The state of research output in South Africa with respect to economy size and population

J. Martin van Zyl

Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa
e-mail: wwjvz@ufs.ac.za

Key Words: Research, citable documents, South Africa, GDP, population size

Summary: The output in research of South Africa is measured in terms of citable documents and it is investigated if this output is in line with what should be expected from a country with the population and economy of South Africa. The expected relationship is calculated by making use of the data of the top 86 countries. It was found that the research output of South Africa is very much in line with what should be expected.

1. Introduction

The state of research of South Africa measured by the total number of citable documents relative to the size of the economy and the population size will be investigated. Citable documents are those published by a journal and are articles, reviews and conference papers. The data for this research were obtained from the Scopus website for international research SJR — SCImago Journal & Country Rank (2007). Data from 1996 to 2010 were used in the calculations to estimate the expected output. The 86 countries included in the analysis are those for which GDP (Gross Domestic Product) and population size were included in The Economist publication, The World in 2011. This set of data was used to estimate the parameters of the expected regression relationship between research output and economic and population size. The focus is more on the broad trends, rather than on small details. Some trends comparing South Africa with research output from other regions will also be given.

The table with the top 50 countries ranked with respect to research documents is given in the appendix. South Africa is ranked 35th with respect to the total number of publications. It was found that GDP is by far the strongest predictor of the research output of a country, and that South Africa performs more or less as would be expected of a country with its economy and population size. South Africa is ranked 29th in terms of GDP, 23rd in terms of population and 51st in terms of GDP per head. The growth in research output is also in line with comparable African countries.

The *h*-index or Hirsch index was introduced by Hirsch (2005) and it measures not only quantity but also the quality of publications, where quality is measured by the number of times a publication was cited. The *h*-index can be used for an individual researcher, but also for a group of researchers, in this case a country.

The cites per document and the *h*-index for the documents published during 1996 to 2010 are 9.19 and 198 respectively compared to 18.88 and 1229 for the USA. This can be seen as a measure of the impact and quality of the individual publications rather than the quantity published. During the period 1996 to 2010 the percentage of self cites for South Africa was 22.29% compared to 46.43% for the USA.

The percentage of documents with international collaboration by South African researchers was 44.76% in 2010. After a large increase in international collaboration during the years after 2000, this percentage stabilized and stayed more or less constant. More on trends in international collaboration can be found in the report by The Royal Society (2011).

According to the THE ranking (Times Higher Education World University Rankings) for the years 2011 to 2012. Africa has four universities in the top 400, and three of which are in South Africa. The top four universities in Africa are the University of Cape Town performing the best with rank 103, Stellenbosch University, University of the Witwatersrand and Alexandria University in Egypt. Matthews (2011) gave a review of the world rankings of South African universities, the effect of rating researchers in South Africa is analyzed by Inglesi-Lotz and Pouris (2011) and the state of science in the South Africa Development Community (SADC) is given in the paper by Pouris (2010). ASSAF (Academy of Science of South Africa) investigated the status of doctoral study in South Africa for the period 2000 to 2007 (ASSAF, 2010).

Much research is conducted using Bibliometrics methods, WSI (Webometrics, Scientometrics and Informetrics) and e-publishing makes it much easier to analyze trends and quality in research output. Some interesting references are the online Research Trends newsletter published by Scopus, the journal Scientometrics, the Journal of Informetrics, the following two papers on trends in citation statistics Adler, Ewing and Taylor (2009), Evans (2008).

2. International research output

South Africa is ranked 35th in terms of total number of publications, 36th in terms of citable documents, 37th in terms of citations and 35th with respect to the *h*-index. The USA dominates international research by far. The ranking, research output details, GDP and population size of the top 5 countries with respect to total number of research documents together with that of South Africa is shown in Table 1 and the ranking of the top 50 countries is given in Table 5 in the appendix.

The mean ranks with respect to total documents of different regions of the top 100 countries are North America (with mean rank 12), Europe (40.18), Asia (48.95), Africa (70) and Latin America (68.57).

