Formulation of Advanced H-Index: The S-Index

Muhammad Imtiaz Subhani, Amber Osman, Syed Akif Hasan, Muhammad Saad
and Sabeen Mir[*]

ABSTRACT

H-index is one of the famous tool for assessing the research authors, journals and publications ranking. With some demerits and limitations of H-index, this research study puts forward an enhanced version of H-index named as S-index, which gives more accurate results in relevance to published papers and citations of a particular author (sum of various key factors of author's ranking) and exclusively negative impacts of self-citations and multiple authored papers. Pearson Correlation analysis was conducted to illustrate the association of RePec research factors with the respective indices. The proposed S-index has proven the negative relationship with number of self-citations and multiple authored papers unlike the H-index, which lacks these factors. An enhanced model is developed for the suggested s-index which is anticipated to be practical for the future research valuations. The implication of both h-index and the new s-index varies with a refining modification which was required to achieve. The author performance measure can be attained by implementing the new index as it embraces the important factors which contributes towards Author's profile.

Keywords: H-index, S-index, citations, self-citations

JEL Classification: C43

Introduction

Index in research is the means to assess the research performance and evaluate output of authors in different categories. Presently, there are numerous indices and bibliometric indicators in use, which are used to rank the authors and institutions in relevance to research and publications. To date no measuring research tool is perfect and has some or the other limit but is a sufficient measure to provide effective results.

This paper has proposed an index to encounter the demerits and limitations of the current measure research tool. The proposed index includes the measuring criteria for creating research rank of specified author in regards to his/her publications, citations and

many other aspects. The plotting of the proposed S-index formula was based on H-index constraints and its advancements. The important aspect of this paper is that S-index proved to assess the co-authorship or multiple authored papers , which h-index be deficient in (Bouyssou & Marchant, 2011a).

H-index does not calculate the papers with no citations and it causes a setback for the person doing his/her research. Also, H-index database is not well-run regarding the new publications (Norris & Oppenheim, 2010)

Literature Review

In the field of research varied inputs from different bodies, academic institutions, corporate organizations etc has a lot of weightage in terms of new knowledge put forward, which is in the form of research periodicals, publications, seminars, journals, workshops etc that serves as the key insights for the environment and business orientation one is heading in. As its research so it is more fragile in terms of finding new concepts and ideas and keeping resources of previous ones to generate or create something new or provide better information for a better future as a whole. In this domain, measurement or assessment is also relevant as it's the competition which makes one better than the others and in this drive more outputs are also achieved.

Research is all about newness and novelty in the work and requires a lot of hard work and deep scientific and technical learning. As it's not achieved at all without quality, the value is incredible and the awards and rewards are much higher as set by the international standards. In this accord, the contributors in research are given importance and the value is calculated in order to rank these individuals or organizations for the value generation and addition to the business and society.

To evaluate a researcher's work and worth of his work, the index is created to list/record the particular's author's entire papers/publications. This helps in portray the authorship caliber and also dignify the individual's input in a mutual effort.

For purpose, mostly H-index is used to determine the quality of the work the researcher has done and even as a rank order (Abbas, 2010). Quality is quite ascertained in this field with respect to scientific concerns and ranking is due to the assignments holding handsome amount for working.

The current tool used for assessing the author's ranking and research evaluation is H-index, which has the pros and cons to its functionality (Cardin, Corazza, Funari, & Giove, 2011). H-index and citations has a bundled association. H-index is known more with its distinctive assessment (Cronin & Meho, 2006). H-index is an index, which performs the calculation of a value and amount both for a distinct number. It furnishes comprehensive results.

One of the limitations is that it is biased with old scientists research work and gives the same h-index results of that of current scientists' research work though it shouldn't be the same because of the absence of work (inactive profiles) and duration. Similarly, for new researchers it is not a responsive index as it assesses the work on the count of the publications and not on the recent publication factor (Mingers, 2009). This index also ignores the multiple authors query i.e. no matter where the author is placed in the article, the calculation is on the same grounds (Hasan, Subhani, & Osman, 2012; Bouyssou & Marchant, 2011b).

