

A Relation Between h-Index and Impact Factor in the Power-Law Model

Leo Egghe

Hasselt University, Agoralaan, 3590 Diepenbeek, Belgium. E-mail: leo.egghe@uhasselt.be

Liming Liang

Institute for Science, Technology, and Society, Henan Normal University, Xinxiang, 453002, China, and School of Humanities and Social Science, Dalian University of Technology, Dalian, 116024, China. E-mail: liangliming1949@sina.com

Ronald Rousseau

Hasselt University, Agoralaan, 3590 Diepenbeek, Belgium; KHBO (Association K.U. Leuven), Industrial Sciences and Technology, B-8400 Oostende, Belgium; and K.U. Leuven, Department of Mathematics, Celestijnenlaan 200B, 3001 Leuven (Heverlee), Belgium. E-mail: ronald.rousseau@khbo.be

Using a power-law model, the two best-known topics in citation analysis, namely the impact factor and the Hirsch index, are unified into one relation (not a function). The validity of our model is, at least in a qualitative way, confirmed by real data.

Introduction

The h-index (Hirsch, 2005) and the impact factor (Garfield, 2006) are undoubtedly the best-known scientometric indicators. It has been shown in previous articles that the idea of calculating an impact factor or an h-index can be applied to many source-item relations and in many timeframes (Egghe, in press; Egghe & Rousseau, 2006; Frandsen & Rousseau, 2005; Liang & Rousseau, in press; Rousseau, Guns, & Liu, 2008). The reader is referred to these articles for the exact definition of an h-index and an impact factor in the general case, i.e., using any time window.

In this note we will propose a general relation (not a function!) in the power-law model. When stating that this relation is not a function we mean that many impact factors may correspond with one h-index value and vice versa.

Received March 18, 2009; revised May 6, 2009; accepted May 7, 2009

© 2009 ASIS&T • Published online 1 July 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21144

The Power-Law Framework in Which the Model is Developed

We adapt the framework described in Egghe (2005) and consider a given size-frequency function $f: [1, \infty[\rightarrow]0, C]$ of the form

$$f(j) = \frac{C}{j^\alpha} \quad (1)$$

where $C > 0$ and $\alpha > 2$. In a discrete setting $f(j)$ refers to the number of sources with production j (or, in the concrete setting of this article, the number of articles that received j citations). In a continuous framework f is interpreted as a density. So, we have two free parameters: C and α . We recall that in Egghe and Rousseau (2006, Appendix; see also Egghe, 2005, Exercise II.2.2.6, p.134) we proved that this setting is equivalent to a setting using the rank-frequency function $g(r) = \frac{B}{r^\beta}$, with $B, \beta > 0$ (corresponding to the parameters C and α) and $r \in]0, T]$. Here T denotes the total number of sources (concretely: published articles). The relations between the parameters are

$$B = \left(\frac{C}{\alpha - 1} \right)^{\frac{1}{\alpha-1}} \quad \text{or} \quad C = B^{(\alpha-1)}(\alpha - 1) \quad (2)$$

$$\beta = \frac{1}{\alpha - 1} \quad \text{or} \quad \alpha = \frac{1 + \beta}{\beta} \quad (3)$$

Recall also that

$$T = \frac{C}{\alpha - 1} \quad (4)$$

and that the total number of items (citations received) is, in this framework, equal to $\frac{C}{\alpha-2}$, which we assume to be strictly larger than 1.

We will work further in the size-frequency framework. Recall that in Egghe and Rousseau (2006) we showed that, in this framework,

$$h = T^{1/\alpha} = \left(\frac{C}{\alpha - 1} \right)^{1/\alpha} \quad (5)$$

We further recall (Egghe, 2005, p. 115) that, for $\alpha > 2$:

$$IF = \mu = \frac{\alpha - 1}{\alpha - 2} \quad (6)$$

where μ denotes the average production. Note that our model is only applicable if $IF > 1$. Equation 6 clearly shows that when investigating the relation between h and IF , α cannot be fixed as otherwise IF would also be fixed. Further, C will also be taken to be a variable parameter.

