



On the optimality of academic rankings of regions with RePEc data

Mishra, SK

North-Eastern Hill University, Shillong (India)

14 October 2008

Online at <https://mpra.ub.uni-muenchen.de/11098/>
MPRA Paper No. 11098, posted 14 Oct 2008 13:54 UTC

On the Optimality of Academic Rankings of Regions with RePEc Data

SK Mishra
Department of Economics
North-Eastern Hill University
Shillong (India)

I. Introduction: Based on the bibliographical data available with the RePEc (Research Papers in Economics), the Internet Documents in Economics Access Service (IDEAS) publishes every month the up-dated academic rankings of different geographic regions (countries/states in the US). Academic rankings of authors belonging to a particular region make the basic building block of the region's rankings. All authors affiliated to the institutions in a particular region are added to the pool of that region. However, authors with multiple affiliations have their scores split among all regions. For computing the academic rankings of authors, thirty one criteria are used. Among those criteria, six relate to the count of work (such as articles, papers, book chapters, software, etc), fifteen are based on citation counts (excluding self citations), six are based on journal page counts and four are based on the popularity of RePEc services (such as abstract views and downloads). The ranks obtained according to these criteria are then aggregated by computing their harmonic mean, albeit with some modifications. The harmonic mean rewards those who are particularly good in some category, and perhaps too much. For this reason, the harmonic mean is dampened somewhat by adding a constant to each rank, and then subtracting it from the mean. Further, for each author, the aggregate rank is truncated by dropping the best and worst rankings. Once all these attributions are made for the authors affiliated to different regions, a properly weighted aggregate of author ranks is obtained although with a special treatment given to the Hirsch index. Finally, the regions (countries/states of USA) are ranked according to the scores thus computed for the authors affiliated to them (Zimmermann, 2007).

It may be relevant to enquire if the overall rankings of different regions obtained by such aggregation of criterion-wise rankings as obtained by the IDEAS are optimal. The objective of this paper is to discuss a few of the principles of optimality and to test on these principles the overall region-wise rankings obtained by the IDEAS.

II. The Criteria of Ranking: The following are the criteria used by the IDEAS/RePEc for ranking of the authors registered in the RePEc.

1. Measures based on count of work: There are six measures based on count of work (in the RePEc data base), namely:

- (i) ANB Works: The number of distinct works of an author;
- (ii) DNbWorks: Count divided by number of authors on each work;
- (iii) ScWorks: Count with simple impact factor weights;
- (iv) AScWorks: Count with simple impact factor weights divided by number of authors on each work;
- (v) WScWorks: Count with recursive impact factor weights;
- (vi) AWSWorks: Count with recursive impact factor weights divided by number of authors on each work.

2. Measures based on citation counts: There are fifteen measures based on citation counts (excluding self citations) of works in the RePEc data base, namely:

- (i) NbCites: Simple citation count;

- (ii) ANbCites: Citation count divided by number of authors on each work;
 - (iii) ScCites: Citation count with simple impact factor weights;
 - (iv) AScCites: Citation count with simple impact factor weights divided by number of authors on each work;
 - (v) WScCites: Citation count with recursive impact factor weights;
 - (vi) AWSnCites: Citation count with recursive impact factor weights divided by number of authors on each work;
 - (vii) DCites: Citation count discounted by age;
 - (viii) ADCites: Citation count discounted by age and divided by number of authors on each work;
 - (ix) DScCites: Citation count with discounted impact factor weights;
 - (x) ADScCites: Citation count with discounted impact factor weights divided by number of authors on each work;
 - (xi) WDScCites: Citation count with recursive discounted impact factor weights;
 - (xii) AWDScCites: Citation count with recursive discounted factor weights divided by number of authors on each work;
 - (xiii) HIndex: h-index;
 - (xiv) NCAuthors: Count of citing registered authors;
 - (xv) RCAuthors: Rank weighted count of citing registered authors.
- 3. Measures based on journal page counts:** There are six measures based on journal page counts (in the journals registered in the RePEc data base), namely:
- (i) NbPages: Simple page count;
 - (ii) ScPages: Page count divided by number of authors on each work;
 - (iii) WSCPages: Page count with simple impact factor weights;
 - (iv) ANbPages: Page count with simple impact factor weights divided by number of authors on each work;
 - (v) AScPages: Page count with recursive impact factor weights;
 - (vi) AWSnPages: Page count with recursive impact factor weights divided by number of authors on each work.
- 4. Measures based on the popularity of RePEc services:** There are four measures based on viewing the abstracts and download of the full papers, articles, etc (available in the RePEc data base), namely:
- (i) AbsViews: Total abstract views in the past 12 months;
 - (ii) AAbsViews: Total abstract views per author in the past 12 months;
 - (iii) Downloads: Total downloads in the past 12 months;
 - (iv) ADownloads: Total downloads per author in the past 12 months.

