DO COUNTRIES WITH LOWER SELF-CITATION RATES PRODUCE HIGHER IMPACT PAPERS? OR, DOES HUMILITY PAY?

Klaus Jaffe

SUMMARY

Studying the 62 countries producing most papers reported in the SCImago data base in the period 1996-2008, it was found that countries with low per capita publication numbers show variable rates of self-citations and produce publications with lower citation impact. In contrast, countries with larger numbers of citations per paper have also high per capita publication numbers and their researchers appear to be humbler, showing lower rates of country and author self-citations. Notable exceptions are China, USA and Iran, which show abnormally high country self-citation rates, partially explained respectively by large populations, large total number of publications and language barriers. An increase of self-citation rates in almost all countries during the last decade, calls for exploring science policies that increase international scientific impact, such as more international cooperation, and science education with broader outlooks.

Introduction

"The principle of science, the definition, almost, is the following: The test of all knowledge is experiment. Experiment is the sole judge of scientific 'truth'. But what is the source of knowledge? Where do the laws that are to be tested come from? Experiment, itself, helps to produce these laws, in the sense that it gives us hints. But imagination is also needed to create from these hints the great generalizations -to guess at the wonderful, simple, but very strange patterns beneath them all, and then to experiment to check again whether we have made the right guess" (Feynman, 1964). These are fine words, but often, the amazing intellectual creation of humans: Science, is handicapped by this very same imagination. Many scientists believe so strongly in the dictates of their own mind that they shun reality (Jaffe, 2010). This is especially evident when comparing soft with hard sciences regarding falsifiability (Fanelli, 2010) or skepticism (Jaffe *et al*., 2010); where in the softer sciences, subjective imagination is more frequently in better regard than experiment.

The balance between objective experiment and subjective imagination can vary enormously, inviting the question of whether an optimal balance for scientific creativity between these two elements exists. Attitudes favoring general creativity are strongly correlated with economic development, and include tolerance and openness to other ideas (Florida, 2005). Comparing specific attitudes in different countries showed that scientific and economic development correlated with attitudes favoring skepticism, tolerance and openness (Jaffe, 2005). Thus science seems to be favored when subjective imagination is kept in check.

Self-citation is a necessary tool in normal scientific publishing. Yet the degree to which authors cite their own work (author self-citation) or cite works performed by re-

KEYWORDS / Country Citation / Impact / Productivity / Scientific Publications / Self-Citation /

Received: 05/20/2011. Modified:08/22/2011. Accepted: 08/24/2011.

Klaus Jaffe. Chemist, Universidad Simón Bolívar (USB), Venezuela. Ph.D. in Biology, University of Southhampton, UK. Professor, USB, Venezuela. Address: Departamento de Biología de Organismos, USB. Apdo. 89000, Caracas 1080, Venezuela. email: kjaffe@usb.ve

¿PAÍSES CON MENORES TASAS DE AUTOCITACIÓN PRODUCEN PUBLICACIONES DE MAYOR IMPACTO? O ¿PAGA SER HUMILDE?

Klaus Jaffe

RESUMEN

El estudio de los 62 países de mayor producción científica del mundo reportados en la base de datos de SCImago para el período 1996-2008 revela que los países con un bajo número de publicaciones per cápita muestran tasas variables de autocitas de otras publicaciones de sus países y producen publicaciones que reciben un bajo número de citas. En contraste, los países con mayor número de citas por artículo, también muestran un elevado número de publicación per cápita y sus investigadores parecen ser más humildes, mostrando una menor tasa de autocitas de autor y de país. Excepciones notables son China, EE.UU. e Irán, que muestran tasas muy altas de auto-citación del país, lo que se explica en parte por su gran población, su gran número total de publicaciones y barreras del idioma, respectivamente. Un aumento importante de la auto-citación en la última década en casi todos los países, llama a explorar políticas científicas que busquen aumentar el impacto científico internacional, tales como mayor cooperación internacional y una educación de la ciencia con perspectivas más amplias.

PAÍSES COM MENORES TAXAS DE AUTOCITAÇÃO PRODUZEM PUBLICAÇÕES DE MAIOR IMPACTO? OU, VALE A PENA SER HUMILDE?