Table 1. Research ranking of the top 5 countries and South Africa with respect to total number of documents, 1996 – 2010.

	Country	Documents	Citable documents	Citations	Citations per Document	*h*-index	GDP (Billions US$)	Population (millions)	GDP per head
1	United States	5,322,590	4,972,679	100,496,612	20.18	1229	14996	312.3	48010
2	China	1,848,727	1,833,463	7,396,935	5.66	316	6460	1345	4800
3	United Kingdom	1,533,434	1,392,982	24,535,306	17.42	750	2403	62.7	38360
4	Japan	1,464,273	1,429,881	16,452,234	11.72	568	5621	126.5	44440
5	Germany	1,396,126	1,321,606	20,437,971	15.79	657	3127	83	37680
35	South Africa	93,926	88,329	862,984	10.58	198	346	49	7050

More detailed data about research output of South Africa is given in Table 2 **(**SCImago, (2007)**.**

The United States has 1229 publications with at least 1229 citations (*h*=1229) and South Africa has 198 publications with at least 198 citations (*h*=198), meaning that South Africa has fewer than 198 publications with 1229 citations. Thus the quality of research results of the United States measured by the *h*-index is much higher than the quality of South African research.

The calculation of the *h*-index can briefly be explained as follows: Suppose a researcher has n publications, each cited $C_1 \geq .... \geq C_n$ times. A plot with $C_1,...,C_n$ on the vertical axis against the number of publications *k*, $k = 1,...,n$, on the horizontal axis is made, then h is the smallest integer number of publications such that this number received at least h citations. If a 45° degree straight line is drawn from the origin, then h is the corresponding smallest number of publications on the horizontal axis, where the straight line crosses the plotted line.

The *h*-index has weaknesses, especially when used to rate individual researchers. It depends on the number of publications and thus also the period for which the researcher was active. The number of authors of a publication is not taken into account and there are large differences in the index between subject fields because of differences between impact numbers in subject fields (Baptista, *et al.* (2006), Iglesias and Pecharromán, 2006, Schubert and Glänzel, 2007). The asymptotic properties of the h-index were considered by Beirlant and Einmahl (2010). Barcza and Telcs (2009) gave conditions for power law upper tails of the distribution of the index for journals. Pratelli, Baccini *et al.* (2011) considered statistical properties of the index in smaller samples and applied it to rate top researchers in Statistics and Probability.

Table 2. Research output and data of South Africa, 2000 – 2010.

Year	Documents	Citable Documents	Cites	Cites per Document	Cited Documents	% International Collaboration	% Region Output (Southern Africa)	% World
2000	4385	4109	6564	14.97	3707	30.97	67.89	0.36
2001	4479	4187	61324	13.69	3822	28.73	68.3	0.34
2002	5091	4747	58352	11.46	4186	28.38	70.12	0.37
2003	5443	51	67153	12.34	4456	41.52	67.02	0.38
2004	5934	5556	73789	12.43	4802	45.05	68.44	0.38
2005	6564	6127	7202	10.97	5163	46.3	68.41	0.38
2006	738	6971	66906	9.07	5699	46.07	66.95	0.4
2007	7827	7361	5686	7.26	5873	48.17	65.4	0.41
2008	8802	824	49243	5.59	6149	46.75	67.37	0.44
2009	984	9254	32615	3.31	5802	46.79	67.69	0.47
2010	10832	10078	14889	1.37	4432	44.76	66.58	0.5

It was found that the relationship between the log number of citable documents and the log of GDP and population size is linear. Note that when the word log is used in this document it is meant as the log to base ten, not the natural logarithm. After experimenting, also using regions as dummy variables, it was found that the best multiple regression model is:

$$E(Y \mid X_1, X_2) = 2.2717 - 0.3625X_1 + 1.2479X_2, \qquad (2.1)$$

where Y is the log of number of citable documents, X_1 is the log of population size in millions and X_2 is the log of GDP in billions of US$, where log is taken as the log base ten. Both independent variables are highly significant. This is in line with the theory of production functions of the Cobb-Douglas form with inputs capital and labour.

