

Impact factor distribution revisited with graphical representation

Giann-wien Hsu¹ · Ding-wei Huang²

Received: 16 October 2015 / Published online: 19 March 2016
© Akadémiai Kiadó, Budapest, Hungary 2016

Abstract In this work, we propose the graphical representation for the empirical data of the impact factor rank-ordered distribution. The characteristics of the distribution can be directly visualized. Within the subject category of journal citation reports, the impact factor rank-ordered distribution systematically presents a clear evidence of the two-exponent behavior and the S-shaped decrease. The sharp convex decrease is related to the first exponent, which dictates the distribution of lower ranks. The mild concave decrease is related to the second exponent, which dictates the distribution of higher ranks. The relevance of Matthew effect is discussed.

Keywords Impact factor · Rank-order distribution · Scaling · Curvature · Matthew effect

Introduction

The impact factor counts the citation rate of articles published in a journal. Scholarly journals are often ranked by their impact factor, which has been widely recognized as an indicator to the journal's prestige. The detailed shape of impact factor rank-ordered distribution is substantially an important issue in informetric research and has been investigated. To our knowledge, most previous studies focused on the abstract mathematical formulation. In this work, we propose using the concrete and graphical representation to present the empirical data.

Recently the two-exponent law for the impact factor distribution has been proposed (Mansilla et al. 2007). The two-exponent law can be taken as an extension to the Lavalette law, where the two exponents are symmetrical. When one of the exponents is neglected,

✉ Ding-wei Huang
dwhuang@cycu.edu.tw

¹ General Education Center, National Tainan Junior College of Nursing, Tainan, Taiwan

² Department of Physics, Chung Yuan Christian University, Chungli, Taiwan

the distribution reduces to the Zipf law. The empirical data from several disciplines can be successfully described (Campanario 2010; Mishra 2010). As the dataset is limited and the data are not smooth enough, different fitting schemes lead to slightly different values of the exponents (Brzezinski 2014). Further mathematical properties have been explored (Egghe 2009; Sarabia et al. 2012). Besides the impact factor distribution, the two-exponent law has also been applied in other rank-ordered distributions from a large amount of phenomena (Martínez-Mekler et al. 2009). In this work, we propose a graphical representation to demonstrate the two exponents visually. This new graph has a balanced representation in the both ends of the distribution. Different mathematical formulations can be visually distinct. We explore this graphical representation systematically over a range of different disciplines.

We also propose another graphical demonstration to resolve a recent controversy on the detailed shape of the impact factor distribution (Egghe 2011). By definition, the impact factor rank-ordered distribution is monotonically decreasing. Only few journals achieve large impact factors. The distribution shows a steep descent in the first few ranks and becomes gentle as the ranks further increase. Previous studies all agreed that the distribution was convexly decreasing in the lower ranks. However, different opinions were proposed in regards to the distribution tendency in the higher ranks. The controversy lies in how the distribution approaches the end. Some researchers (Egghe 2009; Egghe and Waltman 2011) argued that the distribution turned concavely decreasing as it approached the end. The entire distribution was first convex and then concave, i.e., an S-shaped decrease. In contrast, other researchers (Guerrero-Bote et al. 2007; Lanchó-Barrantes et al. 2010) argued that there was no concavely decreasing. The distribution was convexly decreasing entirely. Without resorting to the mathematical manipulations, we propose a simple graph to reveal the curvature of the rank-ordered distribution. Our result confirms the S-shaped distribution and shows that convexness and concaveness can be distinguished unambiguously.

In the following, the graphical demonstration for the two-exponent behavior is presented in “[Modified Zipf plot to demonstrate the two exponents](#)” section. Graphical demonstration for the S-shaped decrease is presented in “[Scaled plot to demonstrate the S shape](#)” section. “[Discussions](#)” section discusses the relationship between the S shape and the two exponents. We suggest that both the two-exponent behavior and the S-shaped decrease can be understood as the manifestation of the Matthew effect.