In the measures enumerated above, *impact factors* of different types (simple, recursive and discounted) appear, which are computed as described in Zimmermann (2007). The *h-index* (Hirsch, 1995) is defined as: a scientist has index h if h of his/her H papers have at least h citations each, and the other $(H-h)$ papers have no more than h citations each. Here H is the total number of papers authored or co-authored by the scientist that are available in the bibliographical data base. The *h-index* can take on an integer value only. Thus, an author with h -index= h would have at least h^2 citations (at least h papers with at least h citations each). This index was developed for physics, where scientists write prolifically and cite their peers generously. Some physicists have h above 100, but in Economics it is very rare to have an h above 20, mainly due to the fact that economists write fewer, but more involved papers (Zimmermann, 2007). Additionally, empirical work (and oftentimes theoretical work as well) in

economics has much less universal relevance and acceptability than the empirical (or theoretical) work in physics enjoys. This fact limits the *h-index* of authors in economics. Tol (2008) improved h-index to take on a real value and called it the *rational successive g-index*.

III. Principles of Representation and the Measures of the Degree of Representation by Overall Rankings: A perusal of the ranking criteria listed above immediately gives an impression that the criteria of citation counts dominate over other criteria by the sheer number (fifteen) of measures included under them. On the other hand, the criteria based on abstract review and downloads include only four measures. Thus, it appears to be quite likely that the overall rankings of the authors with higher scores on citation criteria would dominate over the overall rankings of the authors having higher scores on productivity or popularity criteria simply due to the number of measures used in those criteria. The biases arising out of the unequal number of measures included under different criteria may not be eliminated even if averaging is done at the criteria level and then such criteria-averages are used to obtain the overall ranking by aggregation. It is also likely that such criteria-level averaging would introduce a new kind of bias into the overall scores and the rankings based on them.

We must, therefore, lay down certain principles of optimal representation of a multitude of criteria by a unique composite criterion. We must devise some measure of the degree of representation of individual rank scores (31 in number in the present context) by the overall rank scores. Although there can be many alternative principles of representation, it is generally accepted that correlation, which is a measure of joint movement of two variables, may be used as a fair principle of representation. Among many possible correlation-based measures of the degree of representation, the following are of an immediate interest to us.

1. The array of overall scores, $[S]$, is a linear combination of weighted individual rank score arrays, $[r_j]$, such that the sum of the squared coefficients of correlation (ρ) between S and r_j is maximum. Symbolically, $S = \sum_{j=1}^m w_j r_j : \max \sum_{j=1}^m \rho^2(S, r_j)$. If S is obtained in this manner, it is the factor score array associated with the first Principal Component (Kendall and Stuart, 1968). From S we obtain an n-element array of rank, $[R]$, $R = R(S)$, which is the sequence of positional values, the non-decreasing and non-interleaving natural numbers, uniquely assigned the elements of S , the latter arranged into a pre-assigned scheme of order. We will designate this R by the name of M_1 .
2. The array of overall scores, $[S]$, is a linear combination of weighted individual rank score arrays, $[r_j]$, such that the sum of the squared coefficients of rank correlation ($\tilde{\rho}$) between $R(S)$ and r_j is maximum. Or, $R = R(S); S = \sum_{j=1}^m w_j r_j : \max \sum_{j=1}^m \tilde{\rho}^2(R, r_j)$. Spearman's formula may be used to obtain the rank correlation or, alternatively, the Karl Pearson's formula can be applied on the arrays of rank values. We will designate this R by the name M_2 .
3. The array of overall scores, $[S]$, is a linear combination of weighted individual rank score arrays, $[r_j]$, such that the *minimal* squared (or equivalently, absolute) coefficient of rank correlation ($\hat{\rho}$) between $R(S)$ and r_j is maximum. Or, alternatively expressed, $R = R(S); S = \sum_{j=1}^m w_j r_j : \max(\min_j(\hat{\rho}^2(R, r_j)))$. We will designate this R by the name M_3 .

There can be two other measures of the degree of representation analogous to M_1 and M_2 . If we use the absolute rather than squared value of the coefficient of correlation, we can obtain M_4 analogous to M_1 and M_5 analogous to M_2 .

