Journal of Informetrics 6 (2012) 721–728

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Statistical inference on the *h*-index with an application to top-scientist performance

A. Baccini [a,*], L. Barabesi [a], M. Marcheselli [a], L. Pratelli [b]

[a] *Department of Economics and Statistics, University of Siena, P.zza S.Francesco 7, 53100 Siena, Italy*
[b] *Naval Academy, viale Italia 72, 57100 Leghorn, Italy*

ARTICLE INFO

Article history:
Received 15 May 2012
Received in revised form 21 July 2012
Accepted 24 July 2012

Keywords:
h-index
Point and set estimation
Simultaneous pairwise confidence sets

ABSTRACT

Despite the huge amount of literature concerning the *h*-index, few papers have been devoted to its statistical analysis when a probabilistic distribution is assumed for citation counts. The present contribution mainly aims to divulge the inferential techniques recently introduced by Pratelli et al. (2012), by explaining the details for proper point and set estimation of the theoretical *h*-index. Moreover, some new achievements on simultaneous inference – addressed to produce suitable scholar comparisons – are carried out. Finally, the analysis of the citation dataset for the Nobel Laureates (in the last five years) and for the Fields medallists (from 2002 onward) is considered in order to exemplify the theoretical issues.

1. Introduction

On August 3rd 2005, Jorge E. Hirsch uploaded an article to the arXiv.org e-Print archive (http://arxiv.org/abs/physics/0508025), in which he introduced the so-called *h*-index by means of the following definition (as given in the fifth and last uploaded version): "a scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each" (see also Hirsch, 2005). On August 18th, Nature – which until that time had campaigned for a moderate use of the Impact Factor – published an article (Ball, 2005), where the *h*-index was presented as "transparent, unbiased and very hard to rig" and able to "pick out influential individuals". Arguably, this immediate success of the *h*-index is due in large part to its mathematical simplicity and its ease of calculation. In addition, the *h*-index is probably so diffused because it is perceived by non-technical readers as a unique numerical value measuring a very complex phenomenon such as the quality/impact/production of a researcher. Ranking scientists according to *h*-index is apparently very simple, and differences among researchers appear directly measurable.

Many papers discuss advantages and disadvantages of using the *h*-index, see *e.g.* Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2009), Costas and Bordons (2007), Egghe (2010) and Rousseau (2008). Some scientometric literature deals with the theoretical foundations of the *h*-index. In this respect, three main lines of research have been explored. The first line is the deterministic approach suggested by Hirsch (2005), according to which the *h*-index is the result of a linear growth model of publication and citation. More interestingly, a second line of research consists of the derivation of the *h*-index from Lotka's law (Egghe, 2005, 2006; Egghe & Rousseau, 2006; Ye, 2011). In contrast with these mathematical model approaches, Glänzel (2006) started the third line of research, emphasizing for the first time the relevance of a "statistical background" for the *h*-index. Glänzel required that the number of citations of a paper were a random variable and he derived some properties of

* Corresponding author. Tel.: +39 0577235233; fax: +39 0577232661
E-mail address: alberto.baccini@unisi.it (A. Baccini).

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2012.07.009

the h-index by assuming a Paretian model for the number of citations. The importance of this approach is stressed, among others, by Rousseau (2008) and Panaretos and Malesios (2009).

When the full statistical perspective is considered, *i.e.* by assuming a statistical model for the citation-count distribution, one must take into account the original definition provided by Jorge E. Hirsch gives rise to an empirical index and that the corresponding theoretical index has to be properly defined. Obviously, this process is – in some way – statistically unsound, since the "estimator" is defined in advance to the "parameter" which must be estimated. In any case, once the definition of the theoretical index is suitably carried out, the statistical properties of the empirical index must be assessed. Even if Glänzel (2006) produced the first effort in this direction, the decisive step was made by Beirlant and Einmahl (2010) who handled the empirical h-index as the estimator of a suitable statistical functional of the citation-count distribution. Beirlant and Einmahl (2010) also gave the consistency of the empirical h-index with respect to this functional and the conditions for its large-sample normality. In addition, they provided a variance estimation procedure when the underlying citation-count distribution displays Pareto-type or Weibull-type tails. However, Beirlant and Einmahl (2010) stated their theory by assuming a continuous citation-count distribution, even if the citation number is obviously an integer. Hence, Pratelli, Baccini, Barabesi, and Marcheselli (2012) further developed the results by Beirlant and Einmahl (2010), by achieving similar findings when the citation count follows a distribution supported by the integers. In addition, Pratelli et al. (2012) provided a suitable expression for the variance of the empirical h-index, which allows for simple and consistent nonparametric variance estimation. On the basis of these results, large-sample nonparametric confidence intervals may be implemented.

