

Economic-Mathematical Model of the Hirsch Index Calculation

Dr. Andrew Novotorov

State College of Florida
USA

Dr. George Brikach

Gomel State Technical University
Belarus

Abstract

We are offering a simple method of adjustment of Hirsch Index using “Demand-Supply Law” of the Excel platform where quantity of published articles is a “supply” and quantity of cited articles is a “demand”. The optimization model will assist authors to see adjusted h-index score which is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications.

The h-index is an author-level metric that attempts to measure both the productivity and citation impact of the publications of a scientist or a scholar. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. The index can also be applied to the productivity and impact of a scholarly [1] as well as a group of scientists, such as a department or university or country [2]. The index was suggested in 2005 by Jorge E. Hirsch, a physicist at UCSD, as a tool for determining theoretical physicists' relative quality[3] and is sometimes called the Hirsch index or Hirsch number. However, many authors suggested different ways of Hirsch index adjustment and calculation. For example, M. Abdel-Aty in 2013 recommended equation that depends on the number of the single-author published papers and their citations to be added as a new factor to the known h-index. These factors give an advantage and make a clear evidence of innovative authors and reduce the known h-index for authors who are gaining citations by adding their names to multi-author papers [6].

It is well known that h-index or Hirsch index is a science-metric indicator that was proposed by the American physicist Jorge Hirsch of the University of San Diego, California in 2005. Hirsch index is a quantitative characteristic of productivity of a scientist, a group of scientists, scientific organization or country as a whole. It is based on the number of publications and the number of citations to those publications. The index is calculated based on the distribution of citations to works of the researcher. Although qualitative method such as h-index being used and grown increasingly by different institutions and different dimensions, confusion exists over how different authors with different contributions have the same h-index and how quality can be assessed when authors with multi-author papers compared to authors with single-author papers [4]. According to Hirsch, a scientist has index h if h of his or her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each [5].

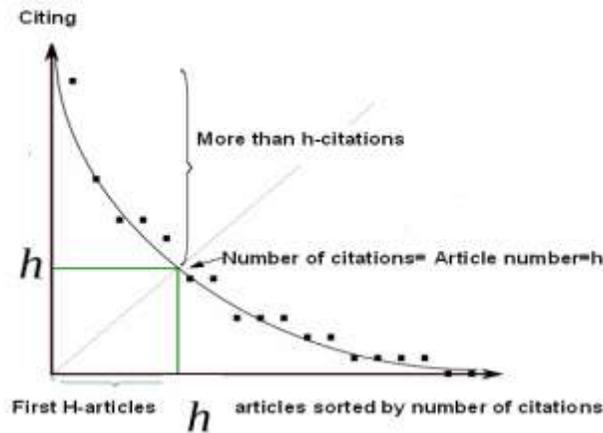


Fig. 1. Graphical representation of the Hirsch index

If you consider the algorithmic representation of index calculation, which is formulated by Hirsch, with widely known graphical representation on its calculation (fig. 1), it may be noted that there is no connection between them, in the form of a mathematical formula, it seems that the task is not formalized.

This paper will describe the alternative method of the Hirsch index calculation, which formalizes the index calculation algorithm correctly, and that can be used for practical calculations by any scientists without resorting to services of various types of information-analytical systems such as Scopus and Web of Science. The proposed method is especially effective for Hirsch index calculation in a case where the author personally and manually specifies a quantitative list of published and cited works, in which for some reason published articles have not been found by search engines like Scopus and Web of Science.

This method consists of two approaches in calculation. The meaning of the first proposed technique of calculation is based on a simple comparison of graphical representation of the Hirsch Index (Figure 1) and a graphical representation of the classical law of supply and demand (figure 2).

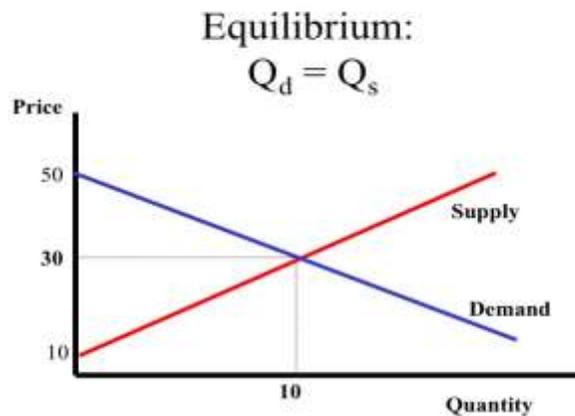


Fig. 2. Graphical representation of the Supply and Demand Law

Graphic analogy between curves in Fig. 1 view of “Hirsch index” and in Fig. 2 of the law of «supply and demand” shows their obvious similarities. Therefore, from an economic point of view, it was assumed that the number of published articles is equivalent to offerings in the market, and the number of citations is an analogue of market demand for published articles. So that a simple comparison indicates that the more articles published and less articles cited the lower the supply-demand rate. The more published articles cited the higher the rate.