IV. The Data Base: We use the data published by the IDEAS for the month of September 2008. The country-wise (state-wise in case of USA) rankings for all 31 criteria are reproduced in Tables 1(A) and 1(B). In Table 1(A) we also present the overall rankings of different countries/states of US as computed by the IDEAS. We will denote the array of IDEAS overall ranking scores by M_0 . These overall rankings are also reproduced in Table 1(B). These are the rank scores that we purport to test for their optimality. Additionally, Table 1(A) reproduces the rankings on the basis of the simple counts of work (NbWorks) reported by the Ideas. We do not use these rankings in further analysis.

V. Computational Aspects: It is rather straightforward to obtain M_1 . The Principal Component Analysis can be carried out on the individual rankings matrix that has 168 rows (countries/states of US) and 31 columns (criteria-wise ranking scores) to obtain the factor scores array associated with the first Principal Component. This array is S . Then, countries/states can be ranked according to the ordered values of S . For doing this, several statistical software/programs (such as STATISTICA or SPSS) are available.

However, for obtaining M_2 through M_5 we do not have any closed form method. They must be obtained by direct nonlinear optimization. The direct optimization method can also be used for obtaining M_1 , although doing so is not economical.

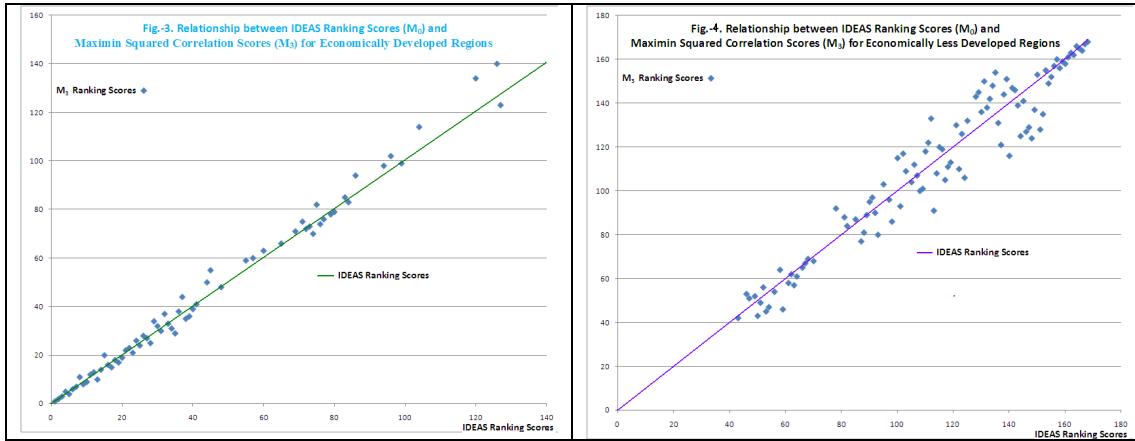
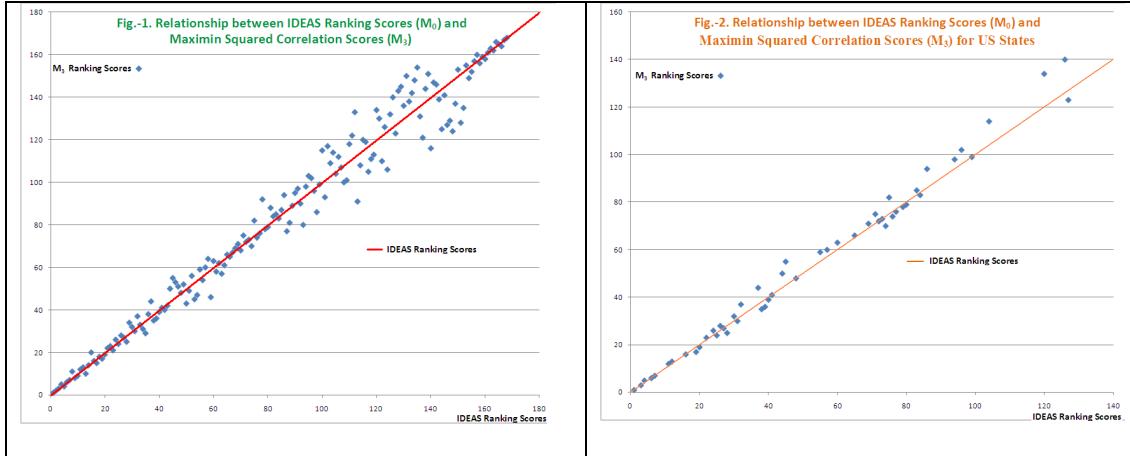
It may be noted that the nonlinear optimization problem to obtain M_2 through M_5 (as well as M_1 if one wants to obtain it directly) is extremely involved. Choice of appropriate values of decision variables (that is weights or w_j ; $j = 1, 2, \dots, 31$) so as to maximize an involved objective function like $\sum_{j=1}^m \tilde{\rho}^2(R(\sum_{j=1}^m w_j r_j), r_j)$ or $\min_j(\tilde{\rho}^2(R(\sum_{j=1}^m w_j r_j), r_j))$ is extremely difficult and cannot possibly be accomplished by any traditional method of nonlinear optimization. Hence, we proceed by using a relatively new method of global optimization, namely the method of Differential Evolution (Storn and Price, 1995; Mishra, 2006). This method has been successful in optimizing very difficult nonlinear multimodal optimization problems.