Modified Zipf plot to demonstrate the two exponents

The conventional Zipf plot is a log-log plot of impact factor f_i versus rank i , where the index i runs through all n journals. A power law distribution becomes a straight line in the Zipf plot, which emphasizes to display vividly the data points in the lower ranks. Owing to the nature of a logarithmic scale, the data points in the higher ranks merge together and the resolution is poor. In the modified Zipf plot, we propose to plot half of the data points in the conventional way, i.e., f_i versus i , with index i running through 1 to the median rank \bar{n} . For the other half of the data points, we plot the reverse ranks, i.e., f_i versus $(n + 1 - i)$, with index i running through the median rank \bar{n} to n . A few examples are shown in Fig. 1. The two-exponent distribution is displayed as a simple triangle. The upper side is dictated by the exponent a of a power law as $f_i \sim i^{-a}$. The lower side is dictated by the exponent b of another power law as $f_i \sim (n + 1 - i)^b$. The two exponents can be observed directly as

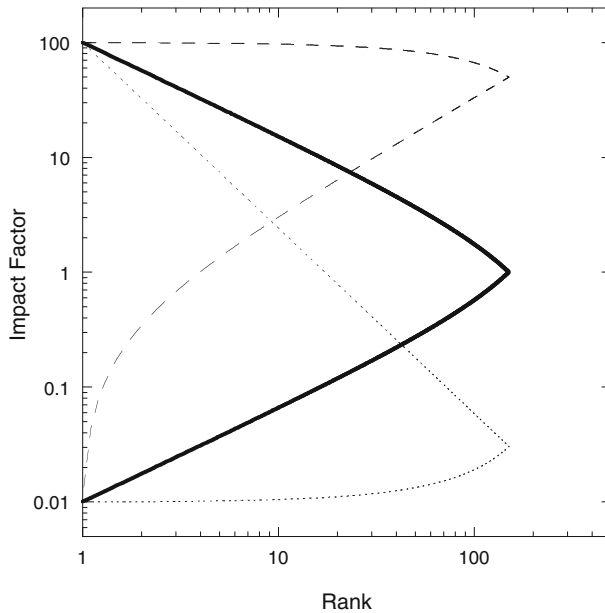


Fig. 1 Modified Zipf plot of three typical distributions. The parameters are $f_{\max} = 100$, $f_{\min} = 0.01$, and $n = 300$. **Bold solid line** shows the Lavalette law. **Dotted line** shows the Zipf law. **Dashed line** shows the linear relation

the slopes of the two sides. When $a = b$, the Lavalette law becomes an isosceles triangle. When $b = 0$, the distribution reduces to the Zipf law. Figure 1 also shows the results from a linear relation (dashed line) and a power law with a negative exponent (dotted line). The linear relation can be taken as a result of random distribution. With fixed maximum f_{\max} and minimum f_{\min} of impact factor, the random distribution distorts toward a large median \bar{f} . And the Zipf law distorts toward a small median. These three distributions can be easily distinguished.

The impact factor stands basically for a journal’s average citation counts. Therefore it might not give a fair judgement across different research fields. In this work, we examine the impact factor rank-ordered distribution within each of the subject categories of journal citation reports (JCR) in the year 2011. To have sufficient statistics, we consider only those subject categories consisting of more than 100 journals. We select the top five and the bottom five subject categories in both science (SCI) and social science (SSCI). The selected twenty categories are listed in Table 1. There are 3556 journals in total. The highest impact factor 101.78 is obtained by the journal *CA: A Cancer Journal for Clinicians* in the SCI subject category (*Oncology*). The lowest impact factor 0.005 is obtained by the journal *The Naval Architect* in the SCI subject category (*Engineering, Civil*).