Panaretos and Malesios (2009) remarked that "while there exists a vast literature on the empirical h-index and its applications, relatively little work has been done on the study of the theoretical h-index as a statistical function, allowing to construct confidence intervals, test hypotheses and check the validity of its statistical properties". Hence, the aim of the present paper is to divulge the available statistical tools for the inference on the h-index, by trying to explain issues and details which might be obscure for non-statisticians. Moreover, new results on simultaneous inference are introduced in order to achieve suitable scholar comparisons. Finally, an extensive application to real data is given, in order to highlight the importance of producing interval estimation, in addition to point estimation.

2. Methodology

2.1. The empirical and theoretical h-index

Let us assume that X be an integer-valued random variable representing the citation number for a paper of a given scholar. Moreover, let us assume that S be the survival function corresponding to the random variable X, *i.e.* $S(x) = P(X > x)$. Therefore, $S(x)$ constitutes the probability that a paper of the scholar receives more than x citations. The random variable X is usually required to be "heavy-tailed" in the scientometric applications (see *e.g.* Glänzel (2006, 2010)), even if the results given in this section hold in general. Hence, if the scholar has published n papers, the random variables $X_1, \ldots, X_n$ represent the citation counts for his/her n papers. In order to develop the theory, it is assumed that $X_1, \ldots, X_n$ be identically and independently distributed.

On the basis of the Hirsch's definition given in the Section 1, the empirical h-index – say $\widehat{H}$ – may be mathematically expressed as

$$\widehat{H} = \max\{j \in \mathbb{N} : n\widehat{S}(j-1) \geq j\}, \tag{1}$$

where $\widehat{S}$ represents the empirical survival function, *i.e.*

$$\widehat{S}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(x,\infty)}(X_i), \tag{2}$$

while I_E turns out to be the usual indicator function of a set E, *i.e.* $I_E(x) = 1$ if $x \in E$ and $I_E(x) = 0$ otherwise. Obviously, $\widehat{S}(x)$ is the empirical rate of citation counts greater than a given x and hence $\widehat{S}$ is the "natural" estimator of S. Moreover, it is apparent that the quantity $n\widehat{S}(j-1)$ represents the number of papers receiving at least j citations. Thus, one can immediately realize that expression (1) formally states the empirical h-index in accordance with the definition provided by Hirsch (2005). Pratelli et al. (2012) however emphasized that (1) gives rise to the following alternative and equivalent (but more convenient) expression for $\widehat{H}$, *i.e.*

$$\widehat{H} = \sum_{j=1}^{n} I_{[j/n,1]}(\widehat{S}(j-1)). \tag{3}$$

Obviously, $\widehat{H}$ is a random variable since it depends on the random variables $X_1, \ldots, X_n$. Moreover, it should be noticed from (3) that $\widehat{H} = f(\widehat{S})$, *i.e.* the empirical h-index is actually a functional of the empirical survival function. This remark allows

for a suitable definition of the theoretical h-index – say h – which may be inherently defined as $h = f(S)$ by adopting the statistical "correspondence principle". More precisely, on the basis of expression (3) the theoretical h-index may be set to

$$h = \sum_{j=1}^{n} I_{[j/n,1]}(S(j-1)), \tag{4}$$

as suggested by Pratelli et al. (2012). Obviously, h depends on n and it is easily verified that $h \to \infty$ and $h/n \to 0$ as $n \to \infty$, a quite unusual behavior for a statistical parameter.