The Findings: As indicated in the previous sections, we have computed five different types of overall ranking scores: M_1 , M_2 , M_3 , M_4 , and M_5 . The IDEAS ranking scores (M_0) are already given. These alternative overall ranking scores are presented in Table-2. The correlation coefficients of these overall ranking scores (M_1 through M_5) with the individual ranking scores (for 31 different criteria) are presented in Table-3. The inter-correlation matrix of different types of overall ranking scores among themselves is presented in Table-4.

A perusal of Table-4 reveals that the array of overall ranking scores given by the IDEAS is very highly correlated with the arrays of alternative scores. The correlation between M_0 and M_3 is the least, yet as high as 0.9866. Since M_3 is obtained in a very specific manner (by maximization of the minimal squared correlation between M_3 and the individual ranking scores, r_j) it may be specifically mentioned that M_3 assigns largest weight (0.6769) to h-index rankings, followed by downloads rankings (0.4916), AWS Cites (0.2611), ADScites (0.2464), AWS Pages (0.1393), RCA Authors (0.1330) and ScPages (0.1266). Weights assigned to other individual ranking scores

(r_j) are rather much smaller. Vis-à-vis M_3 , other overall ranking scores (M_1 , M_2 , M_4 and M_5) assign equitable weights (between 0.15 to 0.19) to all individual ranking scores. A perusal of Table-2 also reveals that except for a few stray instances, M_0 , M_1 , M_2 , M_4 and M_5 rankings for different countries (as well as the states of USA) are identical.

However, the rankings of M_3 are noticeably different from those of M_0 . The differences between the two also have a clearly discernible pattern, best depicted in Fig.-1.



With an increasing value of the IDEAS ranking scores (M_0), the M_3 ranking scores increase, but its (M_3 's) spread or dispersion from M_0 also increases until M_0 reaches a score of 150 or so. After that, once again, they come closer to M_0 . Associated with them are such countries as Malta, Uzbekistan, Botswana, Niue, Georgia, Macedonia, Montenegro, Bahrain, Algeria, Senegal, Afghanistan, Cameroun, Oman, Paraguay, Nauru and Sudan.

A perusal of Fig.-2 through Fig.-4 also reveals that M_3 ranking scores favour the US states and the economically developed regions while M_0 is more favourable to the economically less-developed countries. Elsewhere it has been found that academic and professional participation of economists is significantly higher in developed regions than in the less developed regions (Mishra, 2008).

Concluding Remarks: We raised the question whether the method, best described by Zimmermann (2007), used by the IDEAS/RePEc to obtain academic rankings of different regions (countries and US states) in terms of the academic performance of economists associated with them can be considered optimal. We devised five different types of ranking procedure based on the *principles of representation* of numerically large and varied types of ranking criteria by a single index of overall ranking scores. We used the data published by the IDEAS for the month of September 2008.

We found that the overall ranking scores obtained by the IDEAS are almost optimal on the four (of the five) principles of representation. However, it is not so when the principle of representation is maximization of the minimal squared correlation of overall ranking scores with the constituent individual ranking scores. The overall ranking scores based on maximization of minimal squared correlation beget larger impact (weights) of a select few scientometric criteria such as *h-index*, download counts, and certain specific (co-authorship discounted) measures of impact-weighted citation and productivity of authors affiliated to the regions under consideration. As a consequence, it has some bias in favour of economically developed regions, while the overall ranking scores obtained by the IDEAS are slightly biased in favour of the economically less developed regions. The IDEAS rankings, therefore, have a tendency to discount for the disadvantages faced by the economists associated with the less privileged regions.