Data from these twenty categories are shown in Fig. 2. The triangular shape is obvious for all data, especially for the subject categories with high impact factor. For the upper side of the triangle, the smallest exponent is observed in the SSCI subject category (*Public, Environmental and Occupational Health*). For the lower side of the triangle, the smallest exponent is observed in the SCI subject category (*Mathematics*). None of the distributions conform to the Zipf law, nor the linear relation. The superior of two-exponent distribution is obvious. Yet the deviation from the Lavalette law can also be discerned. The slope in the

Table 1 List of the twenty subject categories with their median impact factors \bar{f} , journal numbers n , and the fitting values of the two exponents (a , b)

	\bar{f}	n	a	b
SCI High Impact Factor				
<i>Cell Biology</i>	3.278	179	0.57	0.73
<i>Immunology</i>	2.992	138	0.71	0.67
<i>Biochemistry and Molecular Biology</i>	2.863	287	0.50	0.69
<i>Neurosciences</i>	2.749	244	0.52	0.74
<i>Oncology</i>	2.569	192	0.76	0.80
SCI Low Impact Factor				
<i>Mathematics</i>	0.561	289	0.41	0.39
<i>Engineering, Civil</i>	0.681	118	0.43	0.82
<i>Mathematics, Applied</i>	0.724	245	0.44	0.51
<i>Engineering, Mechanical</i>	0.747	121	0.61	0.75
<i>Veterinary Sciences</i>	0.821	142	0.36	0.95
SSCI High Impact Factor				
<i>Psychology, Clinical</i>	1.387	110	0.46	0.75
<i>Psychiatry</i>	1.490	115	0.51	0.62
<i>Public, Environmental and Occupational Health</i>	1.278	131	0.34	0.60
<i>Management</i>	1.183	168	0.39	0.80
<i>Business</i>	1.138	112	0.41	0.84
SSCI Low Impact Factor				
<i>Linguistics</i>	0.518	158	0.42	1.06
<i>Political Science</i>	0.618	147	0.38	0.87
<i>Education and Educational Research</i>	0.714	205	0.41	0.75
<i>Law</i>	0.760	135	0.43	0.95
<i>Economics</i>	0.779	320	0.45	0.80

lower side of the triangle is always steeper than that in the upper side. For all data, the upper side of the triangle seems to have a universal exponent around $a \sim 0.5$. In contrast, the exponent in the lower side of the triangle assumes different values around the range $0.5 < b < 1$. And the subject categories with low impact factor have a larger exponent, i.e., $b \sim 1$. Judged by visual inspection alone, the two-exponent law can be expected to provide a good framework to describe the empirical data. The modified Zipf plot provides an easy way to estimate the values of the two exponents. The fitting values of a and b are listed in Table 1.

Scaled plot to demonstrate the S shape

The argument for an S-shaped distribution involves the curvature, which is the second derivative of the curve function. In a strict sense, however, the rank-ordered distribution is not a mathematical function. We propose to determine the curvature in a graphical demonstration without involving any mathematical task.

The rank-ordered distributions before the median rank are shown in Fig. 3, with both axes scaled. The rank is scaled by the number of journals. The impact factor is scaled by

