Study on A new Ratio and Negative Exponential Curve of Sci-Tech Progress in China

Xiufeng Li & Shouhe Chen*

Agricultural Information Institute, Chinese Academy of Agricultural Sciences

ABSTRACT: Sci-tech achievements largely represent the Science and technology progress. The ratio on the number of the research papers divided by the number of the sci-tech achievements in the same field (It is simply called 'the Ratio') is proposed as a new concept in this article, which represents each sci-tech achievement needs how many research papers. We find that the Ratio shows the law of the negative exponential descending curve with the time series, by quantitative measures of the time series data of 8 industries in China using the SAS software. The Ratio is predicted to reach a steady value at a certain stage. We analyze the degree of difficulty and speed of the development of different industries, in line with the diversity of the Ratios of different industries.

KEYWORD: A new Ratio; Negative exponential curve; Sci-tech achievements; China

1 INTRODUCTION

Price published a series of papers and books to demonstrate the law of exponential growth of scientific development, since 1950(e.g.1956, 1961, 1963). Scholars have conducted in-depth studies and discussions on the topic, and the growth of journals, articles, authors, R&D expenditure, international collaboration and so on in many scientific fields always demonstrates an exponential increase.

Nalimov V. V. & Nul'chenko Z. M. published the paper More on the concept of exponential growth (in Russian) in *Nauchno-tekhnicheslcaya informatsiya* in 1969 (Vlachý,1978). Ware(1973) developed a statistical exponential model for characterizing the growth patterns of data base utilization and for estimating future utilization levels of demand for information retrieval organizations. Shestopal & Burman, two former Soviet information scientists, put forward a more "general model" on the basis of Price's research in 1978(Su & Han, 1998). Tague et al.(1981) studied the evidence, implications and forecasts of the Law of exponential growth. Rousseau (1994) proposed two double exponential models to describe the first-citation or response distribution and had fitted real data to these models and provided some rationales why these distributions might be appropriate. Qin(1994) investigated the number of collaborative papers corresponded to an exponential increase over time though the philosophical transactions 1901-1991. Egghe(1997) used an exponential aging model to prove that the

Price Index is a function of the mean reference age and also a function of the median reference age(found experimentally by Glänzel & Schoepflin, 1995). Su and Han(1998) derived a "literature variable exponential growth model" from Price's literature growth model F(t)= a e^{bt} through replacing *bt* by a polynomial of degree n-l and showed that the new model was more convincing. Jin & Rousseau(2005) studied the Chinese publications in Web of Science of 1991–2003, and showed that the number of Chinese publications had exponential growth, at the same time, international collaboration publications also had exponential growth. International collaboration publication output has shown exponential growth aroused by the growth of science(He, 2009; Leydesdorff & Sun, 2009). Egghe (2010) used an exponential literature growth model to prove both regularities found in the paper of Larivière et al. (2007) (The first is that the mean as well as the median reference age increases in time. The second is that the Price Index decreases in time).

However, there is relatively little research on negative exponential model in the field of literature study. Bernal, an English physical scientist, proposed a classical citation distribution model ——Bernal negative exponential aging model at 1958 and Brookes proposed an improved Bernal model——the negative exponential model of cumulative citation distribution at 1970(Yu & Li, 2009). Katz (1994) investigated that the frequency of research collaboration between domestic universities in the

United Kingdom, Canada and Australia decreases exponentially with the distance separating the research partners.

In China, there is no uniform definition of sci-tech achievements. The Chinese Academy of Sciences provides sci-tech achievements refer to the creative results possessing certain academic significance or practical value achieved through observation, research and development or dialectical thinking activities focusing on a specific science and technology research subject.

The Ministry of science and technology of P.R. China founded the Website of National Achievements of Science and Technology (NAST, http://www.tech110.net/) in 1999 which is informative and authoritative, and covers national science and technology information service network. It includes sci-tech achievements in all regions and industries from the mainland of China since 1978 affirmed by provincial and ministerial-level governments. Sci-tech achievements largely represent the Science and technology progress. The change of the number of annual industrial sci-tech achievements directly reflects the science and technology progress of each industry in China.

According to the Website of NAST, China's sci-tech achievements consist of basic theory achievements, applied technology achievements, and soft science achievements. The completion of a sci-tech achievement would take certain time, and researchers publish papers to report phase progresses during that period. Accordingly, there must have some kind of relationship between the number of sci-tech achievements and the number of papers.