The standardized regression coefficients are 0.09721 and 0.1071 showing that GDP is a far more important predictor than population size. The mean square error is 0.2195, $R^2 = 0.7286$ and $R^2_{adj} = 0.7218$. The expected number of citable documents base log 10 is 4.827 versus the observed 4.946, thus very close to what is expected. Countries outperforming most with respect to the expectation given their resources are India and Zimbabwe. Some of the underperformers are Angola, Indonesia, United Arab Emirates, Saudi Arabia and Kazakhstan.

The individual relationship between GDP, population size and total number of citable documents was considered.

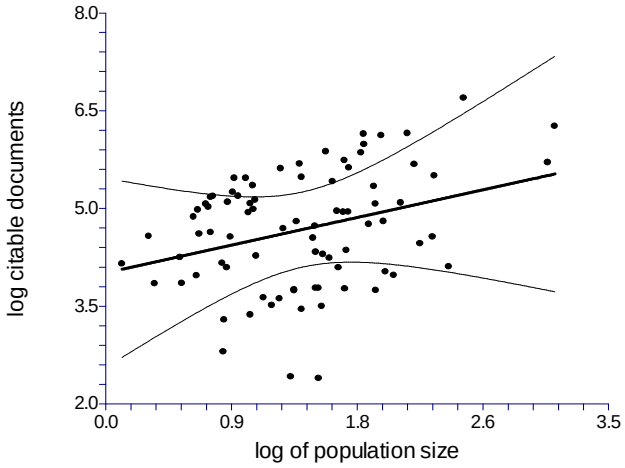

Figure 1. Plot and 95% confidence interval of the log of population versus log of the number of citable documents. Population in millions, log is base ten.

The regression equation is

$$E(Y \mid X) = 4.0151 + 0.4851X\ , \qquad (2.2)$$

where Y is the log of the number of citable documents and X the log of the population size in millions. The data of 83 countries were used, the correlation is 0.3250 and the mean square error is 0.7147. The slope regression coefficient is highly significant with $p < 0.01$. According to this regression South Africa should expect to have 4.8350 citable documents on a log scale, while the observed number is 4.9461, thus in line and better with what would be expected.

Underperformers are Afghanistan and Angola, and the USA is over performing with respect to population size.

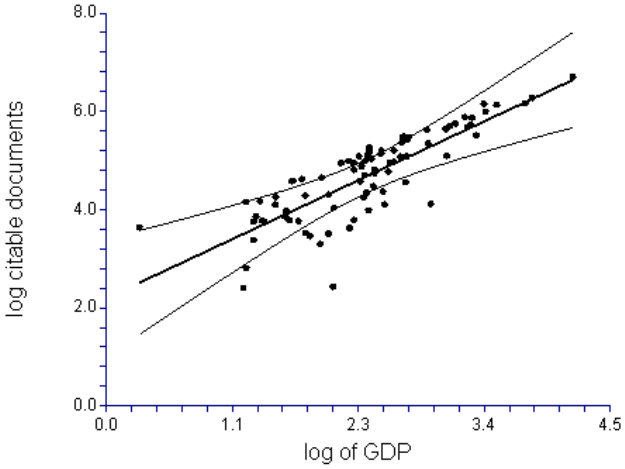

Figure 2. Plot and 95% confidence interval of the log of GDP in billions versus log of number of citable documents. Log is base ten.

The regression equation for the relationship between citable documents and GDP is

$$E(Y \mid X) = 2.1988 + 1.0658X\,, \qquad (2.3)$$

where Y is the log of the number of citable documents and X the log of the GDP in 2011 in billions. The data of 83 countries were used and the correlation is 0.8305. The slope regression coefficient is highly significant with $p < 0.01$. According to this regression South Africa should expect to have 4.9049 (on a log scale) citable documents, while the observed number is 4.9461, thus a little below the expected performance. Underperformers are Afghanistan, Angola, Indonesia while Zimbabwe is especially over-performing with respect to GDP.