Another issue with the h-index in relevance to the citations count is that once the h-index count is assessed, it doesn't cut down. The new researchers have no advantage to this tool. H-index relies on its own database, which has name spelling flaws resulting in inaccurate results (Hasan, Subhani, & Osman, 2012; Norris & Oppenheim, 2010).

H-index takes citations into account which deals with the published articles (Habibzadeh & Yadollahie, 2009). Other Bibliometric indicators have been used in contrast

to assess the achievements and rankings of the researchers (Marchant, 2009). Other quantitative dimension have been associated with publications and citations , which are not included in the h-index, which depicting the author's performance, which is risky for one (Costas & Bordons, 2007). H-index and G-index have been compared at various instances and the difference in both is that one evaluates the authorship and the other gives credit based on the author publication count respectively. Hence, both are dissimilar and with different limited indexes (Egghe, 2006).

Adding to the description about H-index, which calculates particular author's performance via name search – a flawless approach. Researchers who have attained a better citation count will have a more h-index in comparison with researchers having publications with no citations.

The search done in Web of Knowledge and Scopus has name initial search option, which Google scholar does not offer. Blind search is the left option, which goes into the closest broad search category (Jacso, 2008). Co-authorship can be incorporated by scheming the papers as per the authors taken in an inverse manner. It depends on the datasets collection order (Schreiber, 2009). Another complication is h and g index is that authors with small number of publication s and more citations will not be of any value to both of the indices as the numbers of publications are less (Norris & Oppenheim, 2010). Even if the work is being used in the author's articles, which is not being cited will also be not considered by the g-index. G-index is better in comparing the author's research performance with others and the upcoming years of success (Norris & Oppenheim, 2010).

In reference to the online databases, Google has developed page rank, which retrieves information and an improved location of web pages (Ding, Yan, Frazho, & Caverlee, 2009). Google has even come up with citation metrics and other tools in the research calculation via its vast and fast retrieval of database though it lacks certain points for showing up the results.

Another bibliometric research output tool was discussed, which is designed to provide quantity and quality of the publication with rational data set range and works with advanced scientific fields (Claro & Costa, 2011).

To date H-index has been preferred than the rest of the research output indicators and indices. It leads because it is good at interpreting the total citation count, total paper count and citations per paper. Hence, it is also taken as a predictor of future research performance and more of predicting productivity (Hirsch, 2007). One of the prime negative appeals to this index is the no count of co-authorship in a paper (Hasan, Subhani, & Osman, 2012).

H-index is more comprehensive than the other indices and indicators and is widely being used with the increase of academic research but it's important for authors to maintain their publication in an active mode i.e. longer time period (Norris & Oppenheim, 2010). An author research capability and future research performance is assessed by H-index via citations of each paper and papers published and cited as a whole.

An index is required for covering the co-authorship with the number of publications and citation count of a researcher to evaluate better index value. Also to exclude the self-citations as it's taken as a minus from the research quality point of view and also disregarding by the top research ranking institutions.

Hypotheses

The proposed research is based on the following proposition with the support of four hypotheses:

P = S-index measures impact of author with a better rational than h-index.

The proposed hypotheses are the following:

H1: S-index is negatively correlated with the number of multiple authored papers
H2: S-index is negatively correlated with self-citations number
H3: H-index has no relation with the number of multiple authored papers
H4: H-index is positively correlated with self-citations number

Research Methods

Method of Data Collection

The research is conducted on secondary data. Data was required for the author's research performance on the basis of various important factors. RePec & Endnote-Thomson Reuters provides enormous records of data in every factor which is required for the evaluation of performance of the author. The data for this research has been collected from RePec and Endnote databases.

Sampling Criteria

The sampling criteria for the data collection were the authors on the RePec top 5% ranking list as of October'2012. Total 21 authors have been taken for the sample, 20 of them have been selected with the criteria of 10th, 20th, 30th, 40th, . . 210th ranked author from the list. The 21st sample author is the supervisor and producer of this research index.

Sample Size

The sample size consisted of 21 authors from the RePec top ranking list.