We attempt to explore the relationship between the development of science and technology and the number of research papers, by analyzing the relationship between the number of the sci-tech achievements and the number of the research papers in the same field. We have not found the similar research papers at home and abroad yet.

2 A NEW CONCEPT AND THE DATA

For convenience of the analysis, we propose a new concept ——i.e. The ratio on the number of the research papers divided by the number of the sci-tech achievements in the same field(It is simply called 'the Ratio') in this article, which represents each sci-tech achievement needs how many research papers, similar to the scientific productivity proposed by Lotka (1926) and one scientific paper requiring how many authors illustrated by Lotka's Law. The Ratio can be marked with letter Y, which shows the number Y of the research papers is required to support 1 sci-tech achievement. If the number of the research papers and the number of sci-tech achievements are marked with letter z and f respectively, a formula can be expressed as follow:

$$Y = z / f \qquad (1)$$

The data of sci-tech achievements of China come from the Website of NAST, while the data of research papers of China are retrieved from China Academic Journal Full-text Database (CAJFD) (http://epub.cnki.net/kns/brief/result.aspx?dbPrefix=CJFQ) of China National Knowledge Infrastructure (CNKI). The CNKI was launched in 1999 and built the most comprehensive system of China academic knowledge resources. CAJFD is the most comprehensive, full-text database of Chinese journals in the world, which contains more than 9,800 journals in the mainland of China and 10 Series databases, and continuous scholarly articles in all research fields since 1979.

Because the classification between the Websites of NAST and CAJFD is not completely consistent, it is difficult to make quantitative analyses for all industries. We choose 8 industries whose classification is similar as the research samples: they are agriculture, crop farming, forestry, animal husbandry, fishery, medical and health industry, petroleum processing and chemical industry, mining industry. And we retrieve the annual numbers of sci-tech achievements and the annual numbers of journal research papers in 1980-2010 as data samples, and then calculate the corresponding Ratio.

3 METHOD AND DATA ANALYSES

3.1 *Very high correlation coefficients between the number of sci-tech achievements and the number of research papers*

We calculate that the number of research papers is highly related to the number of the sci-tech achievements in the same field, and the correlation is highly significant (Table 1), by using the SAS software.

Table 1. The correlation coefficients and significance level between the number of sci-tech achievements and the number of research papers in each industry.

	agriculture	crop farming	forestry	animal husbandry	fishery	Medical and health industry	Petroleum Processing and chemical industry	mining industry
correlation coefficient	0.86999	0.87567	0.85883	0.9023	0.8691	0.9163	0.7876	0.8379
significance level	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

3.2 *The Ratio shows the law of the negative exponential descending curve*

The Ratio of each industry generally displays a negative exponential curve of descending over time, as shown in Figure1 (taking agriculture as an example).
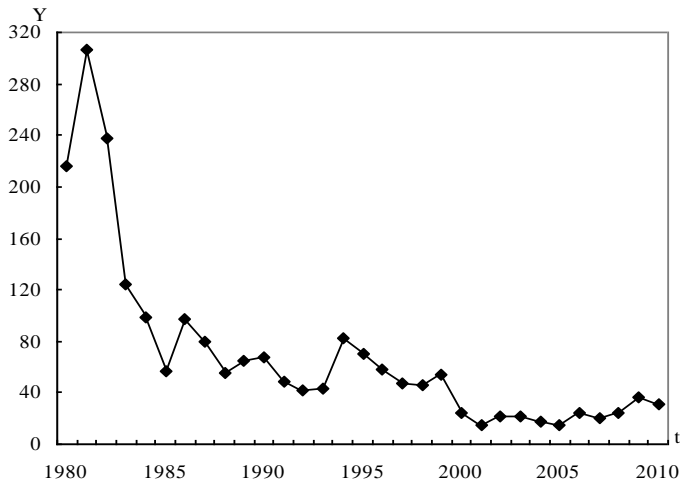

Figure 1. The trend of curve of the Ratio for agriculture .

Let Y be the Ratio as the explained variable, and t is the age as the explanatory variable, then the regression model can be set as follows:

$$Y(t)=a*e^{-b(t-1979)} \quad (a>0,b>0) \tag{2}$$

where a is the coefficient, b is the aging coefficient, and the value of t is the year of 1980, 1981,…,2010.