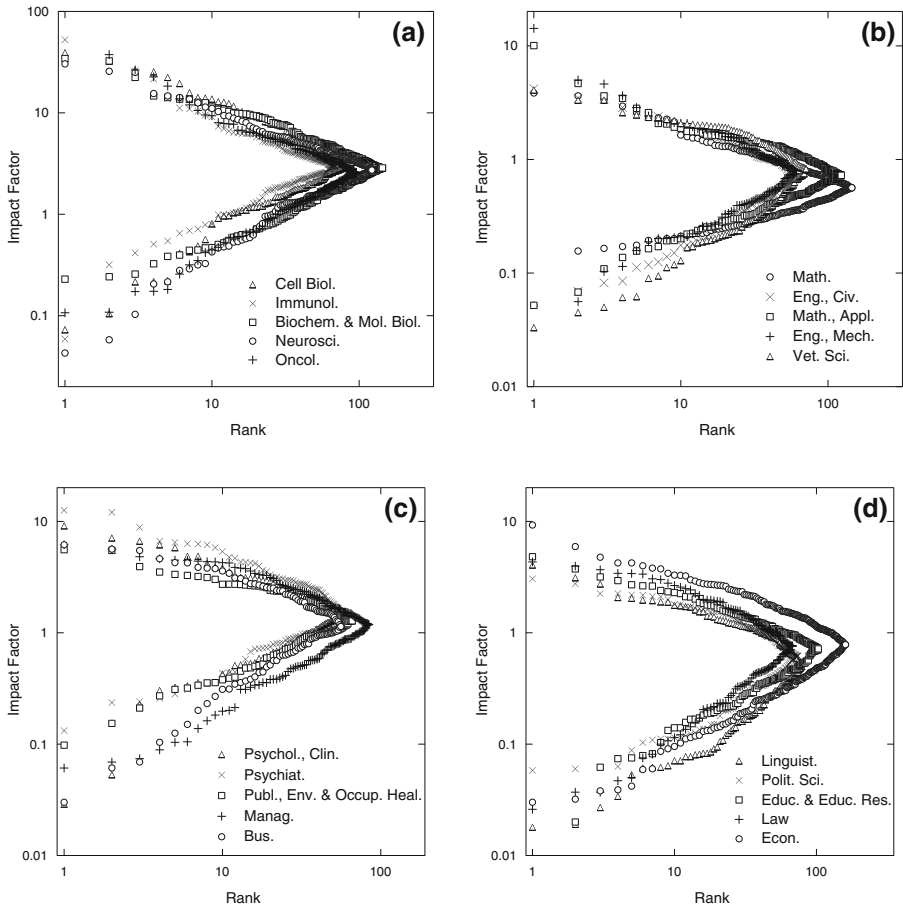


Fig. 2 Modified Zipf plot for each subject category: **a** SCI high impact factor; **b** SCI low impact factor; **c** SSCI high impact factor; **d** SSCI low impact factor

the difference between the maximum and median impact factors (with a shift of the median impact factor). In contrast to the conventional plot of f_i versus i , we plot $(f_i - \bar{f}) / (f_{\max} - \bar{f})$ versus (i / n) for $i = 1, \dots, \bar{n}$. In the figures, there are two fixed points: the maximum impact factor at $(0, 1)$ and the median impact factor at $(0.5, 0)$. A grey straight line connecting these two fixed points shows the linear decreasing. If the distribution is convexly decreasing, the data lie completely beneath the grey line. On the contrary, if the data lie above the grey line, the distribution must turn concavely decreasing in accordance. As shown in Fig. 3, all the twenty distributions are convexly decreasing.

The rank-ordered distributions after the median rank are shown in Fig. 4. The impact factor is now scaled by the median impact factor. We plot (f_i / \bar{f}) versus (i / n) for $i = \bar{n}, \dots, n$. The median impact factor is fixed at $(0.5, 1)$ and the minimum impact factor at $(1, 0)$. Similarly, a grey straight line connecting the two fixed points shows the linear decreasing. The concaveness and convexness can be revealed in a similar way. Distributions for the top five subject categories in science (SCI) are shown in Fig. 4a. All data lie above the grey line. Unambiguously, all distributions are concave. Distributions for the