Research Model Developed

Figure 3.1: Research Model

Proposed formula

The proposed formula on the basis of this research model (Figure.3.1) has been formulated and named as the new index; the 's-index':

$$Index = \log \left(\frac{N.P + N.Cit + h + N.JPg + N.AbsV_{t-1} + N.Dn_{t-1}}{N.SAP + (-N.MAP) + (-N.SCit)}\right)^2$$

Where;

N.P	= number of papers
N.Cit	= number of citations
h	= h-index
N.JPg	= number of journal pages
N. AbsV$_{t-1}$	= number of abstract views of previous month
N. Dn$_{t-1}$	= number of downloads in the previous month
N.SAP	= number of self authored papers
N.MAP	= number of multiple authored papers
N.SCit	= number of self citations

Example No.1 Calculation

Index= log $(\frac{N.P + N.Cit + h + N.JPg + N.AbsV_{t-1} + N.Dn_{t-1}}{N.SAP + (-N.MAP) + (-N.SCit)})^2$

Index= log $(\frac{71 + 27 + 3 + 6400 + 277 + 84}{0 + (-71) + (-23)})^2$

Index= log $(\frac{6862}{-94})^2$

Index= log $(-73)^2$

Index= log 5329

Index= 3.726

Example No.2 Calculation

Index= log $(\frac{N.P + N.Cit + h + N.JPg + N.AbsV_{t-1} + N.Dn_{t-1}}{N.SAP + (-N.MAP) + (-N.SCit)})^2$

Index= log $(\frac{273 + 29113 + 81 + 3352 + 3716 + 856}{88 + (-185) + (-755)})^2$

Index= log $(\frac{37391}{-852})^2$

Index= log $(-43.886)^2$

Index= log 1925.994

Index= 3.284

Statistical Technique

Pearson Correlation was used for data analysis. The Pearson Correlation has been used because it's a very useful statistical test for finding the linear relationship between two or more quantitative variables.

Results

Findings and Interpretation of the Results

The results from the Pearson Correlation analysis show the relationship between the selected RePec research factors with the h-index and s-index. There is variation in the findings of both the indices as h-index has positive relationship with the number of self citations, whereas s-index has the negative relationship. Also, h-index does not have any relation with the number of multiple co-authorship papers, whereas s-index has a negative relationship. The results matrixes for both indices are below:

Table 4.1: Correlation matrix for the h-index

Factors	h-index	
	Pearson Correlation	**Sig. Value**
Number of publications	.216	.347
Number of Citations	.931[**]	.000
Number of Journal Pages	-.259	.256
Number of Abstract Views	.885[**]	.000
Number of Downloads	.812[**]	.000
Number of self-authored papers	.159	.491
Number of multiple-authored papers	.221	.336
Number of Self Citations	.637[**]	.002

The results of correlation testing shows that h-index has a significant relationship with four factors i.e. N.Cit (number of citations), N.AbsV (number of abstract views), N.Dn (number of downloads) and S.Cit (self citations number), because it's significant values are .000, .000, .000, and .002 respectively that are < .05 and therefore having a significant relationship. The testing values showed that h-index is not correlated with number of multiple co-authorship papers and have a non significant relationship with it, although it is positively correlated with number of self citations factor as its correlation value is 0.637.

Table 4.2: Correlation matrix for the s-index

Factors	s-index	
	Pearson Correlation	**Sig. Value**
Number of publications	-.401	.072
Number of Citations	.143	.537
Number of Journal Pages	-.176	.446
Number of Abstract Views	-.043	.854
Number of Downloads	-.059	.798
Number of self-authored papers	-.116	.618
Number of multiple-authored papers	-.624[**]	.003
Number of Self Citations	-.460[*]	.036

The results of correlation testing shows that s-index has a significant relationship with only two factors i.e. N.MAP (multiple co-authorship papers) and S.Cit (self citations number), because it's significant values are .003 and .036 respectively that are < .05 and therefore

having a significant relationship. Although, s-index has a negative correlation with number of multiple co-authorship papers and number of self citations. It is moderately correlated with number of multiple co-authorship papers as its correlation value is -0.624 that is between the ranges of -0.5 to -0.8, and its correlation is weak with number of self citations factor as its correlation value is -0.460 that is between the ranges of 0 to -0.5.