The regression is done by SAS software, to calculate the estimated values of coefficients a and b and the significance level for each industry (specific values shown in Table 2). The fitting effects are well, and all have passed the test of significance.

General programming of SAS software as follows:

```
title 'exponential:y=a*(exp(-b*(t-1979)))';
proc nlin   data=sasuser.idustryname;
parms a=0 b=0;
model y=a*(exp(-b*(t-1979)));
run;
```

Note: the data of the 8 industries are calculated separately, and the 'idustryname' in the programming is each of the industry's name successively.

Table 2. The fitting results of the 8 industries.

		agriculture	crop farming	forestry	animal husbandry	fishery	medical and health industry	Petroleum Processing and chemical industry	mining industry
a	Estimate	274.3	273.1	645.0	194.6	178.5	634.2	106.6	138.6
b	Estimate	0.1285	0.1638	0.2337	0.0660	0.0652	0.0841	0.0633	0.0747
	F Value	119.99	136.75	15.71	181.69	54.51	163.48	79.98	54.56
	Pr > F	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	Adjusted R Square	0.8885	0.9008	0.5035	0.9235	0.7883	0.9157	0.8412	0.7828

3.3 *The change of the Ratio and empirical analysis*

The negative exponential curve decreases gradually with the increase of time t, and reaches to the flat state. In theory, if the t is unlimited, Y will eventually converge to zero, and the mathematical expression is

$$\lim_{t \to +\infty} Y（t）=\lim_{t \to +\infty} a*e^{-b(t-1979)} = 0 \tag{3}$$

which obviously does not correspond to the fact. As predicted by Price(1963) — the exponential growth could not last forever, so would the negative exponential curve, and Figure 2 shows that the curve indeed starts to level off. Therefore, the reasonable situation is that: Y could tend to be in a stable condition, and will no longer follow the negative exponential function model, when t reaches a certain stage, then Formula (1) becomes Y = z / f = c, —— i.e. z = f*c (c is a natural number) ——that is, there is a stable linear relationship between the number of research papers z and the number of sci-tech achievements f. This means that a subject will be in a relatively mature state of development without new technology-driven, and the law of exponential model of the literature growth will switch to linear model (Wang et al.,1994; Larivière, V. et al.,2008), then the Ratio stabilizes.
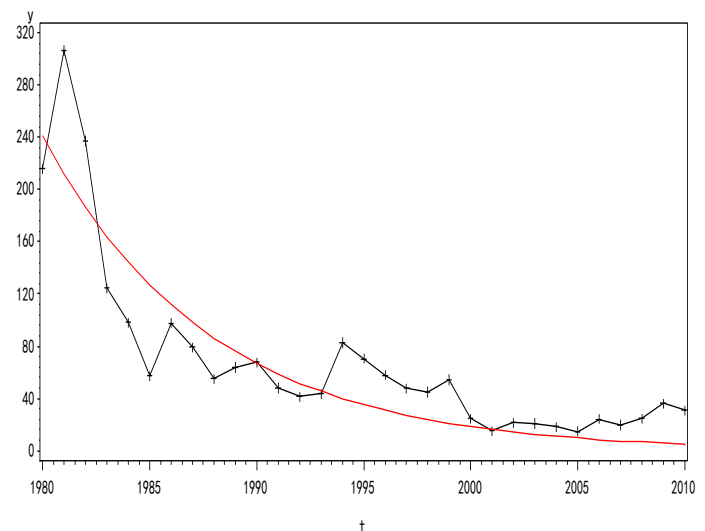

Figure 2. The fitting curve of the Ratio for the agriculture; the smooth line: y=274.3*(exp(-0.1285*(t-1979))).

It can be considered that the curve of the Y will stabilize when the absolute value of the slope s of each curve is less than 1. After calculation, Forestry reached this condition in 2001, and the theoretical value of the Y was about 4 at that time(note: due to the lowest fitting degree of the Forestry Function, and the accuracy is poor, so it may be not so early in fact). The Medical and Health Industry will be the latest to stabilize in 2027, and the theoretical value of the Y will be about 11. The specific theoretical values of the 8 industries are shown in Table 3.