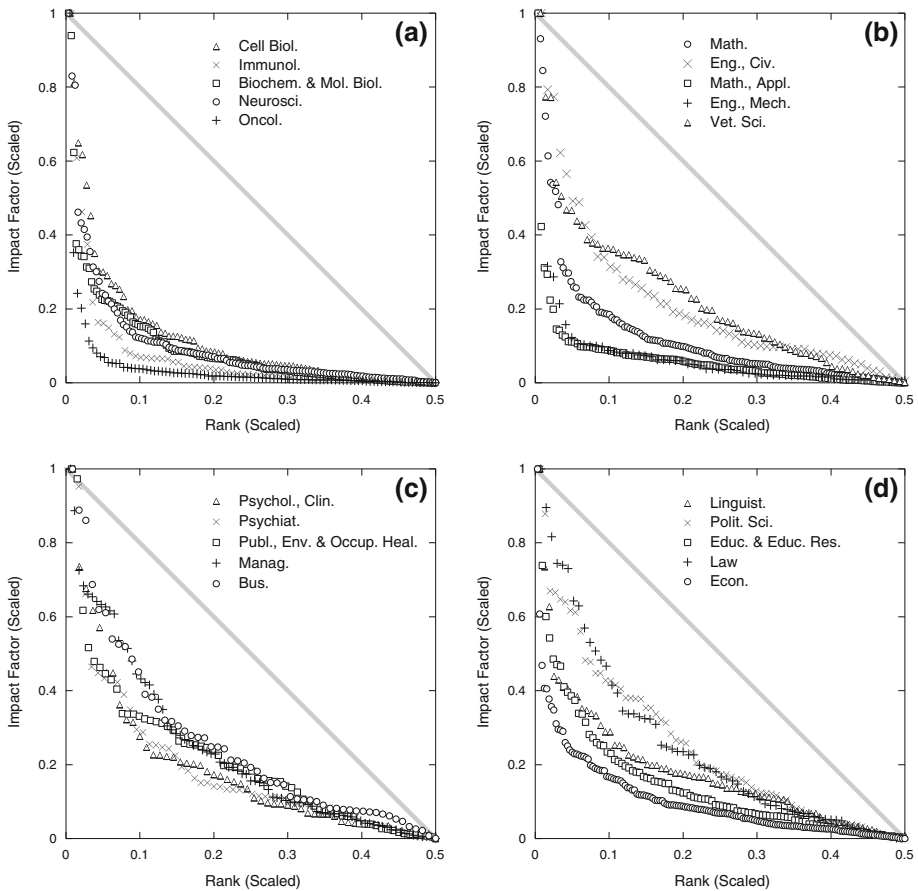


Fig. 3 Scaled distribution in lower rank for each subject category: **a** SCI high impact factor; **b** SCI low impact factor; **c** SSCI high impact factor; **d** SSCI low impact factor

bottom five subject categories in science (SCI) are shown in Fig. 4b. One of the distributions (*Veterinary Sciences*) is clearly below the grey line. Two of the distributions (*Civil Engineering* and *Mechanical Engineering*) show the onset of the grey line above at a scaled ranking of 0.65, which results in a shortened range of concaveness. The other two distributions (*Mathematics* and *Applied Mathematics*) present a much larger deviation toward the end. As a result, the distributions turn concave only near the scaled ranking of 1. In summary, nine of the ten distributions show concave decreasing, which can be established more easily in the subject category with high impact factor.

Compared to science (SCI), most of the journals in social science (SSCI) have lower impact factor. Distributions for the top five subject categories in social science (SSCI) are shown in Fig. 4c. Basically the data show concave as in Fig. 4a. One distribution (*Management*) shows a delay of onset, where the data follow the grey line until the scaled ranking of 0.6. In Fig. 4d, we plot the distributions for the bottom five subject categories in social science (SSCI). In contrast to other distributions, these data follow the grey line more closely. One distribution (*Linguistics*) lies entirely below the grey line. Two distributions (*Political Science* and *Law*) completely follow the grey line. The other two

