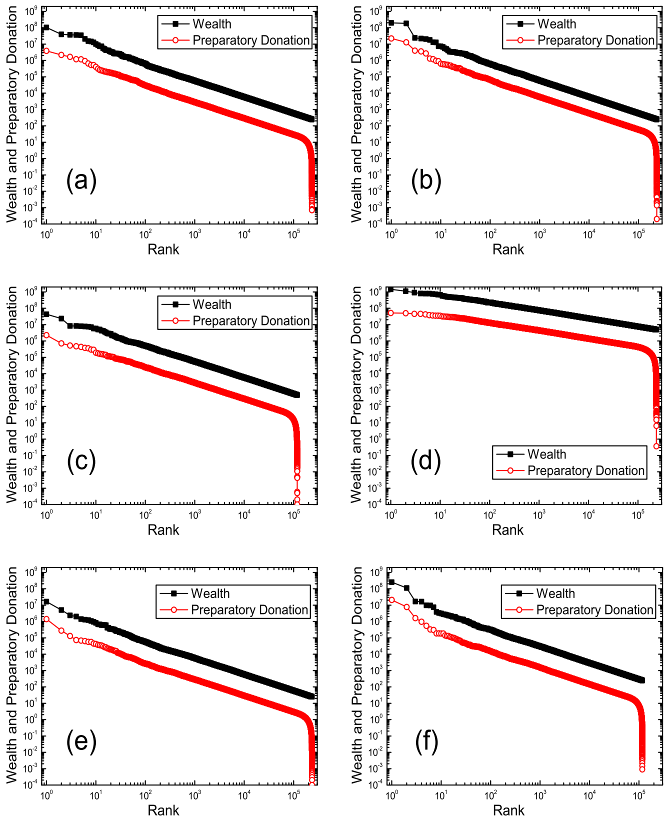

Fig. 3: (Colour on-line) The Zipf plot of wealth and primal donation for a given setting (a) $N=4\times10^8, \beta=1, w_{\max}=10^{11}, p=5.8588\times10^{-4}, \lambda=0.1$. Each of the other subplots shows differences from (a) by varying only one parameter respectively, (b) $\lambda=0.2$; (c) $N=2\times10^8$; (d) $\beta=0.5$; (e) $w_{\max}=10^{10}$; (f) $p=2.9294\times10^{-4}$.

where $w_{\min}\leq w\leq w_{\max}$, and the power law's exponent $\alpha=1+1/\beta$.

Since the donation is a random share of wealth as given by eq. (2), the primal-donation distribution function is the integral over $ws=z$ of the product of the power law function given by eq. (3) and the uniform distribution function. After a series of manipulations, the expression can be written as

$$p(z)=\begin{cases}0, & z\leqslant 0,\\ \dfrac{(\alpha-1)(w_{\min}^{-\alpha}-w_{\max}^{-\alpha})}{\alpha\lambda(w_{\min}^{1-\alpha-w_{\max}^{1-\alpha}})}, & 0<z\leqslant\lambda w_{\min},\\ \dfrac{(\alpha-1)(\lambda^{\alpha}z^{-\alpha}-w_{\max}^{-\alpha})}{\alpha\lambda(w_{\min}^{1-\alpha}-w_{\max}^{1-\alpha})}, & \lambda w_{\min}<z\leqslant\lambda w_{\max},\\ 0, & z>\lambda w_{\max}.\end{cases} \tag{4}$$

The formula shows clearly that the density function is divided into two parts. When $0\leqslant z<\lambda w_{\min}$, all items in the expression of $p(z)$ are given before simulation, so it is a constant, resulting in a uniform distribution. But when $\lambda w_{\min}<z\leqslant\lambda w_{\max}$, the expression of $p(z)$ contains two terms: the former one is a constant, the latter one is a power function of z. As $w_{\max}$ is large enough so that $w_{\max}^{-\alpha}$ approaches its limit of 0, $p(z)$ will be simplified into a power function with an exponent α, which is identical to that of the presumed wealth distribution.