Table 4.3: Hypotheses Assessment Summary

	Proposition	Empirical Conclusion
P1	s-index measure impact of author with a better rational than h-index	Failed to be rejected

	Hypotheses	**Coefficient of co-relation**	**Sig. Value**	**Empirical Conclusion**
H1	s-index is negatively correlated with the number of multiple authored papers	-.460	.036	Failed to be rejected
H2	s-index is negatively correlated with self-citations number	-.624	.003	Failed to be rejected
H3	h-index has a relation with the number of multiple authored papers	.221	.336	Failed to be accepted
H4	h-index is positively correlated with self-citations number	.637	.002	Failed to be rejected

Conclusion

The research proposed an idea for an index which considers the weightage for multiple authored papers and gives a negative impact of self citations on an authors rank. The applied testing on the gathered secondary data shows the objected positive findings. The new s-index is successful in bringing up a balance of the impacts and also showing the hypothetical relationship between the variables. Whereas, h-index was deficient in showing any relation with the number of multiple authored papers. Also, the impact of the total number of citations should be considered positive for any authors' performance, but the impact of self citations number should be different. S-index translates this idea very clearly.

Discussions

The data taken for the testing comprised of 21 authors ranking data. The results showed convincing results for s-index, as its relationship is moderately correlated with number of multiple authored papers. That means if authors' number of co-authorship papers increases then the value of s-index will decrease but with a moderate change. However, data variability is weak when comparing s-index value with the self citations number but still there is a negative correlation between them. Concluding, Higher the s-index value is, the higher the author ranking will be.

Implications

Lacking of h-index has been taken into consideration because it's the index which generally been used in order to determine author rank. Whereas study has focused on various factors which consist on the relevancy and determination of author hard work and it allowed formal procedure to know where author stands in the world stage of ranking. The accepted hypotheses of this research study opens door to a new s-index which can be successfully applied on any author's data and provides with an authentic ranking place in respective field.

References

Abbas, A. M. (2010).Weighted Indices for Evaluating the Quality of Research with Multiple Authorship. *Scientometrics*, 88(1): 107-131.

Bouyssou, D., & Marchant, T. (2011a). Bibliometric rankings of journals based on impact factors: An axiomatic approach. *Journal of Informetrics*, 5(1):75-86.

Bouyssou, D., & Marchant, T. (2011b). Ranking scientists and departments in a consistent manner. *Journal of the American Society for Information Science and Technology*, 62(9): 1761-1769.

Cardin, M., Corazza, M., Funari, S., & Giove, S. (2011). A Fuzzy-based Scoring Rule for Author Ranking - An Alternative of h-index. *WIRN*, 36-45.

Claro, J. & Costa, C. A. V. (2011). A made-to-measure indicator for cross-disciplinary bibliometric ranking of researchers' performance. *Scientometrics*, 86, 113–123.

Costas, R. & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, 1 (3), 193-203.

Cronin, B. & Meho, L. I. (2006). Using the H-index to Rank Influential Information Scientists. *Journal of the American Society for Information Science and Technology*, 57(9): 1275-1278.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60 (11), 2229–2243.

Egghe, L. (2006). An improvement of the H-index: The G-index. *ISSI Newsletter*, 2(1), 8–9.

Habibzadeh, F., & Yadollahie, M. (2009). Read the Articles; Don't Count Them. *Archives of Iranian Medicine*, 12(3): 302-303.

Hasan, S. A. & Subhani, M. I. & Osman, A., (2012). H-Index: The key to research output assessment. *MPRA Paper*, 39097, University Library of Munich, Germany.

Hirsch, J. E. (2007). Does the h index have predictive power? *PNAS*, 104(49): 19193–19198.

Jacso, P. (2008). The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. *Online Information Review*, 32(2): 266-283.

Marchant, T. (2009). An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors. *Scientometrics*, 80(2), 325-342.

Mingers, J. (2009). Measuring the research contribution of management academics using the Hirsch-index. *Journal of the Operational Research Society*, 60(9):1143 –1153.

Norris, M. & Oppenheim, C. (2010). The h-index: a broad review of a new bibliometric indicator. *Journal of Documentation*, 66(5): 681-705.

Schreiber, M. (2009). A Case Study of the Modified Hirsch Index hm Accounting for Multiple Coauthors. *Journal of the American Society for Information Science and Technology*, 60 (6), 1274–1282.