Table 3. The theoretical values of the Ratio of the 8 industries starting to stabilize.

	agriculture	crop farming	forestry	animal husbandry	fishery	medical and health industry	Petroleum Processing and chemical industry	mining industry
$s^* < 1$ starting year	2008	2003	2001	2019	2018	2027	2010	2011
c^{**}	7	6	4	14	14	11	15	13

Note: *:$\Delta t = 1$, the slope of the curve 's' is considered as $| (\Delta Y)/(\Delta t) |$ instead;

   **:The values of 'c' are identified as the natural numbers, and it is integral part of the average of the value between the starting year and the previous year.

3.4 The diversity analyses of the Ratios of different industries

3.4.1 The diversity of the number of research papers required to support the sci-tech achievements

The value of coefficient a is the intercept of the exponential function curve in the Y axis, so coefficient a can be thought as the theoretical Ratio in the initial year (i.e. 1979 in this article)—— that is, the number a of the research papers is required to support 1 sci-tech achievement of each industry in the initial year. The diversity of the coefficient a suggests that the number of research papers required to support 1 sci-tech achievement in the initial year for different industry is different, as showed in Table4.

Table 4. The theoretical value of research papers required to support 1 sci-tech achievement in the initial year for different industry.

agriculture	crop farming	forestry	animal husbandry	fishery	medical and health industry	Petroleum Processing and chemical industry	mining industry
274.3	273.1	645.0	194.6	178.5	634.2	106.6	138.6

The higher the number is in table 4, the relatively fewer sci-tech achievements of the subjects in the initial stage of development of the industry, and its development is more difficult. The comparison between the Ratio Y of different period and the value c of its stable state, can measure the degree of difficulty for the development of the industry, and the degree can be marked as the letter d, so it can be mathematically represented as

$$d = \frac{c}{Y} \times 100\% \tag{4}$$

What's more, if d is greater, the Y is closer to the stable state, the development becomes more difficult.

3.4.2 The diversity of the descending speed of the Ratio curve

The negative exponential fitting function curve shows that, all these Ratio curves of various industries have the descending tendency. Because the coefficient of exponential function a and the coefficient of the explanatory variable b together determine the degree of inclination of the curve, the difference of a and b indicates that the different degrees of inclination of the negative exponential fitting function curves of various industries—— that is, the descending speeds of the curves are different when these curves tend to the stable state, as shown in Table 5 and Figure 3.

Table 5. The diversity of the absolute value of the Ratio fitting curves' slope in different times for the industries.

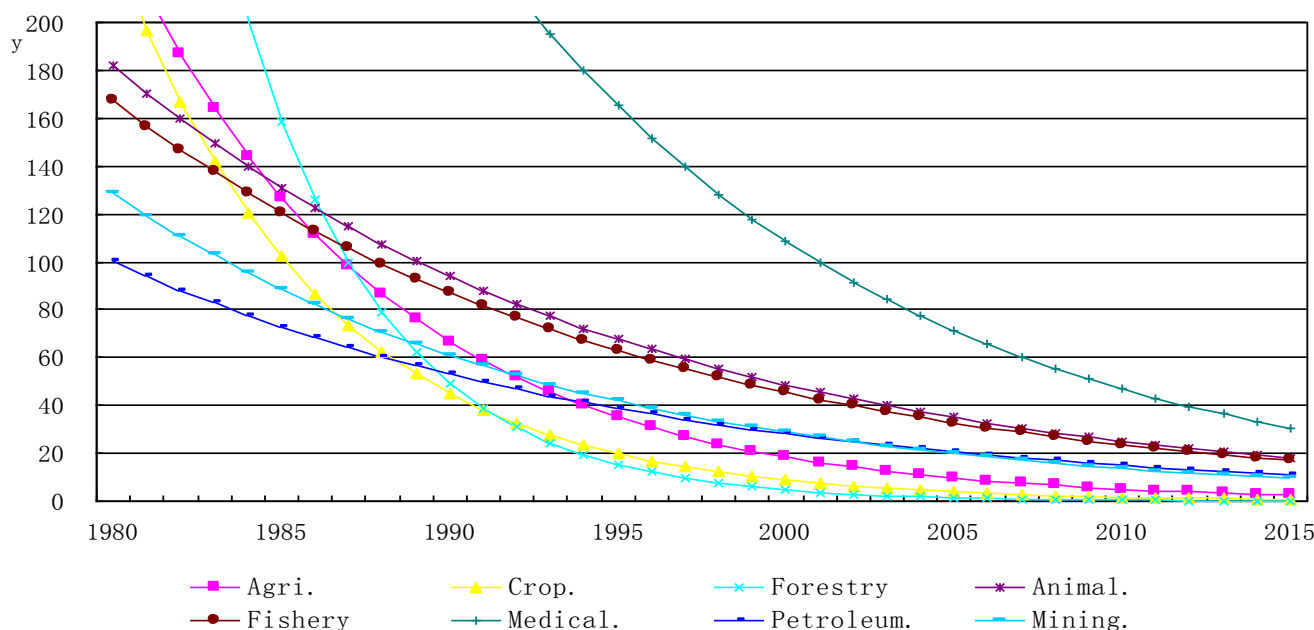|	Agriculture	Crop farming	Forestry	Animal husbandry	Fishery	Medical and health industry	Petroleum processing and chemical industry	Mining industry
1985	16.30	16.74	37.09	8.64	7.87	32.20	4.62	6.61
1995	4.51	3.25	3.58	4.47	4.10	13.89	2.45	3.13
2005	1.25	0.63*	0.35*	2.31	2.14	5.99	1.30	1.48