A Mathematical Relation

From Equation 6 we deduce that

$$\alpha = \frac{2IF - 1}{IF - 1} \quad (7)$$

This relation can also be found in Egghe (2005, p. 115). Combining Equations 5 and 7 yields

$$h = \left(\frac{C}{\frac{2IF-1}{IF-1} - 1} \right)^{\frac{IF-1}{2IF-1}} \quad (8)$$

or

$$h(C, IF) = \left(C \frac{IF - 1}{IF} \right)^{\frac{IF-1}{2IF-1}} = \left(C \left(1 - \frac{1}{IF} \right) \right)^{\frac{IF-1}{2IF-1}} \quad (9)$$

Keeping C fixed and calculating the first derivative of $h(IF)$, it is not difficult to show that $h'(IF)$ is positive, which proves that h is an increasing function of IF . Although we are able to calculate the second derivative $h''(IF)$ we are not able to determine its sign. Through numerical calculations we found that $h''(IF)$ is first convex and then concave. The inflection point is, however, always situated between $IF = 1.27$ and $IF = 1.9$ (for C between 1 and 10,000). This means that in practice this function can always be considered to be concave. Note also that $\lim_{IF \rightarrow \infty} h(IF) = \lim_{IF \rightarrow \infty} \left(C \frac{IF-1}{IF} \right)^{\frac{IF-1}{2IF-1}} = \sqrt{C}$. This is not only easy to check, but it also follows naturally from the power-law theory. Indeed, if IF tends to ∞ , then, by Equation 7, α tends to 2, and, by Equation 5, h tends to $\sqrt{T} = \sqrt{C}$ (by Equation 4, where $\alpha = 2$). Figures 1 and 2 illustrate the mathematical form of $h(C, IF)$ for various values of C .

We further note that if h and IF are given, C can be calculated from Equation 9, leading to

$$C = \frac{IF}{IF - 1} h^{\frac{2IF-1}{IF-1}} \quad (10)$$

Real Data

In this section we investigate if it is possible to find real curves resembling those shown in Figures 1 and 2. We collected the IF (year 2007) for all 304 journals in the eight JCR physics categories: applied; atomic, molecular, and chemical; condensed matter; fluids and plasmas; mathematical; multi-disciplinary; nuclear; particles and fields. Defining a review journal as a journal for which, according to the Web of Science (WoS) classification, more than 50% of its articles are review articles, we removed 22 review journals. Moreover, 122 journals with IF smaller than 1 were also removed (our theory is only valid if $IF > 1$). For the remaining 160 journals we calculated an h-index and a rational h-index, denoted as h_{rat} , based on the same time period as the IF, i.e., using citations in the year 2007 and publications in the year 2005 and 2006. More information on the rational h-index can be found in Guns and Rousseau (2009) and Ruane and Tol (2008). For fitting purposes we prefer the rational h-index as it corresponds better to our real-valued model. Using Equation 10 we obtained the corresponding C -value. Note that we claim that there is an $h(IF)$ -curve for each C -value separately. As it is not feasible to check this for real data, we grouped our data into

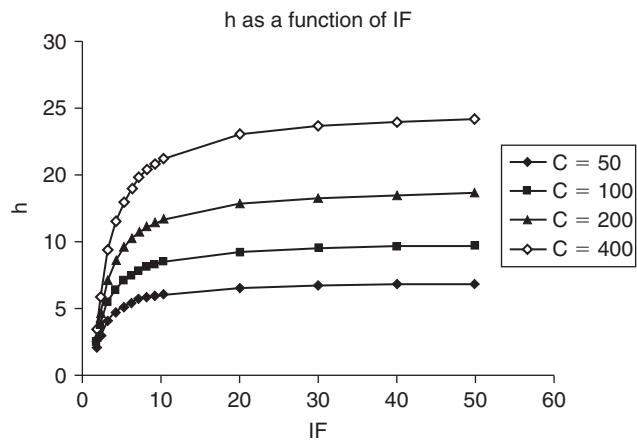


FIG. 1. The h-index (h) as a function of the impact factor (IF) for different values of C .

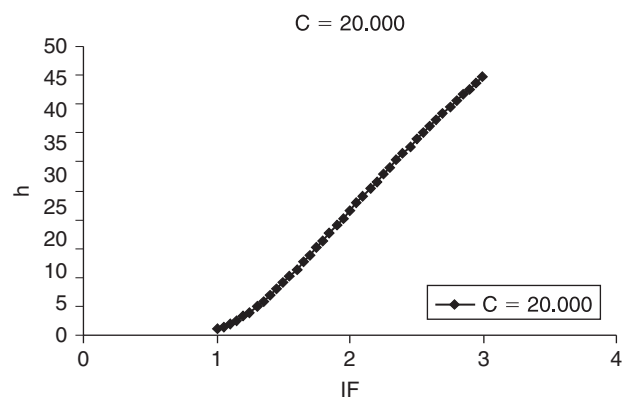


FIG. 2. For small values of IF and high C -values the inflection point (in the model) becomes visible.

TABLE 1. Fitting results.