Klaus Jaffe

RESUMO

O estudo dos 62 países de maior produção científica do mundo relatados na base de dados de SCImago para o período 1996-2008 revela que os países com um baixo número de publicações per capita mostram taxas variáveis de autocitas de outras publicações de seus países e produzem publicações que recebem um baixo número de citações. Em contraste, os países com maior número de citações por artigo, também mostram um elevado número de publicações per capita e seus investigadores parecem ser mais humildes, mostrando uma menor taxa de autocitações de autor e de país. Exceções notáveis são a China, EE.UU. e o Irã, que mostram taxas muito altas de autocitações do país, o que se explica em parte por sua grande população, seu grande número total de publicações e barreiras do idioma, respectivamente. Um aumento importante da autocitação na última década em quase todos os países, é um convite a explorar políticas científicas que busquem aumentar o impacto científico internacional, tais como maior cooperação internacional e uma educação da ciência com perspectivas mais amplas.

---

searchers in their own country (country self-citation) varies widely. Self-citations have various functions and what scientometrics can extract from statistics tracking self citations is complex and variable (Aksnes, 2003). Part of this variation might be related to attitudes of scientists concerning openness to the ideas of other scientists, self-valuation and tolerance, i.e humility. Here, this possible relationship is explored in some detail by comparing self-citation rates in different countries.

Methods

One dimension of the balance between the subjective and the objective is the openness towards worldviews regarding the balance between world-wide interests and personal, local or national interests. Here it is assumed that this openness is at least partially reflected in a greater citation rate of scientific research performed by others, regardless of the country of origin. This openness in citations is by definition negatively correlated with country and author self-citation rates (the ratio between self-citations to total citations in a document). Data for self-citation (Table I) was calculated from over 19.5×10^6 documents indexed by SCImago (2007) of the Universidad de Granada, Spain, based on academic publications from nearly 18,000 titles compiled by Scopus (2010). The rate of country self-citations was calculated by dividing the total number of country self-citations, according to the country of residence of the corresponding author given by SCImago, by the total number of citations for each country. This self-citation rate was compared with other indices such as number of publications per capita and average citation rates of documents for each country. Only data for the 62 countries with more than 10000 publications, as recorded by Scopus between 1996 and 2008 are presented. Data for gross domestic product (GDP) and population size for each country are from the World Bank (2009). Other data are from references as cited in the text. All data can be accessed freely at the websites of Scopus and the World Bank. Statistical analysis was performed using the commercial software *Statistica 8* (StatSoft, 2007).

Data from Scopus was checked for their robustness by comparing the statistical correlation between GDP and number of publications with those calculated from reports by the Science Citation Index and by Google Scholar. No statistically significant differences between these correlations could be found. Part of these comparisons has been published elsewhere (Jaffe, 2005).

Results and Discussion

Country self-citation

Figure 1 shows that low rates of country self-citation correlated with high per capita publication numbers and high citation impact (citations per document) of the publications. This trend was statistically highly significant (Spearman correlation coefficient= -0.322, and probability of rejecting the null hypothesis $p<0.01$) when comparing rates of self-citations with the number of publications per

inhabitants, a still larger correlation index of -0.438 ($p<0.0003$) was found for country self-citations and number of total citations received by the country. It was found that countries with low numbers of scientific publications per capita show higher rates of self-citations and produce publications with lower citation impact. At the same time, countries with widespread scientific activity, indicated by large numbers of publications per capita, seem to be humbler and tend to show lower rates of self-citations. Although countries with low numbers of scientific publications per capita show in average higher rates of self-citations, many exceptions exist. For example, Latin American countries (bold names in Figure 1) showed large variations in country self citation rates, with Brazil and Argentina using much higher self citation rates than expected from the regression, whereas Colombia and Venezuela showed much lower rates.

The exceptions to this rule were the very high self-citation rates in China, USA and Iran. These oddities are not due to statistical fluctuations, as self-citation rates were calculated based on over 1.2 and 4.3×10^6 documents for China and the USA respectively. These results can be understood if we take into account the total number of publications per country and assume that countries with larger absolute academic productivity will have larger country self-citation rates as it will be more likely by pure chance that cited documents are from the same country. The data show a positive correlation between the total number of country self-citations and total number of citable documents (correlation coefficient= 0.58, $p<0.0001$). Other researchers reported oriented citation bias for the USA, partially explained by the large cumulative citation number for papers originating from this country (Pasterkamp *et al*., 2007). A network analysis of co-authorship (Royal Society, 2011) using the documents compiled by Scopus, revealed a similar special status for the USA. Most countries cooperate with researchers in the USA, but researchers in the USA have a low cooperation rate compared with those from other countries.