Note: *according to Table 4, the 2numbers are invalid.

Figure 3. The comparison of the Ratio fitting curves for the industries.

Note: The range of the values of 'a' is too large for the industries, so in order to show the diversity of the curves, parts of the curves are omitted.

During the same time period, if the absolute value of the curve's slope is greater in the Figure 3, the curve descends faster——i.e. the value of the Ratio Y decreases fast, which suggests the increase in the number of the sci-tech achievements is faster than that of the research papers, hence, the science and technology of the subjects in the industry develops quickly, and these subjects will take the lead more easily before reaching the mature stage.

4 RESULTS AND DISCUSSION

We calculate that the number of research papers is highly related to the number of the sci-tech achievements in the same field. We find that the Ratio Y shows the law of the negative exponential descending curve with the time series, and the Ratio will reach a steady value at a certain stage, by quantitative measures of the data of 31 years of 8 industries in China. The regression functions calculate the annual theoretical results of the Ratio of the 8 industries. And we determine the theoretical values c of the Ratio of the 8 industries starting years to stabilize, which have some certain reference and guiding significance. The stabilization of the Ratio can be tentatively understood that "the growth rule of science and technology size is the coordinate growth between science and technology and economy in essence" demonstrated by Sun(2012).

We introduce d to measure the degree of difficulty for the development of each industry, according to the diversity analyses of the Ratios of different industries. We ascertain the industry in which the science and technology develops quickly and will take the lead more easily before reaching

the mature stage, in the light of the descending speeds of the curves.

Zhou & Leydesdorff (2006)proved the emergence of China as a leading nation in science, and China has become a major player in the production of scientific papers. Now most fields of science and technology in China catch up the level of developed countries basically, what's more, a few fields reach the international advanced level(Chen, 2013). Glänzel et al.(2006) showed that "China is challenging the leading sciento-economic powers and the time is approaching when this country will represent the world's second largest potential in science and technology". Accordingly, we may deduce that the relationship on the time series data between Chinese research papers and sci-tech achievements, can reflect the relationship between science and technology progress and the number of papers in the world.

It is to be noted that, the data are not as satisfactory as possible because of the limitations. The regression results will have more explanatory power, if the increase of the number of industries and the improvement of data can be made. In addition, we consider the law has international universal significance, since the similar data of sci-tech achievements and research papers of other countries do not collected yet in this article, further empirical research needs to be done.

5 CONCLUSIONS

Sci-tech achievements largely represent the science and technology progress. We explore the relationship between the development of science and

technology and the number of research papers in China. We propose a new concept as 'The ratio on the number of the research papers divided by the number of the sci-tech achievements in the same field', which represents each sci-tech achievement needs how many research papers. The Ratio shows the law of the negative exponential descending curve with the time series. What's more, the Ratio is predicted to reach a steady value at a certain stage. We also analyze the degree of difficulty and speed of the development of different industries.