Quartile	C-range of real data	Best-fitting C-value	R^2
I	23–760	366.04	0.74
II	761–2,700	1,471.6	0.87
III	2,701–23,000	4,311.33	0.93
IV	23,000–1,000,000	52,405.2	0.35

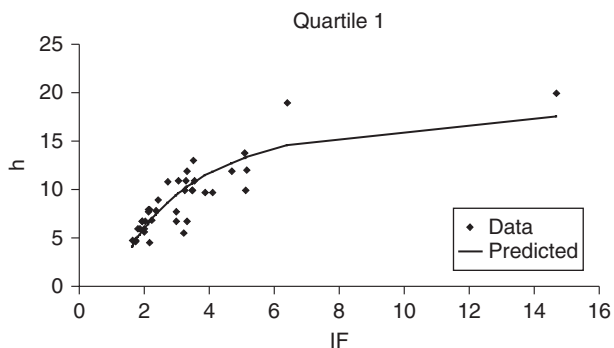


FIG. 3. The rational h-index as a function of the impact factor (first quartile).

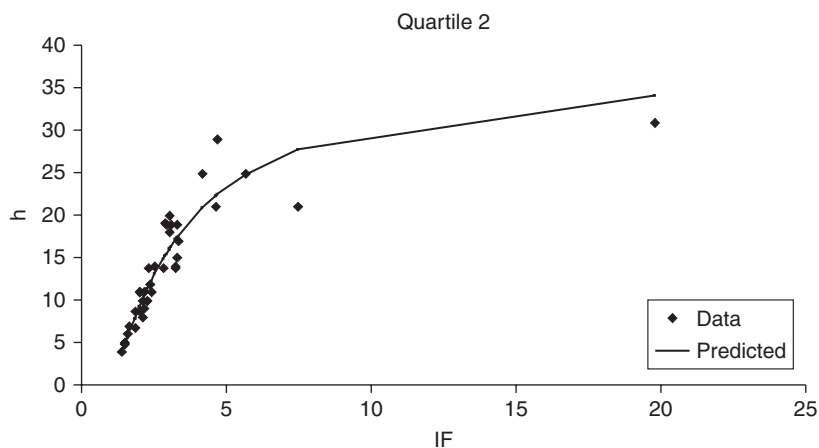


FIG. 4. The rational h-index as a function of the impact factor (second quartile).

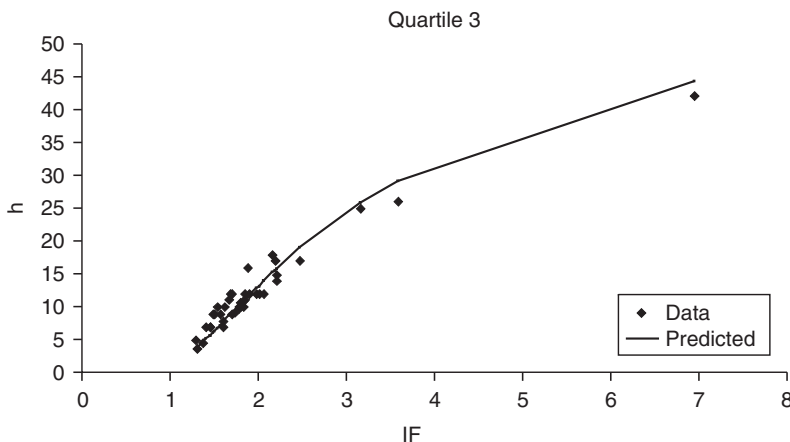


FIG. 5. The rational h-index as a function of the impact factor (third quartile).

four quartiles, depending on the C -value. For each of the four groups we determined a best-fitting curve (Equation 9), using the nonlinear least squares procedure; see Table 1 for results and R^2 -values. Fitting results are shown in Figures 3–6. The term *predicted* refers to the piecewise linear line obtained by connecting the points with coordinates consisting of the real IF -value and the h -value derived from the best-fitting curve. Clearly, our model corresponds to the observed data, at least in a qualitative way.

In a first version we had among the journals in the first quartile the *Journal of Cosmology and Astroparticle Physics*. Its coordinates in Figure 3 were (6.067, 3.857), which made it a clear outlier. We checked the data and found that according to the Journal Citation Reports this journal’s 372 articles had been cited 2,257 times in 2007 (only for articles published in the years 2005 and 2006), leading to a (high) impact factor of 6.067, while in the Web of Science, we found the same 372 articles, but this time they were cited only 93 times leading to a (low) h -index of 3.857. Clearly one of the two data sources is wrong. Hence we deleted this journal from our data. In any case, we consider this fact (finding—in this way—an error in the Thomson Reuters database) as a point in favor of our model.