The emergence of the uniform tail in the primal donation can be understood in the following way. For a certain amount of donation, only the richer whose maximum desire is greater than the value of donation can contribute to it. According to the postulation d), *i.e.* eq. (2), the part s that each agent takes from his or her wealth w is assumed to be a uniform distributed random variable with an upper limit λ. Thus the maximum desire for a donator is λw. When donation is less than $\lambda w_{\min}$, everybody has donation capability. Given an amount of donation $z\in(0, \lambda w_{\min})$, all agents would contribute to this amount with the same quota. As a result, a uniform distribution over this domain is produced.

Regarding the formation of the power law character of the upper part, the major cause is obviously the presumed pattern of wealth distribution. When donation z is greater than $\lambda w_{\min}$, people whose wealth is less than z/λ cannot contribute to this amount at all. As donation increases, the total number of people who are able to donate decreases. Given a certain amount of donation, the change in the number of people who have capability at this point corresponds to the density of the primal-donation distribution, and the number equals the integral of the wealth distribution curve from the corresponding point to infinity. Therefore, the density of donation distribution has the same mode as the wealth distribution. This means that when donation increases linearly, the number of potential donators decreases at an approximate power speed.

The above two parts of the donation distribution are segmented by a turning point, which is signified by $\lambda w_{\min}$. Immediately, we have one of the determinants of the location of the point, *i.e.*, the maximum of donating desire λ. As shown in fig. 3, when λ gets larger, the turning point will shift upward. According to the derivation of postulation b), the wealth of the poorest $w_{\min}$ can be expressed in terms of N, β, and $w_{\max}$. The impacts of these factors on the turning point are illustrated respectively in fig. 3. The last determinant of the turning point is p which represents the possibility for each person to be a sampled donator. Although p does not appear in the expression of the point, it sets the horizontal position of the point with other factors together. Since the horizontal location is dependent on the number of actually observed donators, as p decreases, the number will get smaller, so that the location of the turning point will shift left. This is testified by comparing (a) and (f) of fig. 3.

Synthetical explanation on donation distribution. – Although the primal-donation model can reproduce the main Zipf-like character of donation distribution, it can not reproduce the pattern of actual donation and

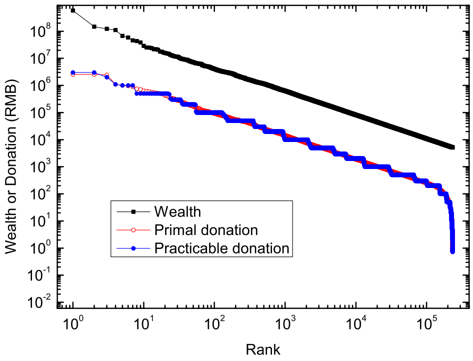

Fig. 4: (Colour on-line) The Zipf plots of simulation results of wealth, primal and practicable donations.

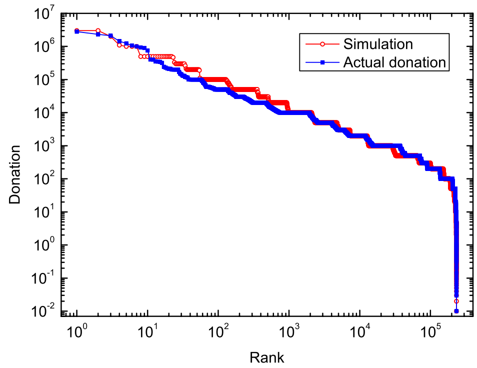

Fig. 5: (Colour on-line) Comparison between actual donation and simulation result.

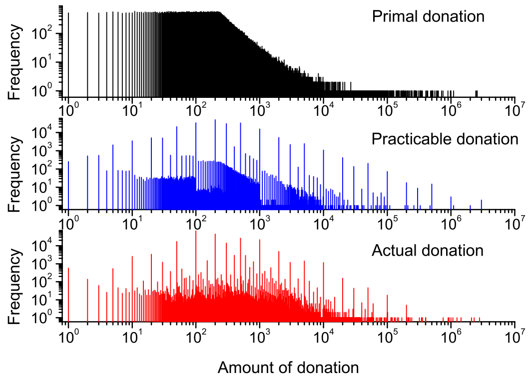

Fig. 6: (Colour on-line) Distributions of actual and simulated donations.

leaves some minor details unexplained. In order to simulate the empirical result, we first calibrated the parameters of the primal-donation model. The setting of parameters in the primal model is given in the following.