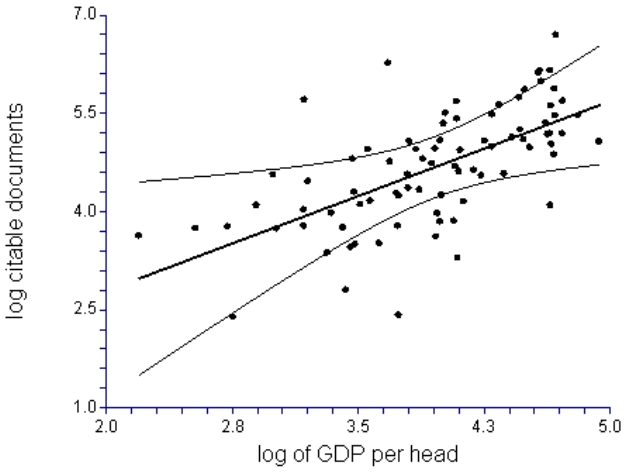

Figure 3. Log of GDP per head versus log of number of citable documents, log base ten.

The regression equation for the relationship between citable documents and GDP per head is $E(Y \mid X) = 0.8567 + 0.9654X$, where Y is the log of the number of citable documents and X the log of the GDP per head in 2011. The data of 83 countries were used and the correlation is 0.6429.

The slope regression coefficient is highly significant with $p < 0.01$. According to this regression South Africa should expect to have 4.9049 citable documents on a log scale, while the observed number is 4.5737. This is below the number that would be expected. Angola is under performing while China and India are over performing with respect to GDP per head.

In the following chart, trends in the number of citable documents of Egypt, Nigeria and South Africa are shown.

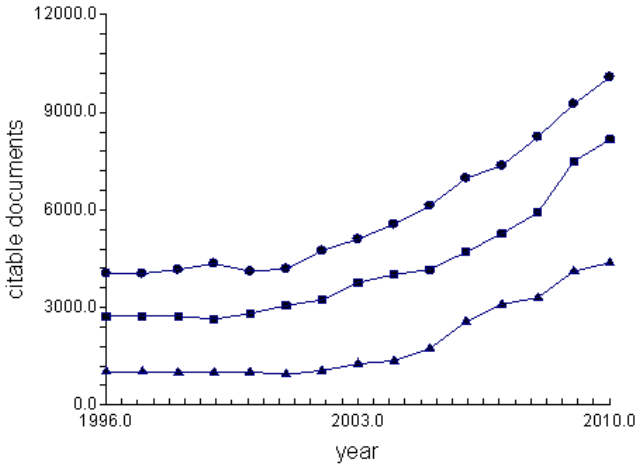

Figure 4. Number of citable documents Egypt (square symbol), Nigeria (triangle symbol) and South Africa (dot symbol).

The trends indicate growth, but in percentage terms the growth of South Africa is the lowest with 149.08%, Nigeria showed 324% growth and Egypt showed 199.56% growth. The data for the plot is given in Table 3.

Table 3. Number of citable documents: Egypt, South Africa and Nigeria, 1996 – 2010.

Year	South Africa	Nigeria	Egypt
1996	4046	1029	2724
1997	4043	1026	2730
1998	4160	1001	2734
1999	4350	997	2641
2000	4109	1016	2814
2001	4187	943	3060
2002	4747	1058	3233
2003	5100	1268	3764
2004	5556	1361	4013
2005	6127	1737	4168
2006	6971	2566	4706
2007	7361	3100	5267
2008	8240	3301	5915
2009	9254	4113	7486
2010	10078	4371	8160

To get an indication of the South African trends versus international trends between 1996 and 2010, a comparison between China, the USA and South Africa is shown in Figure 5.