The result for China can be explained with a similar correlation between self-citations and population size (correlation coefficient= 0.65, $p<0.0001$). A similar conclusion for China, using a different experimental approach was obtained by Minasny *et al*. (2010). The result for Iran has no such explanations, but could be due to language (Moed, 2005), which limits Iranian researchers to publish and cite in local journals (Biglu, 2007).

A multiple regression study with country self-citation rates as the dependent variable is presented in Table II. It confirms that self-citation rates are explained by several factors, of which at least three can be identified here. The total number of documents produced in a country is a strong predictor for country self-citations, followed by population size of the country. Yet the relationship between self-citation rates and the number of citations per document published is also a strong factor identified by these statistics.

Self-citation rates increase over time. Comparing the average self-citation rates in the 62 countries in 1996 to those in 2009 showed that the mean self-citation rate in 2009 was 28.9% higher than in 1996. This difference was statistically highly significant (t-test for dependent samples: t= 9.8, $p>>0.0001$), and suggests that newer publications everywhere are decreasing their international citation rates. Only countries with a relatively small scientific output, i.e. Cuba, Argentina, Latvia, Venezuela, Jordan and Georgia, decreased their country self-citation rates in that period.

Author self-citation

Does country self-citation differ from author self-cita-

TABLE I
DATA USED IN THE PRESENT ANALYSIS ORDERED BY COUNTIES FROM EUROPE, ASIA, AMERICA AND AFRICA*

Country	Population in millions (2009)	Nr. of documents in Scopus	Document per 1000 inhabitants	Nr. of selfcitations	Citations per 1000 inhabitants	Rate of self citation
Norway	4.8	94,617	20	215,909	14.55	0.1745
Poland	38.2	209,744	5	373,845	6.91	0.2838
Portugal	10.6	72,826	7	137,362	10.49	0.2177
Romania	21.5	42,320	2	45,957	5.24	0.2470
Russia	141.8	405,499	3	577,757	4.61	0.3112
Slovakia	5.4	35,274	7	49,365	6.39	0.2331
Slovenia	2	29,493	15	44,163	7.37	0.2297
Spain	46	449,406	10	1,225,409	11.59	0.2650
Sweden	9.3	250,129	27	708,092	16.95	0.1766
Switzerland	7.7	247,655	32	643,764	19.44	0.1466
Turkey	74.8	171,048	2	259,644	6.37	0.2929
Ukraine	46	74,325	2	75,267	3.46	0.2969
United Kingdom	61.8	1,244,316	20	4,476,611	15.48	0.2482
Australia	21.4	401,930	19	1,101,340	14.05	0.2208
China	1317	1,223,278	1	2,240,814	4.83	0.5176
Hong Kong	7	103,462	15	180,238	10.88	0.1815
India	1140	393,536	0	729,613	6.09	0.3462
Iran	72	68,401	1	109,888	6.57	0.4194
Israel	7.3	154,402	21	340,029	14.75	0.1586
Japan	127.7	1,224,465	10	3,920,215	10.53	0.3139
Jordan	5.8	10,751	2	8,143	5.2	0.1751
Malaysia	27.5	29,166	1	23,912	6.2	0.1944
New Zealand	4.3	80,299	19	160,983	12.91	0.1738
Pakistan	169.7	24,564	0	26,902	4.53	0.3002
Saudi Arabia	25.4	26,763	1	19,320	5.35	0.1433
Singapore	5	82,159	16	109,807	9.91	0.1563
South Korea	48.8	319,976	7	530,243	8.53	0.2391
Taiwan	23	233,763	10	416,490	8.27	0.2582
Thailand	67.8	41,892	1	51,460	8.7	0.1851
Argentina	39.9	73,705	2	144,908	9.12	0.2328
Brazil	190	236,703	1	517,211	8.35	0.3202
Canada	33	630,525	19	1,803,543	15.54	0.2043
Chile	16.6	37,347	2	69,866	11.26	0.1982
Colombia	44.4	14,754	0	15,017	8.79	0.153
Cuba	11.2	15,277	1	17,155	4.68	0.2601
Mexico	107.4	96,625	1	160,114	8.3	0.2287
United States	307	4,318,928	14	35,474,244	18.08	0.4684
Venezuela	28.4	17,580	1	19,204	7.14	0.1657
Algeria	34.4	11,664	0	8,760	5.12	0.2052
Egypt	81.5	47,420	1	55,012	5.62	0.2295
Kenya	38.7	10,026	0	19,244	11.8	0.1821
Morocco	32	15,952	0	16,051	5.48	0.1992
Nigeria	154.7	20,341	0	17,874	4.36	0.2538
South Africa	49.3	71,731	1	133,165	8.97	0.2285
Tunisia	10.4	17,785	2	16,222	5.15	0.2397