a) The total number of agents in this system N is 8×10^8, for it is generally recognized that the economic active population in China is around 8×10^8.

b) From fig. 2, we know that Zipf's law holds in the wealth distribution from rank 30 to 500. The Zipf exponent is estimated to be $\hat{\beta} = 0.88$. The extrapolation from the effective sample yields that the supposed richest person in China possesses the wealth of 352 billion RMB. According to eq. (1), the wealth of the i-th agent $w_i = 3.52 \times 10^{11} \times i^{-0.88}$(RMB).

c) In our simulation, each agent has a probability of $p = 2.9294 \times 10^{-4}$ to be an observed donator. According to the prior definition, p should be equal to the ratio of valid records D to the population N. In our case, $D = 234352$, $N = 8 \times 10^8$, then we can get the value of the probability.

d) The upper limit of s in eq. (2) is set to be $\lambda = 0.045$. This value is evaluated based on experience. This number means none of the agents will donate more than 4.5% of his or her wealth.

Following the above initial settings and process, we can get the values of primal donation. Even though these values are predeterminate, they cannot be fully presented in reality. The actual presentation of them is also affected by other factors, especially, number preference of donators.

The overwhelming majority of the agents have integer preference when making a donation, they prefer to donate in some simple integers rather than other numbers. Among 234352 donators in our sample, only 2858 donators subscribed their money with fractional currencies, and almost all the other records contains only one non-zero digit. So, we keep only 1% of primal donation unadjusted, and transform the others into integers containing only 1 non-zero digit by rounding them. As shown in table 1, the main part of the agents prefers 1, 2, 3, 5 rather than other digits. Amounts such as 50, 100, 200, 500, 1000, 2000 are more popular. In our simulation, if $m_i = 4$, it definitely turns into 5, and if $m_i = 6, 7, 8, 9$, they will be modified, with 90% probability, to either 5 or 10 at random.

As a consequence of the preceding calibration and modification of primal donation, 234210 records of the practicable donation are left at last. The simulation results of wealth, primal donation and practicable donation are shown respectively in fig. 4. The part greater than 200 RMB is confirmed to follow Zipf's law with exponent 0.88. The practicable donation displays almost the same trend as the primal one only with a somewhat slight fluctuation. The wealth distribution is also plotted in fig. 4 to illustrate the parallel between the upper part of the donation and itself.

The comparisons between the simulation results of practicable donation and the real data are presented in figs. 5 and 6. As shown in fig. 5 the similarity between them is obvious. The two curves have almost the same turning point, and the upper parts have the same slope. A slight deviation can be found in the range of 10000–500000 RMB, where the practicable donation is always greater than the actual one at the same rank. The possible reason for this deviation is that a certain part of larger donations is missed because of this private-transfer way. Some people, who donate large amount of money, are likely to subscribe their donations in more public ways.

Figure 6 shows the effects of number preference, where 3 subplots from top to bottom correspond to primal donation, practicable donation, and actual donation, respectively. The frequencies are counted in bin with a width of 1. For the primal donation, the frequency follows uniform distribution in the left part and power law in the right. On the contrary, instead of uniform arrays, many remarkable bursts emerge at some specific points in the other two subplots, indicating the number preference. The practicable donation and the empirical one resemble each other in the burst pattern.

Conclusion. – Although people subscribe their donations individually, a specific collective pattern emerges, in which Zipf's law governs the upper part and a uniform distribution dominates the lower one. The OLS estimation on a sample of donations in China indicates that the slope of the upper part of the Zipf plot of donations is about -0.88, which approximately equals Zipf's exponent of the Chinese personal wealth. To explain how this pattern is formed, we proposed a stochastic model to generate the primal donations. It is found that the Zipf plots of primal donation are robust for variant parameters. In order to account for the empirical pattern, we modified the primal donation by introducing the number preference and then got the practicable donation. The simulation results match the actual data well. From these results, we can infer that the global Chinese wealth distribution follows Zipf's law with an exponent equal to 0.88.