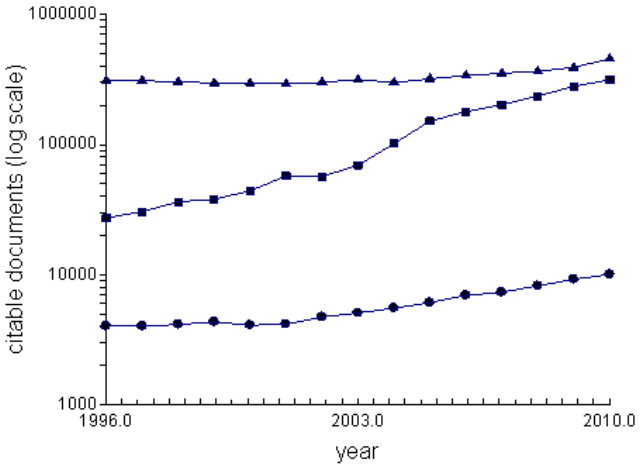

Figure 5. Number of citable documents: China (square symbol), USA (triangle symbol) and South Africa (dot symbol).

Clearly the research output of China is increasing at a much higher rate than the USA or South Africa. It can be mentioned that the expected cross-over where China will start outperforming the USA is between 2013 and 2020 (Leydesdorff (2011), The Royal Society (2011)).

Trends in selected subject areas for the period 2005 – 2010 are shown in Table 4.

Table 4. Number of citable documents published by South Africa in different subject areas sorted by percentage growth over the period 2005 to 2010.

	2005	2006	2007	2008	2009	2010	% Growth
Earth and Planetary Sciences	650	673	665	658	765	749	15.23
Veterinary	124	125	124	133	179	150	20.96
Agricultural and Biological Sciences	1294	1528	1491	1603	1740	1924	48.68
Biochemistry, Genetics and Molecular Biology	535	640	690	748	816	799	49.34
Chemistry	336	354	398	424	447	506	50.59
Mathematics	256	306	350	366	443	394	53.90
Engineering	354	391	487	401	521	546	54.23
Chemical Engineering	175	174	197	188	220	273	56
Physics and Astronomy	335	389	450	466	511	536	60
Environmental Science	513	710	686	710	783	824	60.62
Medicine	1480	1591	1756	2109	2132	2380	60.81
Immunology and Microbiology	381	427	545	547	659	657	72.44
Computer Science	143	202	168	204	278	250	74.82
Psychology	113	154	130	200	172	204	80.53
Energy	59	72	76	84	92	121	105.08
Materials Science	181	256	253	273	407	397	119.33
Social Sciences	476	513	533	683	844	1061	122.89
Arts and Humanities	177	212	177	294	323	413	133.33
Pharmacology, Toxicology and Pharmaceutics	113	111	133	194	200	267	136.28
Economics, Econometrics and Finance	72	91	73	128	136	191	165.27
Business, Management and Accounting	66	79	91	135	170	207	213.63

There is growth in all the subject areas, with especially high growth in output in Arts and Humanities, Social Sciences, Economic and Pharmacology subjects. Some of the areas with slow growth are in the basic scientific subject areas, Biology, Chemistry and Mathematics. One gets the impression that the growth is more in the social sciences, medicine and economics, which may be because of better funding models and incentives for researchers in those fields.

3. Conclusions and recommendations

In terms of quantity, the growth in research output of South Africa is in line with what can be expected. Looking at the number of cites per document and also the *h*-index, it seems that the quality is a problem and much lower than the top performing countries. It can be recommended that more effort should be put in increasing the quality of research, rather than the quantity.

At the moment the money paid to universities by the Department of Education for each publication in an accredited journal is exactly the same for a publication in a local journal with a very low impact number as it is for a top high impact international journal. There are big differences between subject fields with respect to citations and this is related to the number of publications which can be achieved in a specific subject field. This aspect is not taken into account by some universities when research funds are distributed on the basis of published papers.

Because of the emphasis on quantity, there is often pressure on researchers just to get any publication and the focus is on quantity rather than quality. There should be a large incentive for researchers to get publications in top journals, and allow them enough time without pressure to achieve this.