References

- Hirsch, J. E. (2005) "An Index to Quantify an Individual's Scientific Research Output", *Proceedings of the National Academy of Sciences*, 102 (46): 16569–16572.
- Kendall, M.G. and Stuart, A. (1968) *The Advanced Theory of Statistics*, vol. 3, Charles Griffin & Co., London.
- Mishra, S. K. (2006) "Global Optimization by Differential Evolution and Particle Swarm Methods: Evaluation on Some Benchmark Functions", SSRN: <http://ssrn.com/abstract=933827>
- Mishra, S.K. (2008) "What Determines the Academic and Professional Participation of Economists?", SSRN: <http://ssrn.com/abstract=1265044>.
- Storn, R. and Price, K. (1995) "Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces", *Technical Report, International Computer Science Institute*, Berkley.
- Tol, R.S.J. (2008) "A Rational, Successive G-Index applied to Economics Departments in Ireland", *Journal of Informetrics*, 2: 149-155.
- Zimmermann, C. (2007) "Academic Rankings with RePEc", *Working Papers Series #2007-36*, Dept. of Economics, Univ. of Connecticut: <http://www.econ.uconn.edu/working/2007-36.pdf>

Note: The Fortran Computer program available at <http://www.webng.com/economics/make-indices.html>

Continuation of Table-2													
Country / State	M₀	M₁	M₂	M₃	M₄	M₅	Country / State	M₀	M₁	M₂	M₃	M₄	M₅
Peru	93	92	92	80	92	92	Wisconsin	27	28	28	27	28	28
Iceland	95	95	95	103	95	95	Indiana	28	27	27	25	27	27
Uruguay	97	97	97	96	97	97	Virginia	30	29	29	32	29	29
Indonesia	98	98	98	86	98	98	Ohio	31	31	31	30	31	31
Iraq	100	101	101	115	101	101	Rhode Isl	32	33	33	37	34	34
Egypt	101	100	100	93	100	100	Arizona	37	38	38	44	38	38
Kuwait	102	103	102	117	103	103	Florida	38	37	37	35	37	37
Côte d'Iv	103	102	103	109	102	102	Georgia	39	36	36	36	36	36
U Arab Em	105	104	104	104	104	104	Tennessee	40	39	39	39	39	39
Tunisia	106	106	106	112	106	106	Iowa	41	41	41	41	41	41
Estonia	107	107	107	107	107	107	N Hampsh	44	44	44	50	44	44
Ethiopia	108	108	108	100	108	108	S Carolina	45	48	48	55	48	48
Venezuela	109	109	109	101	109	109	Washington	48	47	47	48	47	47
Croatia	110	111	111	118	111	111	Kentucky	55	55	55	59	55	55
Bangladesh	111	110	110	122	110	110	Oregon	57	56	56	60	56	56
S Arabia	112	114	114	133	114	114	Colorado	60	57	58	63	58	58
Jamaica	113	117	117	91	117	117	Utah	65	64	64	66	64	64
Kazakhstan	114	113	113	108	113	113	Louisiana	69	69	69	71	69	69
Iran	115	112	112	120	112	112	Wyoming	71	71	71	75	71	71
Bolivia	116	116	116	119	116	116	Alabama	72	72	72	72	72	72
Qatar	117	115	115	105	115	115	Kansas	73	73	73	73	73	73
Tanzania	118	118	118	111	118	118	Nevada	74	74	74	70	74	74
Nigeria	119	122	122	113	122	122	N Mexico	75	76	76	82	76	76
Fiji	121	121	121	130	120	120	Hawaii	76	75	75	74	75	75
Sri Lanka	122	119	119	110	119	119	Maine	77	77	77	76	77	77
Viet Nam	123	123	123	126	123	123	Vermont	79	79	79	78	79	79
Latvia	124	124	124	106	124	124	Oklahoma	80	80	80	79	80	80
Morocco	125	125	125	132	125	125	Delaware	83	82	82	85	82	82
Macao	128	128	128	143	128	128	Nebraska	84	83	84	83	84	84
Martinique	129	129	129	145	131	131	W Virginia	86	85	85	94	85	85
Malawi	130	131	131	136	130	130	Mississippi	94	94	94	98	94	94
Burkina F	131	132	132	150	132	132	Arkansas	96	96	96	102	96	96
Barbados	132	130	130	138	129	129	Montana	99	99	99	99	99	99
Réunion	133	133	133	142	133	133	Alaska	104	105	105	114	105	105
Guadeloupe	134	134	134	148	134	134	N Dakota	120	120	120	134	121	121
Kyrgyzstan	135	138	138	154	139	139	Idaho	126	126	126	140	126	126
Lithuania	136	135	135	131	135	135	S Dakota	127	127	127	123	127	127