*Data extracted from documents compiled by Scopus from 1996 to 2008 as reported in SCImago website.

tions? Author self-citations, "those where authors cite their own works - account for a significant portion of all citations. These self-references may result from the cumulative nature of individual research, the need for personal gratification, or the value of self-citation as a rhetorical and tactical tool in the struggle for visibility and scientific authority" (Fowler and Aksnes, 2007). These authors found among Norwegian scientists that the more one cites oneself the more one is cited by other scholars; yet Aksnes (2003) also reported for this sample that the highest share of self-citations was found among the least cited papers. That is, author self-citation averages for countries show the same negative correlation with average citations per documents as shown here (Schubert *et al*., 2006); although they did not analyze their data that way, when joining data from these authors with the ones presented here we get the relationship shown in Figure 2. Clearly, author self-citations follow the same trend regarding citation impact as country self-citation. A similar correlation was found by Glänzel *et al*. (2004) who found a negative correlation in 50 countries between the share of author self-citations and an index of citation impact, the mean expected citation rate. This trend varies between disciplines (Snyder and Bonzi, 1998; Glänzel and Thijs, 2004), but always a low author self-citations correlates with higher impact factors. This correlation (Figure 2) is highly significant as indicated by a Spearman correlation coefficient (-0.89, $p<0.0001$) between the average author self-citation rate for each country and average citations per document of the 40 countries selected by Schubert *et al*. (2006).

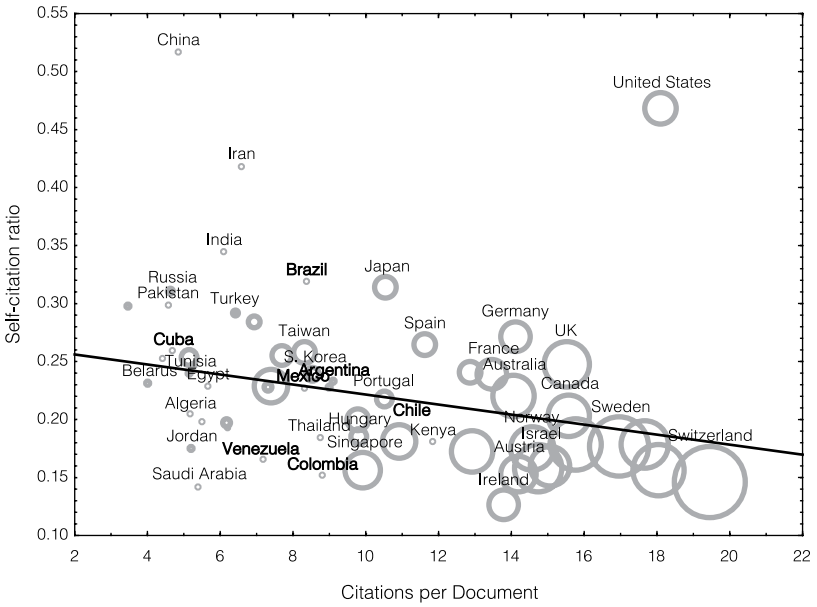

Figure 1. Proportion of total self-citations to total citations of publications, recorded by Scopus for each country between 1997 and 2008, versus the average citation impact (average number of citations per document) of the publications of that country. The size of each sphere is proportional to the number of scientific papers per capita in the country, ranging from 0.13 to 32.16 publications per 1000 inhabitants. (Data from SCImago, 2007 and World Bank Statistics, 2009).