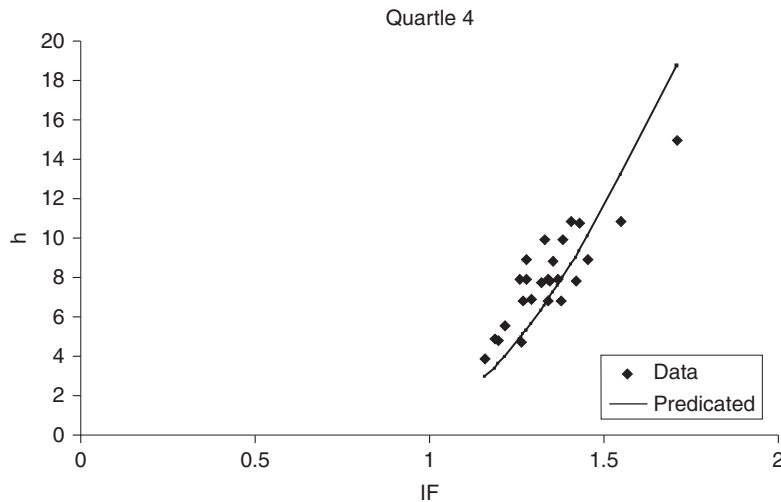


FIG. 6. The rational h-index as a function of the impact factor (fourth quartile).

It is remarkable that for our data high C -values correspond to a low impact factor. This is not a theoretical necessity (at least not in our model), but it may be interesting to find out why this is the case. For the fourth quartile we could not find a fit (R^2 is negative). For this reason we only considered C -values in the range 23,000–1,000,000 (this data range refers to 25 of the 40 original data). Higher C -values can be considered to be unrealistic, and not covered by our model. Recall that C is the density of the size-frequency function for $j = 1$ (see Equation 1). In this way we obtained a fit, although it is not very good. Yet it is clear that the data show a convex trend, as predicted by our model.

We also tried to find a fit for the second and third quartile combined. Although this was possible, the R^2 -value was not very high ($R^2 = 0.4$). Moreover, data points that clearly fitted the equation for one quartile became outliers in the combined case. This result confirms the fact that the relation between IF and h is not a function.

Schubert and Glänzel (2007) too established a relation between the h-index and the impact factor. In their model $h = cT^{1/3}IF^{2/3}$ when $\alpha = 2$. At the moment the relation between their approach and ours is not yet clear to us. We further note that Vanclay (2008) shows a graphical relation between the standard impact factor (IF) and the h-index. This h-index is, however, not the two-year one used by us. In his graph the relation between IF and h can be considered to be roughly linear. In relation to our theory nothing much can be derived from this graph.

Conclusion

We have shown a direct relation between any h-index and the corresponding impact factor in the power-law model. This relation is, however, not a function. It is shown that this model corresponds, at least to a large extent, to real data.

Acknowledgment

The work of Liming Liang and Ronald Rousseau is supported by the National Natural Science Foundation of China through Grant No. 70673019. This work was finished while R.R. was a guest at the Carlos III University of Madrid. He thanks Professor E. Sanz Casado and his research group for their hospitality during his visit.

References

- Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Oxford, England: Elsevier.
- Egghe, L. (in press). The Hirsch-index and related impact measures. *Annual Review of Information Science and Technology*.
- Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, 69, 121–129.
- Frandsen, T.F., & Rousseau, R. (2005). Article impact calculated over arbitrary periods. *Journal of the American Society for Information Science and Technology*, 56, 58–62.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, 295, 90–93.
- Guns, R., & Rousseau, R. (2009). Real and rational variants of the h-index and the g-index. *Journal of Informetrics*, 3, 64–71.
- Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16569–16572.
- Liang, L.M., & Rousseau, R. (in press). A general approach to citation analysis and an h-index based on the standard impact factor framework. *Proceedings of ISSI 2009*.
- Rousseau, R., Guns, R., & Liu, Y.X. (2008). The h-index of a conglomerate. *Cybermetrics*, 12(1), Paper 2. Retrieved May 29, 2009, from www.cindoc.csic.es/cybermetrics/articles/v12i1p2.html
- Ruane, F., & Tol, R.S.J. (2008). Rational (successive) h-indices: An application to economics in the Republic of Ireland. *Scientometrics*, 75, 395–405.
- Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, 1, 179–184.
- Vanclay, J.K. (2008). Ranking forestry journals using the h-index. *Journal of Informetrics*, 2, 326–334.