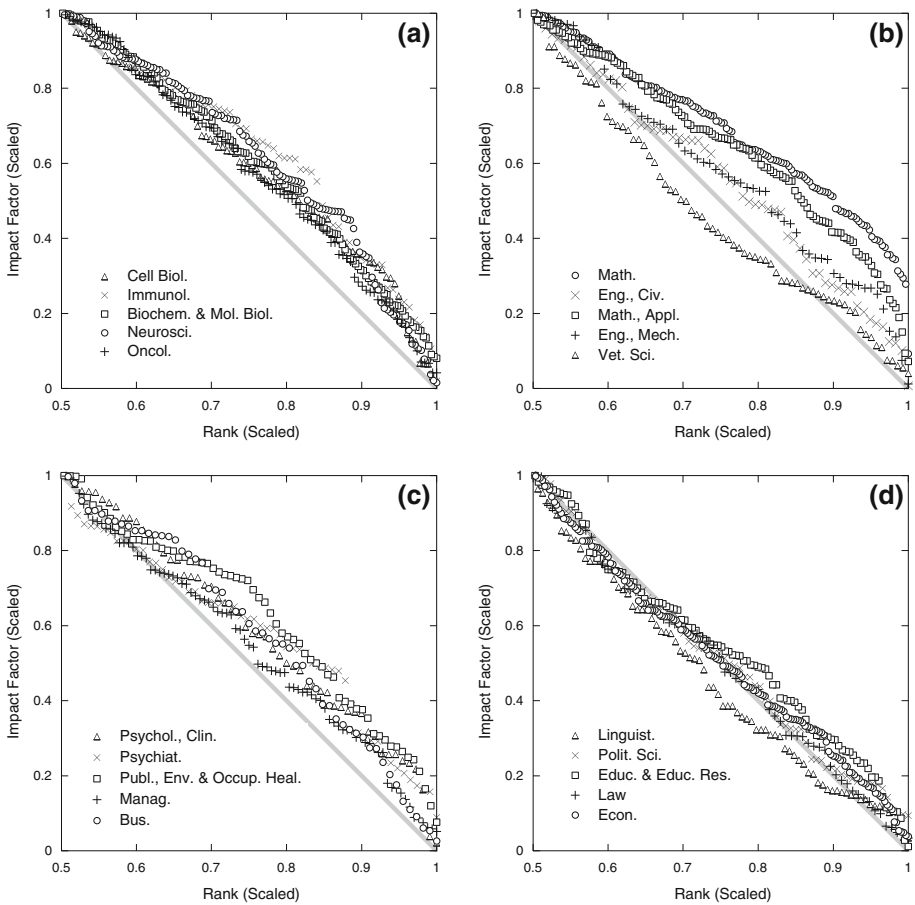


Fig. 4 Scaled distribution in higher rank for each subject category: **a** SCI high impact factor; **b** SCI low impact factor; **c** SSCI high impact factor; **d** SSCI low impact factor

distributions (*Education and Educational Research* and *Economics*) show a significant delay of onset. The data rise above the grey line after the scaled ranking of 0.7. In conclusion, the five distributions with high impact factor show concave decreasing clearly; while, the other five distributions with low impact factor seem to be consistent with a linear decrease.

In summary, the impact factor rank-ordered distribution presents a sharp convex decreasing followed by a mild concave decreasing. The S-shaped distributions are confirmed by the graphical demonstration.

Discussions

We show that the characteristics of impact factor rank-ordered distribution can be unambiguously revealed by the graphical representation. Within a subject category of journal citation reports (JCR), the distinct character of two exponents can be observed

systematically and directly. For the two-exponent law, the median impact factor is roughly equal to the geometric average of the maximum and the minimum. In contrast, the median in a random distribution is equal to the arithmetic average of the maximum and the minimum. For the Zipf law, the median is approximately equal to the minimum in the order of magnitude. The distribution presents a power-law behavior both in the lower ranks and the higher ranks. The two exponents are different and can be viewed as the slopes of the two sides in a triangle. In general the higher ranks have a larger exponent, i.e., the slope of the lower side is steeper. The power-law behavior can be related to the Matthew effect, which refers to a saying from the Gospel of Matthew in the Bible: “*For everyone who has will be given more, and he will have an abundance. Whoever does not have, even what he has will be taken from him.*” Although the effect is known as *the rich get richer and the poor get poorer*, often only the first half of the saying is used in the so-called cumulative advantage. When the number of ranks is large, the conventional Zipf law focuses on the distribution of the lower ranks. Data in the higher ranks are either incomplete or uninteresting. In the case of impact factor distribution, the number of ranks is limited, e.g., $110 < n < 320$ in this study. The so-called cumulative disadvantage can be as effective as the cumulative advantage. Thus the two-exponent law can be a balanced manifestation of the Matthew effect.