4. Appendix

Table 5. Research ranking of the top 50 countries with respect to total number of documents, 1996 – 2010.

	Country	Documents	Citable documents	Citations	Citations per Document	h-index	GDP (Billions US$)	Population (millions)	GDP per head
1	**United States**	5,322,590	4,972,679	100,496,612	20.18	1,229	14996	312.3	48010
2	**China**	1,848,727	1,833,463	7,396,935	5.66	316	6460	1345	4800
3	**United Kingdom**	1,533,434	1,392,982	24,535,306	17.42	750	2403	62.7	38360
4	**Japan**	1,464,273	1,429,881	16,452,234	11.72	568	5621	126.5	44440
5	**Germany**	1,396,126	1,321,606	20,437,971	15.79	657	3127	83	37680
6	**France**	1,021,041	964,320	14,156,535	15.09	604	2490	63.2	39370
7	**Canada**	790,397	748,787	12,187,113	17.55	580	1616	34.3	47070
8	**Italy**	762,290	720,911	9,861,600	14.45	515	1888	60.3	31320
9	**Spain**	583,554	547,858	6,573,014	13.12	412	1337	46.1	28990
10	**India**	533,006	507,792	3,211,864	7.27	256	1832	1202	1520
11	**Australia**	520,045	485,249	7,083,995	16	450	1190	22.5	52830

12	**Russian Federation**	480,665	476,490	2,456,003	5.21	285	1737	141.5	12280
13	**Netherlands**	435,083	409,982	7,805,760	20.05	509	743	16.6	44630
14	**South Korea**	430,438	422,745	3,344,131	9.82	287	1094	49.6	22050
15	**Brazil**	328,361	318,294	2,409,214	9.57	262	2052	194.9	10530
16	**Switzerland**	309,549	292,254	6,007,936	21.77	506	513	7.9	65050
17	**Taiwan**	308,498	301,775	2,391,691	9.57	229	466	23.2	20040
18	**Sweden**	304,831	292,150	5,410,618	19.09	448	449	9.5	47300
19	**Poland**	265,139	259,850	1,853,462	7.87	258	496	38.1	12310
20	**Belgium**	237,081	224,898	3,621,954	17.1	398	444	10.6	41760
21	**Turkey**	231,178	219,280	1,380,599	7.54	176	760	74	10270
22	**Israel**	186,281	177,814	2,898,025	16.66	368	227	7.7	29410
23	**Austria**	164,308	155,111	2,324,954	16.01	336	376	8.4	44250
24	**Denmark**	162,761	154,612	3,015,221	20.42	373	292	5.6	52320
25	**Finland**	153,964	149,390	2,447,743	17.64	330	229	5.4	42740
26	**Greece**	142,767	135,434	1,350,053	11.34	228	290	11	26350
27	**Hong Kong**	129,792	124,880	1,464,726	12.79	248	220	7.1	30820
28	**Mexico**	125,646	122,268	1,005,002	9.49	201	1119	113.8	9830
29	**Norway**	122,768	116,973	1,749,741	16.63	288	431	5	86740
30	**Czech Republic**	122,379	118,930	942,579	8.82	206	184	10.2	18050
31	**Iran**	120,350	117,469	499,322	7.68	106	488	75.9	6430
32	**Singapore**	109,346	105,665	1,092,233	11.82	218	237	5.2	45200
33	**New Zealand**	101,286	95,295	1,309,197	14.8	247	148	4.4	33490
34	**Portugal**	100,111	96,937	960,473	12.14	199	211	10.7	19810
35	**South Africa**	93,926	88,329	862,984	10.58	198	346	49	7050
36	**Argentina**	93,883	91,056	886,653	10.56	191	375	40.9	9160
37	**Hungary**	89,305	86,438	923,883	11.14	224	127	9.9	12910
38	**Ukraine**	88,707	88,007	344,658	3.98	121	165	45.3	3630
39	**Ireland**	78,892	74,033	974,485	15.56	234	194	4.1	46750
40	**Egypt**	64,565	63,415	367,134	6.79	115	253	86.2	2940
41	**Romania**	63,809	62,975	282,393	6.04	117	166	21.4	7740
42	**Thailand**	59,332	57,509	442,250	10.18	145	336	68.2	4920
43	**Malaysia**	55,211	53,979	218,280	7.24	106	253	28.8	8780
44	**Chile**	50,379	48,964	505,589	12.69	170	207	17.3	12000
45	**Slovakia**	44,051	42,903	296,161	7.32	131	85	5.4	15540
46	**Croatia**	41,951	40,676	215,609	5.99	118	57	4.5	12670
47	**Slovenia**	38,459	37,586	285,289	8.58	127	47	2	23550
48	**Pakistan**	38,274	36,650	156,030	5.54	93	188	189.6	992
49	**Bulgaria**	37,286	36,609	246,242	7.2	121	46	7.4	6270
50	**Saudi Arabia**	36,780	35,161	200,216	6.42	106	481	27.9	17250