TABLE II
RESULTS FROM A MULTIPLE REGRESSION WITH COUNTRY SELF-CITATION RATES AS DEPENDENT VARIABLE, AS PRODUCED BY STATISTICA 8

Partial correlations	Beta in	Partial	Semipart	Tolerance	R-square	t(57)	p-level
Population	0.365	0.528	0.322	0.781	0.219	4.697	0.000017
Total Documents	0.630	0.706	0.517	0.672	0.328	7.533	0.000000
Documents / Population	0.071	0.066	0.034	0.237	0.763	0.5027	0.617799
Citations / Documents	-0.527	-0.422	-0.241	0.209	0.791	-3.513	0.000874

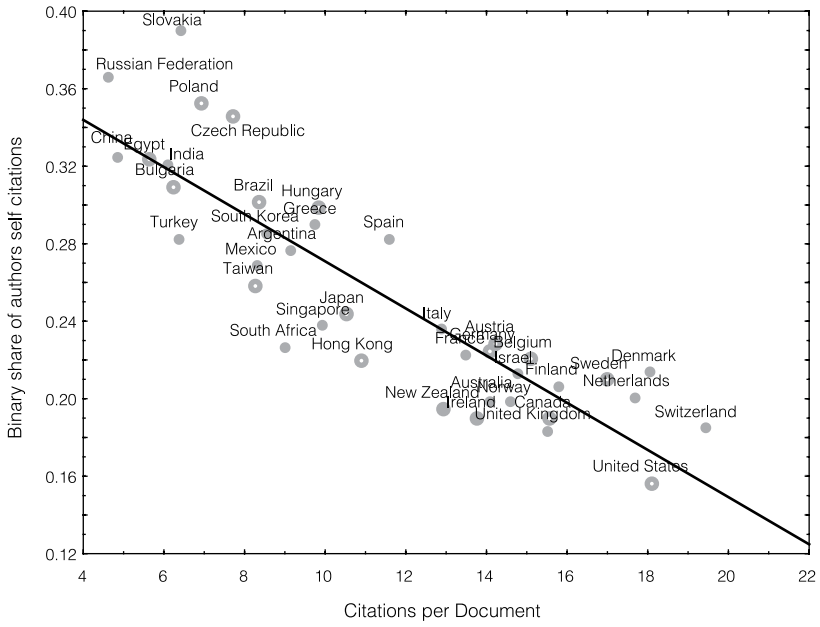

Figure 2. Share of binary author self-citations in all citations of the 40 most active countries in the sciences in 1994-2003 (Schubert *et al*., 2006) versus citation per document of the given country as reported in 1996-2008 (SCImago, 2007).

Conclusion

There is no reason for condemning self-citations in general (Glänzel *et al*., 2006). Self-citations are an integral part of the way we advance in science and a moderate level of self-citations is indicative of consolidated scientific activity in the group and country of the authors. There might be different motivations for author and country self-citations. In addition, the reasons for each might be manifold. High self-citation rates might be explained by relative isolation in research. This fact might explain higher self-citation rates in countries with low density of scientists and or low publication numbers per capita, but not the trends found for country self-citations. Globalization has made research more of an international enterprise (Royal Society, 2011), and citations normally cover work produced in many different countries. Thus other explanations for this trend can not be dismissed. It can be suggested that high rates of self-citation produced by low numbers of non-self-citations might be expected from scientists with a relative lack of interest in the scientific activity of others. This last suggestion is supported by the fact that both, author self-citation and country self-citation correlate negatively with the scientific impact (citation rates) of the paper. Citation rates are modulated by a variety of factors (Bornmann *et al*., 2008) but somehow reflect

levels of visibility among the scientific community, although not necessarily the quality (Arnold and Fowler, 2011), which in turn will affect the influence the work might have on future research.

Statistical analysis has shown clearly that large numbers of country and author self-citation rates are negatively correlated with the average citation impact of the scientific publications produced in a country. This result complements the study by The Royal Society (Royal Society, 2011) that showed that scientific success is correlated with international cooperation. That is, less self-citation rates and more international cooperation produce better science. Correlation is no proof of causation, but everyday experience in science supports a common trend between international cooperation, empathy and citation impact. Country self-citations are increasing nowadays, despite a centrifugal dispersion of scientific activity from the USA and Europe to other countries as reported by Royal Society (2011) and Schubert and Glänzel (2006) among others. These facts are relevant to science policy and science education. If more cooperation and more out-looking attitudes favor higher impact science, then the trend towards increased country self-citations detected here does not seem encouraging. This fact calls for science policies that specifically address this shortcoming. For example, more support for international cooperation might improve country self-citation rates, cooperation statistics, innovation and, eventually, better science.