Based on this principle authors built the economic-mathematical model of "supply and demand" in the form of simulation in Excel platform programming. The following table represents the h-index of one of the US scientists that we obtained in Scopus and we will use this data to come up with the proposed economic-mathematical model to calculate and adjust given Hirsch index.

Table 1. Output form Scopus (number of publications and their citations Hirsch index equal to 17)

United States		Author ID: _____7070		
About Scopus Author Identifier View potential author matches				
• Documents:38				
• Citations:1003 total citations by 720 documents				
• h-index: 17				
The h-index considers Scopus articles published after 1995 in Economics, Econometrics and Finance, Mathematics View More				
Article Title	Authors	Year	Source	Pages
Quantile and probability curves without crossing	2	2010	Econometrica	33
Inference on Counterfactual Distributions	3	2013	Econometrica	32
Ell ₁ -penalized quantile regression in high-dimensional sparse models	2	2011	Annals of Statistics	31
Least squares after model selection in high-dimensional sparse models	2	2013	Bernoulli	26
Square-root lasso: Pivotal recovery of sparse signals via conic programming	3	2011	Biometrika	22
Average and Quantile Effects in Nonseparable Panel Models	3	2013	Econometrica	13
Intersection Bounds: Estimation and Inference	3	2013	Econometrica	12
Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors	3	2013	Annals of Statistics	11
Least squares after model selection in high-dimensional sparse models	2	2013	Bernoulli	8
Local identification of nonparametric and semiparametric models	4	2014	Econometrica	7

As an input data to the proposed model we used the number of publications and the maximum number of citations (in descending order) in the Scopus system, the output of which is presented in table 1. Then from this table the quantitative values of published works and their number of citations were transferred and inserted into the programming Excel platform tool resulting wrapper as table 2, and their outcomes are illustrated as shown in Figure 3.

Table 2. The source data for the Hirsch Index model

Found in Scopus search engine		
The number of published works	The absolute number of references	The normalized number of work references
Supply	Demand	The relative demand (%)
1	33	100,00
2	32	96,97
3	31	93,94
4	26	78,79
5	22	66,67
6	13	39,39
7	12	36,36
8	10	30,30
9	8	24,24
10	7	21,21
11	6	18,18
12	6	18,18
13	5	15,15
14	5	15,15
15	3	9,09

The next step of the developing a method for calculating the index of Hirsch is the task of approximating functions of choice for the model of supply and demand. For supply approximation a linear function has been selected, and to approximate a demand the parabolic function was utilized:

- $Y = kX$ - linear function on supply
- $Y = aX^2 + bX + c$ -parabolic function on demand

Considering that Hirsch index is the balanced point between demand and supply curves (equilibrium), then the mathematical equation for calculation of the index Hirsch can be represented as follows:

(a) $aX^2 + bX + c = kX$ - for parabolic demand function

In order to define a specific numeric value of Hirsch index it is necessary to calculate specific numerical values of coefficients a, b, c-for parabolic equations and k-for a linear equation. This was done in the following way.

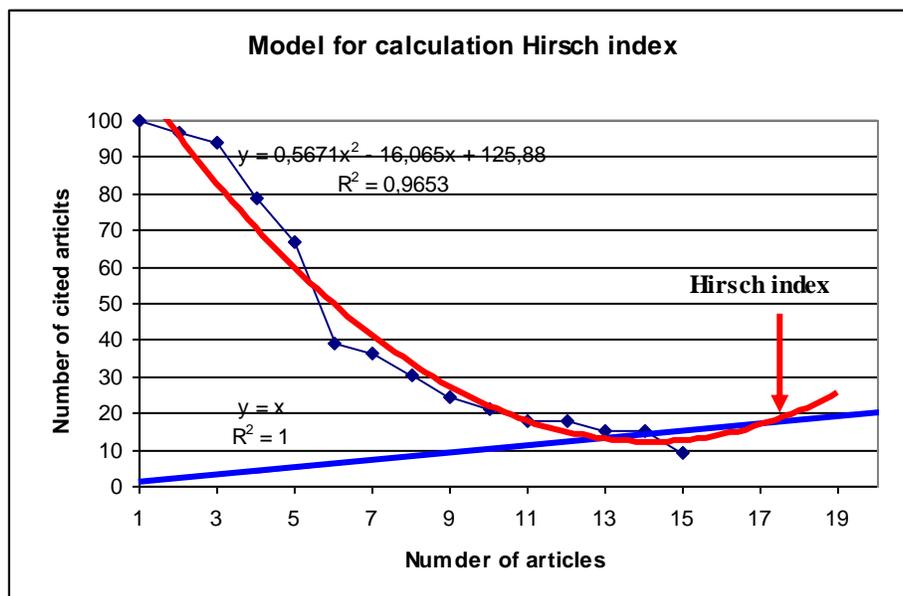
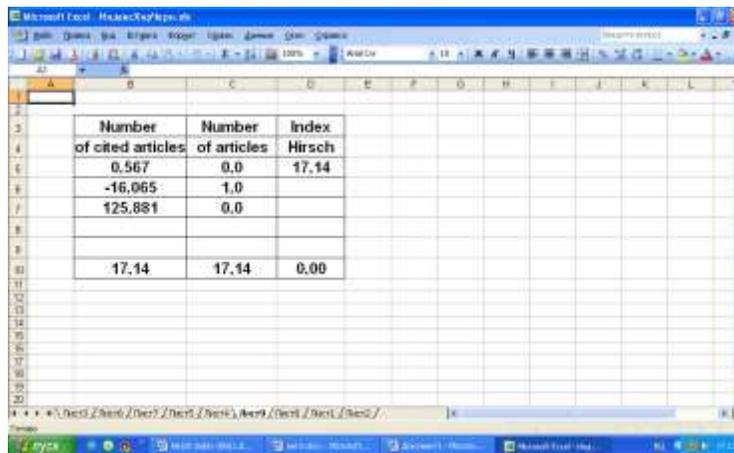