As to the main statistical properties of the empirical h-index, Pratelli et al. (2012) proved that

$$\mathrm{E}[\widehat{H}] = \sum_{j=1}^{n} p_j \tag{5}$$

and

$$\mathrm{Var}[\widehat{H}] = \sum_{j=1}^{n} p_j(1-p_j) + 2\sum_{l=2}^{n}\sum_{j=1}^{l-1} p_l(1-p_j), \tag{6}$$

where

$$p_j = \sum_{l=j}^{n} \binom{n}{l} S(j-1)^l (1-S(j-1))^{n-l}. \tag{7}$$

Thus, $\widehat{H}$ is a biased estimator for h. Indeed, h and $\mathrm{E}[\widehat{H}]$ do not generally coincide, since on the basis of (4) and (5) it follows that h is integer-valued, while $\mathrm{E}[\widehat{H}]$ is real-valued. However, Pratelli et al. (2012) have carried out a large simulation study empirically showing that the bias is negligible. Actually, for each statistical model and each sample size considered in the study, the absolute bias was less than one, while the bias could be either positive or negative. In any case, since Pratelli et al. (2012) showed that

$$\lim_n \mathrm{E}\left[\left(\frac{\widehat{H}}{h} - 1\right)^2\right] = 0, \tag{8}$$

it also follows that $\widehat{H}/h \xrightarrow{P} 1$ as $n \to \infty$, *i.e.* the ratio $\widehat{H}/h$ converges in probability to one. Thus, $\widehat{H}$ may be considered as a "consistent" estimator for h, even if in this setting the usual definition of consistency is pointless since the parameter approaches to infinity as simple size increases (see also a similar comment by Beirlant & Einmahl, 2010).

As previously emphasized, in the present framework some arbitrariness arises in the choice of the theoretical h-index and hence some attention is required in order to properly identify the reference parameter under estimation. Since in many statistical applications the expected value of the estimator coincides with the parameter to be estimated, we argue that $\mathrm{E}[\widehat{H}]$ could be considered as a "natural" competitor of h. This suggestion is also supported by the equivalence of h and $\mathrm{E}[\widehat{H}]$ for large n, *i.e.*

$$\lim_n \frac{\mathrm{E}[\widehat{H}]}{h} = 1, \tag{9}$$

as shown by Pratelli et al. (2012). In any case, the simulation study carried out by Pratelli et al. (2012) has shown that h and $\mathrm{E}[\widehat{H}]$ are very similar even for small n.

2.2. Large-sample properties of the empirical h-index

With the aim of achieving the implementation of large-sample confidence intervals for h or $\mathrm{E}[\widehat{H}]$, the assessment of the large-sample properties of $\mathrm{Var}[\widehat{H}]$ is of primary interest. It is worth noting that, since $\mathrm{E}[\widehat{H}] \to \infty$ as $n \to \infty$ and since scientometricians usually required "heavy-tailed" distributions for the citation counts, the most interesting case should imply that $\mathrm{Var}[\widehat{H}] \to \infty$ as $n \to \infty$.

First, in order to obtain a conservative estimator of $\mathrm{Var}[\widehat{H}]$, it is useful to introduce an operative condition onto the underlying citation distribution, *i.e.* for each $M > 0$ it is assumed that

$$\lim_n \left(\sup_{j \in D_M} \left|\frac{P(X=j)}{P(X=n)} - 1\right|\right) = 0, \tag{10}$$

where $D_M = [n - M\sqrt{n}, n + M\sqrt{n}] \cap \mathbb{N}$. Intuitively, if the random variable X satisfies the condition (10), for a large n its distribution is nearly uniform on an interval of natural numbers centered on n and with size proportional to $\sqrt{n}$. Practically

speaking, assumption (10) actually implies a "slow decrement" of $P(X=n)$ as $n\to\infty$. Actually, the underlying citation distribution is commonly assumed to be Pareto-type or Weibull-type (see *e.g.* Barcza & Telcs, 2009; Beirlant & Einmahl, 2010; Glänzel, 2006, 2010) and these distribution types – or their mixtures – satisfy condition (10). As a matter of fact, if l is a slowly-varying function, *i.e.* $l(tx)/l(t)\to 1$ for each x as $t\to\infty$, a Pareto-type distribution is characterized by a survival function given by $S(x)=x^{-\alpha}l(x)$ and hence it verifies (10) for any $\alpha>0$. This distribution type encompasses families of central importance for describing heavy-tailed discrete data, such as the discrete stable distribution (see *e.g.* Marcheselli, Baccini, & Barabesi, 2008; Zhu & Joe, 2009). Analogously, a Weibull-type distribution is characterized by a survival function given by $S(x)=\exp(-x^{\tau}l(x))$ and accordingly it verifies (10) for any $\tau<1/2$.