Like such a case of donation and wealth, many power-law phenomena coexist in complex systems. Most of the previous researches studied them separately, in spite of the facts that they might be related to each other. For instance, as proxies of a firm size both capital and revenue are found to follow a power law distribution [30]. By regarding the capital as a source of revenue, Zhang *et al.* have demonstrated that firm size distributions could be well explained [31]. Among societies all over the world, income *per capita* and the corruption level have a reverse relationship, and the empirical distributions of them have almost the same pattern with heavy tails [11,32]. Our work provides one additional evidence. We believe that it should be valuable for exploring the complexity of a system to seek for the casuality between the distributions of different elements within it and put them together into investigations.

* * *

This research was supported by National Natural Science Foundation of China under Grants of No. 70771012 and No. 60534080. We are grateful to two anonymous referees for their valuable suggestions.

REFERENCES

[1] Willis J. C. and Yule G. U., *Nature*, **109** (1922) 177.
[2] Bak P., *How Nature Works: The Science of Self-Organized Criticality* (Copernicus Press, NY) 1996.
[3] Lu E. T. and Hamilton R. J., *Astrophys. J.*, **380** (1991) 89.
[4] Roberts D. C. and Turcotte D. L., *Fractals*, **6** (1998) 351.
[5] Zanette D. H. and Manrubia S. C., *Physica A*, **295** (2001) 1.
[6] Estoup J. B., *Gammes Stenographiques* (Institut Stenographique de France, Paris) 1916.
[7] Zanette D. H., *Musicae Sci.*, **10** (2006) 3.
[8] Pareto V., *Le Cours d'Economie Politique* (Macmillan, London) 1897.
[9] Axtell R. L., *Science*, **293** (2001) 1818.
[10] Zipf G. K., *Human Behavior and the Principle of Last Effort* (Addison-Wesley, Cambridge, Mass.) 1949.
[11] Shao J., Ivanov P. Ch., Podobnik B. and Stanley H. E., *Eur. Phys. J. B*, **56** (2007) 157.
[12] Price D. J., *Science*, **149** (1965) 510.
[13] Brockmann D., Hufnagel L. and Geisel T., *Nature*, **439** (2006) 462.
[14] Oliveira J. G. and Barabási A.-L., *Nature*, **437** (2005) 1251.
[15] Dezsö Z., Almaas E., Lukács A., Rácz B., Szakadát I. and Barabási A.-L., *Phys. Rev. E*, **73** (2006) 066132.
[16] Zhou T., Kiet H. A. T., Kim B. J., Wang B. H. and Holme P., *EPL*, **82** (2008) 28002.
[17] Barabási A.-L., *Nature*, **435** (2005) 207.
[18] Vázquez A., Oliveira J. G., Dezsö Z., Goh K. I., Kondor I. and Barabási A.-L., *Phys. Rev. E*, **73** (2006) 036127.
[19] Zhou T., Han X.-P. and Wang B.-H., in *Science Matters: Humanities as Complex Systems*, edited by Burguete M. and Lam L. (World Scientific Publishing, Singapore) 2008, pp. 207–233.
[20] Schweitzer F. and Mach R., *PloS One*, **3(1)** (2008) 1458.
[21] Baird J. C. and Noma E., *Psychol. Res.*, **37** (1975) 281.
[22] Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P., in *Numerical Recipes in C: The Art of Scientific Computing,* second edition (Cambridge University Press, New York) 1992.
[23] Carroll J., Mccarthy S. and Newman C., *Econ. Soc. Rev.*, **36** (2005) 229.
[24] Glazer A. and Konrad K. A., *Am. Econ. Rev.*, **86** (1996) 1019.
[25] Levy M. and Solomon S., *Physica A*, **242** (1997) 90.
[26] Dragulescu A. and Yakovenko V. M., *Physica A*, **299** (2001) 213.
[27] Sinha S., *Physica A*, **359** (2006) 555.
[28] Ding N. and Wang Y. G., *Chin. Phys. Lett.*, **24** (2007) 2434.
[29] `http://www.nber.org/papers/t0342.pdf`.
[30] Gaffeo E., Gallegati M. and Palestrinib A., *Physica A*, **324** (2003) 117.
[31] Zhang J., Chen Q. and Wang Y., *Physica A*, **388** (2009) 2020.
[32] Podobnik B., Shao J., Njavro D., Ivanov P. Ch. and Stanley H. E., *Eur. Phys. J. B*, **63** (2008) 547.