Fig. 3. Graph model of supply and demand, where the "demand" is the number of citations, "Supply"- the number of published works

First, we applied a trend analysis with Excel capabilities that were involved to complete the task, through the arrangements of ad hoc spreadsheet in Excel to fulfill the system of normal equations of first and second degree polynomials [7]. As a result, for a parabolic equation coefficients were as follows: (a) = 0.56, (b) = 16.06, (c) = 125.88 and for linear equation $k = 1$.

Based on tabulation of equations with the obtained values of the coefficients, the graph was built (Figure 3), that can graphically define the coordinates of the equilibrium for Hirsch index point. In this case, its value was about 17, which coincides with the value determined by Scopus system (it has the same value of 17). It speaks about the correctness of our proposed method of calculating the Hirsch index. In addition, for a more precise calculation of Hirsch index, you can use Excel "Goal seek" function or "function selection" (fig. 4) using the following equation:

-for parabolic equation $-0.56X^2 - 16.06X + 125.88 - X = 0$, in which case h value stood at 17.14.



	Number of cited articles	Number of articles	Index Hirsch
	0.567	0.0	17.14
	-16.065	1.0	
	125.881	0.0	
	17.14	17.14	0.00

Fig. 4. Screenshot of "function Selection" option to calculate h-index

It should be noted that for different source data, and hence different scholars who have different number of citations to their articles they will have different coefficients of parabolic equations. As a base number for published articles in our calculations 15 publications with the largest number of citations were selected. Our experience has shown that this is a sufficient number of publications. In case of necessity this number can be extended to a larger quantity of publications. Articles with zero citations do not affect the calculation of the index values (economically, this means that at the time of calculation of the index for the time being they do not have the demand) and therefore the basis of published works are not included.

Also, it may be noted that the evaluation index of the rounding format, which is used in Scopus is somewhat incorrect, because if one scholar index is higher than another, for instance, one's researcher's index is 5.4, but another's one is 5.7. The problem is that if you round the other one, it is the rounded value index and it is equal to 5, but then another will equal to 6. And this is a significant error, because the increase in the index for the 1 requires considerable time and the number of citations, because of the manifestation of demand for publication is inertial character.

And in conclusion, we would like to note the dignity of this model because it is in private use. Secondly, this technique is implemented in Excel that has a simple and intuitive interface. Thirdly, it is absolutely simply because originally, you should just put your data in two column input table, for example the number of your publications and their citations taken from the system, for example, Scopus. And if the automatic search system Scopus for some reason does not detect any number of your publications and links to them, then you can make these quantitative changes manually and without problems, adjust your Hirsch index.

The authors of this study hope that with basic knowledge of Excel application interested scientists can just use this clear and transparent tool, for both the economy and mathematics algorithm method, calculating Hirsch index monitoring and, if necessary, correcting your score, which will increase the efficiency of its scientific work.

References

1. "Google Scholar Metrics for Publications - Google Scholar Blog".googlescholar.blogspot.com.br.
2. Jones, T.; Huggett, S.; Kamalski, J. (2011). *Finding a Way Through the Scientific Literature: Indexes and Measures*. World Neurosurgery 76: 36.
3. Hirsch, J. E. (15 November 2005). *An index to quantify an individual's scientific research output*. PNAS 102 (46): 16569–16572.
4. Bakkalbasi, N. Bauer, K. Glover, J. and Wang, L. (2006). *Three options for citation tracking: Google Scholar, Scopus and Web of Science*. Biomed Digit Library. 3, 7.
5. Hirsch J.E. (2005). *An index to quantify an individual's scientific research output*, Proc. Nat. Acad. Sci. 46, 16569.
6. M. Abdel-Aty (2013). *New Index for Quantifying an Individual's Scientific Research Output*. <https://www.researchgate.net/publication/236935958> New Index for Quantifying an Individual's Scientific Research Output
7. Goldberger, Arthur S. (1964). "Classical Linear Regression". *Econometric Theory*. New York: John Wiley & Sons. pp. 156–212 [p. 158]