If self-citation rates are at least partially related to the balance of openness *vs* selfishness; broad world views *vs* local views; humility *vs* arrogance, then the results presented here are consistent with the view that more openness, broader outlooks, and larger doses of humility correlate with higher impact scientific productivity. Folk wisdom would have summarized the relationship reported here as 'arrogance increases ignorance'. Further research may throw more light into these relationships. Further work might clarify these relationships in the future.

ACKNOWLEDGEMENTS

The author thanks Wolfgang Glänzel and Rodolfo Jaffe for helpful comments on earlier versions of the manuscript.

REFERENCES

Aksnes DW (2003) A macro study of self-citations. *Scientometrics 56*: 235-246.

Arnold DN, Fowler KK (2011) Nefarious numbers. Notices AMS 58: 434-437.

Biglu MH (2007) Tendency to the self-citation among journals in Iran and Turkey. *Inf. World 8*: 297-309.

Bornmann L, Mutz R, Neuhaus C, Daniel HD (2008) Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics Sci. Env. Polit. 8*: 93-102.

Fanelli D (2010) "Positive" results increase down the hierarchy of the sciences. *PLOS One 5*(4) doi:10.1371/journal.pone.0010068

Feynman R (1964) In Richard P. Feynman RP, Leighton RB, Matthew Sands M (Eds.) *The Feynman Lectures on Physics*. 2nd ed. 2005. Addison-Wesley, San Francisco, CA, USA. 560 pp.

Florida R (2005) *The Flight of the Creative Class. The New Global Competition for Talent*. HarperCollins. New York, USA, 326 pp.

Fowler JH, Aksnes DW (2007) Does self-citation pay? *Scientometrics 72*: 427-437.

Glänzel W, Thijs B (2004) The influence of author self-citations on bibliometric macro indicators. *Scientometrics 59*: 281-310.

Glänzel W, Thijs B, Schlemmer B (2004) A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics 59*: 63-77.

Glänzel W, Debackere K, Thijs B, Schubert A (2006) A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics 67*: 263-277.

Jaffe K (2005) Science, religion and economic development. *Interciencia 30*: 370-373.

Jaffe K (2010) *What is Science: An Interdisciplinary Perspective*. University Press of America. Lanham, MD, USA. 120 pp.

Jaffe K, Florez A, Grigorieva V, Masciti M, Castro I (2010) Comparing skills and attitudes of scientists, musicians, politicians and students. *Interciencia 35*: 545-552.

Minasny B, Hartemink AE, McBratney A (2010) Individual, country, and journal self-citation in soil science. *Geoderma 155*: 434-438.

Moed HF (2005) *Citation Analysis in Research Evaluation*. Springer. Netherlands. 353 pp.

Pasterkamp G, Rotmans JI, De Kleijn DVP, Borst C (2007) Citation frequency: A biased measure of research impact significantly influenced by the geographical. *Scientometrics 70*: 153-165.

Royal Society (2011) *Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century*. RS Policy document 03/11. London, UK. http://royalsociety.org/uploadedFiles/Royal_Society_Content/Influencing_Policy/Reports/2011-03-28-Knowledge-networks-nations.pdf

Schubert A, Glänzel W (2006) Cross-national preferences in co-authorship, references and citations. *Scientometrics 69*: 409-426.

Schubert A, Glänzel W, Thijs B (2006) The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics 67*: 503-514.

SCImago (2007) *SJR - SCImago Journal & Country Rank*. (Cons. 11/28/2010) ww.scimagojr.com/countryrank.php

Scopus (2010) www.info.sciverse.com/scopus/about

Snyder H, Bonzi S (1998) Patterns of self-citation across disciplines (1980-1989). *J. Inf. Sci. 24*: 431-435.

StatSoft (2007) *STATISTICA. Data Analysis Software System*. Ver. 8.0. StatSoft, Inc. Tulsa, OK, USA. www.statsoft.com.

World Bank Statistics (2009) http://data.worldbank.org/indicator/SP.POP.TOTL.