Similarly the S-shaped distribution can also be revealed directly by the graphical demonstration. The impact factor rank-ordered distribution presents a sharp convex decrease followed by a mild concave decrease. For those subject categories with high impact factor, the evidence is more clear. For those research fields with lower impact factor, the concave decrease in the higher ranks can still be observed in some categories. While the others are consistent with a linear decrease, i.e., the convex decrease in the lower ranks becoming gentle in the higher ranks. In this study, the majority of the twenty distributions show the concave decrease in the higher ranks. If the data are mixed together, the trend will become less evident. The convex trend in the lower ranks will be dominant by the subject categories with high impact factor; while the concave trend in the higher ranks will be dominant by the subject categories with low impact factor. In previous studies, the distribution for *Mathematics* suggested to provide the most clear evidence of concave decrease. Within all the subject categories in science (SCI) in this study, *Mathematics* has the lowest impact factor. However, we find that the disciplines with high impact factor should provide a much more clear evidence of concave decrease. This puzzle can be addressed as follows. The distribution of *Mathematics* in Fig. 4b does present the largest deviation from the grey line. However, the large deviation has been pushed toward the end of the distribution. Thus the concave decrease can only be observed around very near the end of the distribution. In practice, the distribution has a particularly steep descent for the last two data points. If these two points are excluded, the concavity apparently becomes much more weakened.

With mathematical manipulations, the S shape can be related to the two exponents of the distribution. With a naive interpretation of functional $f(i) \equiv f_i \sim i^{-a}$ in the lower ranks, the curvature obtained by the second derivative gives $a(a+1)$. For all exponent $a > 0$, the curvature is positive and the curve is convex decreasing. Similarly the functional becomes $f(i) \equiv f_i \sim (n+1-i)^b$ in the higher ranks. The curvature becomes $b(b-1)$. For exponent $0 < b < 1$, the curvature is negative and the curve is concave decreasing. Around the threshold of $b = 1$, the curve is linearly decreasing as shown in Fig. 4d. With $a \sim b$, the sharp convex $a(a+1)$ and the mild concave $b(b-1)$ are self-evident from the formulation. The S-shaped distribution can be considered as a consequence of the two-exponent law. Both the two exponents and the S shape are the manifestation of the Matthew effect.

References

- Brzezinski, M. (2014). Empirical modeling of the impact factor distribution. *Journal of Informetrics*, 8, 362–368.
- Campanario, J. M. (2010). Distribution of ranks of articles and citations in journals. *Journal of the American Society for Information Science and Technology*, 61, 419–423.
- Egghe, L. (2009). Mathematical derivation of the impact factor distribution. *Journal of Informetrics*, 3, 290–295.
- Egghe, L. (2011). The impact factor rank-order distribution revisited. *Scientometrics*, 87, 683–685.
- Egghe, L., & Waltman, L. (2011). Relations between the shape of a size-frequency distribution and the shape of a rank-frequency distribution. *Information Processing and Management*, 47, 238–245.
- Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gómez-Crisóstomo, R., & Moya-Anegón, F. (2007). The iceberg hypothesis: import-export of knowledge between scientific subject categories. *Scientometrics*, 71, 423–441.
- Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). The iceberg hypothesis revisited. *Scientometrics*, 85, 443–461.
- Mansilla, R., Köppen, E., Cocho, G., & Miramontes, P. (2007). On the behavior of journal impact factor rank-order distribution. *Journal of Informetrics*, 1, 155–160.
- Martínez-Mekler, G., Martínez, R. A., del Río, M. B., Mansilla, R., Miramontes, P., & Cocho, G. (2009). Universality of rank-ordering distributions in the arts and sciences. *PLoS One*, 4, e4791-1-7.
- Mishra, S. K. (2010). A note on empirical sample distribution of journal impact factors in major discipline groups. Available at Social Science Research Network. <http://ssrn.com/abstract=1552723>.
- Sarabia, J. M., Prieto, F., & Trueba, C. (2012). Modeling the probabilistic distribution of the impact factor. *Journal of Informetrics*, 6, 66–79.