Acknowledgements

The author would like to thank the editor and Sean van der Merwe for helpful comments and suggestions.

References

ADLER, A., EWING, J. and TAYLOR, P. (2009). Citation Statistics. *Statistical Science*, 24, no. 1, 1 -14.

ACADEMY OF SCIENCE OF SOUTH AFRICA (2010), The PHD Study.

BATISTA, P.D., CAPITELI, M.G. and KINOUCHI, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, Volume 68, Number 1, 179-189, DOI: 10.1007/s11192-006-0090-4.

BARCZA, K. and TELCS, A. (2009). Paretian publication patterns imply Paretian Hirsch index. *Scientometrics*, Volume 81, Number 2, 513-519, DOI: 10.1007/s11192-008-2175-8

BEIRLANT, J. and EINMAHL, J.H.J. (2010). Asymptotics for the Hirsch index. *Scandinavian Journal of Statistics, 37*, 355-364.

EVANS, J.A. (2008). Electronic Publication and the Narrowing of Science and Scolarship. *Science*, 321, no. 5887, 395 -399.

HIRSCH, J.A. (2005). An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*. 2005 November 15; 102(46): 16569–16572. Published online 2005 November 7. DOI: 10.1073/pnas.0507655102.

IGLESIAS, J.E. and PECHARROMÁN, C. (2006). Scaling the *h*-Index for Different Scientific ISI Fields, arXiv:physics/0607224

INGLEZI-LOTZ, R. and POURIS, A. (2011). Scientometric impact assessment of a research policy instrument: the case of rating researchers on scientific outputs in South Africa. *Scientometrics* 88: 747 – 760.

LEYEDESDORFF, L. (2011). World shares of publications of the USA, EU-27, and China compared and predicated using the new interface of the Web-of-Science versus Scopus. arXiv:1110.1802v2 [cs.DL].

MATTHEWS, A.P. (2012). South African universities in world rankings. *Scientometrics*, DOI 10.1007/s11192-011-0611-7.

POURIS, A. (2010). A scientometric assessment of the South Africa Development Community: science in the tip of Africa. *Scientometrics* 85: 145 – 154.

PRATELLI, L., BACCINI, A., BARABESI, L. and MARCHESELLI, M. (2011). Statistical analysis of the Hirsch Index, arXiv:1102.2701v1 [math.ST]. Forthcoming *Scandinavian Journal of Statistics.*

RESEARCH TRENDS, www.researchtrends.com.

SCHUBERT, A.and GLÄNZEL, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, 1, 179 – 184.

SCImago. (2007). SJR — SCImago Journal & Country Rank.
Retrieved February 08, 2012, from http://www.scimagojr.com

THE ECONOMIST (2011), The World in 2011.

THE ROYAL SOCIETY (2011). Knowledge, Networks and Nations: Global Scientific collaboration in the 21st century, Report 03/11, Published by Elsevier.

THE, Top Universities in Africa 2011-2012. *Top Universities.* Times Higher Education. Retrieved 10 February 2012.