Although the data is not perfect, we regard that the findings are advisable and significant. We can judge if an industry or subject is a leading one or in its mature stage, and we may evaluate the degree of difficulty in development, on the basis of the numerical values Y, c and d, some measures can be taken accordingly. There are a lot work to do about this topic in the future.

ACKNOWLEDGMENTS

REFERENCES

[1] Chen L. China Became the Country of Science and Technology Power and Innovation Power. Science and Technology Daily. October 12th 2013(in Chinese).

[2] Egghe, L. 1997. Price index and its relation to the mean and median reference age. *Journal of the American Society for Information Science* 48(6): 564–573.

[3] Egghe, L. 2010. A model showing the increase in time of the average and median reference age and the decrease in time of the Price Index. *Scientometrics* 82(2):243–248.

[4] Glänzel, W. & Schoepflin, U. 1995. A bibliometric ageing study based on serial and non-serial reference literature in the sciences. In M. E. D. Koenig & A. Bookstein (Eds.), *Proceedings of the fifth biennial conference of the ISSI*:177–185. River Forest, IL: Rosary College.

[5] Glänzel, W., Debackere, K. & Meyer, M. 2006 'Triad' or 'Tetrad'? On China's New Global Role in Science and Technology. In: *Proceedings of The 2nd Conference of Chinese science and technology policy and Management and International Academic Forum on science of science and Scientometrics*. Da Lian, China.

[6] He, T. 2009. International scientific collaboration of China with the G7 countries. *Scientometrics* 80(3):571–582.

[7] Jin, B.H. & Rousseau, R. 2005. China's quantitative expansion phase: exponential growth but low impact. In:

Ingwersen, P. & Larsen, B. (eds.), ISSI 2005: *Proceedings of the 10th International Conference on Scientometrics and Informetric*: 362–370. Stockholm: Karolinska University Press.

[8] Katz, J. S. 1994. Geographical proximity and scientific collaboration. Scientometrics, 31(1): 31-43.

[9] Larivière, V., Archambault, É. & Gingras, Y.2008. Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology* 59(2):288–296.

[10] Leydesdorff, L. & Sun, Y. 2009. National and international dimensions of the triple helix in Japan: University-industry-government versus international coauthorship relations. *Journal of the American Society for Information Science and Technology* 60(4):778–788.

[11] Lotka, A.J. 1926. The frequency distribution of scientific Productivity. *Journal of the Washington Academy of Science*16(12):317-323.

[12] Price, D.J. 1956. The exponential curve of science. *Discover* 17(6): 240-243.

[13] Price, D.J.1961. *Science Since Babylon*, New Haven: Yale University Press.

[14] Price, D.J.1963. *Little science, Big science*. New York: Columbia University Press.

[15] Qin, J. 1994. An investigation of research collaboration in the sciences through the philosophical transactions 1901–1991. *Scientometrics* 29(2): 219-238.

[16] Rousseau, R. 1994. Double exponential models for first-citation processes. *Scientometrics*, 30: 213-227.

[17] Su, Y. & Han L.F. 1998. A new literature growth model: Variable exponential growth law of literature. *Scientometrics* 42(2):259-265.

[18] Sun, X.J.2012. Analysis of the "exponential growth" paradox and "coordination growth" rule for science and technology size. *Studies in Science of Science*30(6):813-819 (in Chinese).

[19] Tague, J., Beheshti, J. & Rees-Potter, L. 1981. The Law of Exponential Growth: Evidence, Implications and Forecasts. *Library Trends* 30:125-149.

[20] Vlachý, J. 1978. Frequency distributions of scientific performance a bibliography of Lotka's law and related phenomena. *Scientometrics* 1(1):107-130.

[21] Wang, Q.l., Liu, W. & Yang Q.L.1994. Research on Modeling of Growth Regularity of the Number of Literature. Xiamen: *Proceedings of Chinese control and decision Symposium in 1994*(in Chinese).

[22] Ware, G.O.1973. A general statistical model for estimating future demand levels of data-base utilization within an information retrieval organization. *Journal of the American Society for Information Science* 24: 261–264.

[23] Yu, G. & Li, Y.J. 2010. Identification of referencing and citation processes of scientific journals based on the citation distribution model. *Scientometrics* 82(2): 249–261.

[24] Zhou, P. & Leydesdorff, L. 2006. The emergence of China as a leading nation in science. *Research Policy*35(1):83–104.