Table-3. Correlation between the Different Types of Overall Academic Rankings (M) of Countries/States of US And Individual Ranking Scores (r_i)						
SI No.	Ranking Criteria (r_i)	M ₁	M ₂	M ₃	M ₄	M ₅
1	DNbWorks	0.98024140	0.97855130	0.97878522	0.98041463	0.97890726
2	ScWorks	0.99486657	0.99464722	0.98355445	0.99487714	0.99464722
3	AScWorks	0.99224418	0.99169877	0.98232698	0.99226883	0.99164309
4	ANbWorks	0.97913187	0.97714070	0.97920807	0.97931781	0.97747006
5	WScWorks	0.99199715	0.99105340	0.98197772	0.99201808	0.99101037
6	AWScWorks	0.98940233	0.98813025	0.98111216	0.98943793	0.98806191
7	NbCites	0.99446474	0.99325425	0.97408596	0.99438033	0.99312718
8	ANbCites	0.99397668	0.99315414	0.97525267	0.99389989	0.99308454
9	ScCites	0.99104519	0.98983130	0.96918074	0.99092385	0.98964794
10	AScCites	0.99217075	0.99120267	0.97301850	0.99206416	0.99107354
11	WScCites	0.98869078	0.98709271	0.96755042	0.98856725	0.98691401
12	AWScCites	0.99080088	0.98953225	0.97135218	0.99068828	0.98942089
13	DCites	0.99399332	0.99297320	0.97297099	0.99390031	0.99286230
14	ADCites	0.99438252	0.99351705	0.97577415	0.99430127	0.99346037
15	DScCites	0.99128257	0.99026518	0.96932794	0.99115667	0.99008698
16	ADScCites	0.99215427	0.99123890	0.97255123	0.99204155	0.99110164
17	WDScCites	0.98863971	0.98716004	0.96698326	0.98850511	0.98700724
18	AWDScCites	0.99104841	0.99006056	0.97082646	0.99092871	0.98990777
19	HIndex	0.93803227	0.93586584	0.96696920	0.93821363	0.93656923
20	NCAuthors	0.99501553	0.99406763	0.97679107	0.99495518	0.99399720
21	RCAuthors	0.99490884	0.99386367	0.97549010	0.99483262	0.99378315
22	NbPages	0.98549453	0.98486097	0.97694117	0.98555000	0.98496300
23	ScPages	0.98802319	0.98793218	0.97234506	0.98793753	0.98761298
24	WSPCPages	0.98472664	0.98439290	0.96772561	0.98461499	0.98394091
25	ANbPages	0.98529542	0.98421979	0.97837055	0.98536375	0.98438552
26	AScPages	0.98855953	0.98832543	0.97316476	0.98847467	0.98802666
27	AWScPages	0.98439462	0.98406604	0.96698251	0.98427926	0.98358852
28	AbsViews	0.96661155	0.96500321	0.97135568	0.96686987	0.96538031
29	AAbsViews	0.96430780	0.96241640	0.96924657	0.96457676	0.96295815
30	Downloads	0.95962064	0.95750680	0.96693679	0.95990300	0.95789149
31	ADownloads	0.96063736	0.95850649	0.96796432	0.96092055	0.95905062

Table-4. Inter-Correlation Matrix of Different Types of Overall Academic Rankings (M) of Countries/States of US						
Overall Rankings	M ₀	M ₁	M ₂	M ₃	M ₄	M ₅
M ₀	1.00000000	0.99957735	0.99960265	0.98660667	0.99954444	0.99952926
M ₁	0.99957735	1.00000000	0.99998735	0.98714574	0.99997216	0.99996710
M ₂	0.99960265	0.99998735	1.00000000	0.98712550	0.99997469	0.99996963
M ₃	0.98660667	0.98714574	0.98712550	1.00000000	0.98742920	0.98746716
M ₄	0.99954444	0.99997216	0.99997469	0.98742920	1.00000000	0.99999494
M ₅	0.99952926	0.99996710	0.99996963	0.98746716	0.99999494	1.00000000