A "natural" estimator for the quantity p_j may be obtained in by means of a plug-in of the empirical survival function into the expression of p_j, *i.e.*

$$\widehat{p}_j=\sum_{l=j}^{n}\binom{n}{l}\widehat{S}(j-1)^l(1-\widehat{S}(j-1))^{n-l}. \tag{11}$$

Hence, on the basis of the expression of $\mathrm{Var}[\widehat{H}]$, by adopting in turn the statistical "correspondence principle", Pratelli et al. (2012) propose the variance estimator

$$\widehat{V}=\sum_{j=1}^{\min(\lfloor 3\widehat{H}\rfloor,n)}\widehat{p}_j(1-\widehat{p}_j)+2\sum_{l=2}^{\min(\lfloor 3\widehat{H}\rfloor,n)}\sum_{j=1}^{l-1}\widehat{p}_l(1-\widehat{p}_j), \tag{12}$$

where $\lfloor x\rfloor$ denotes the greatest integer less than or equal to x. In expression (12), the truncation of the summation extremes is due to some technical issues in order to improve the estimation (see Pratelli et al., 2012). Under condition (10) it generally holds that

$$\lim_n \mathrm{Var}[\widehat{H}]=\infty, \tag{13}$$

and Pratelli et al. (2012) proved the "consistency" of the estimator $\widehat{V}$, in the sense that

$$\frac{\widehat{V}}{\mathrm{Var}[\widehat{H}]}\xrightarrow{P}1 \tag{14}$$

as $n\to\infty$. In turn, equivalently to the Section 2, the usual definition of consistency is not useful since $\mathrm{Var}[\widehat{H}]$ approaches to infinity as simple size increases. It should be also remarked that estimator (12) is fully nonparametric since it does not require the specification of a model for the underlying citation distribution. In contrast, the variance estimator proposed by Beirlant and Einmahl (2010) assumes semi-parametric modeling and it implies the estimation of the Paretian index for the Pareto-type family, which is a complicated task. In addition, the computation of estimator (12) is straightforward from a practical point of view.

On the basis of the findings by Pratelli et al. (2012), if condition (10) is verified, the large-sample normality of $\widehat{H}$ holds, *i.e.* the following convergences in distribution are achieved

$$\frac{\widehat{H}-h}{\sqrt{\widehat{V}}}\sim\frac{\widehat{H}-\mathrm{E}[\widehat{H}]}{\sqrt{\widehat{V}}}\xrightarrow{d}N(0,1), \tag{15}$$

as $n\to\infty$, where – as usual – $N(0,1)$ represents a standard Normal random variable. The previous result provides the pivotal quantities for the implementation of a large-sample confidence set at the $(1-\gamma)$ confidence level for h given by

$$C=\{[\![\widehat{H}-z_{1-\gamma/2}\sqrt{\widehat{V}}]\!],\ldots,[\![\widehat{H}+z_{1-\gamma/2}\sqrt{\widehat{V}}]\!]\}, \tag{16}$$

where z_γ represents the γ-th quantile of the standard Normal distribution, while $[\![x]\!]$ represents the integer closest to x. Obviously, C turns out to be a confidence set since h may solely assume integer values. Similarly, a large-sample confidence interval at the $(1-\gamma)$ confidence level for $\mathrm{E}[\widehat{H}]$ is given by

$$C'=(\widehat{H}-z_{1-\gamma/2}\sqrt{\widehat{V}},\widehat{H}+z_{1-\gamma/2}\sqrt{\widehat{V}}). \tag{17}$$

It should be again emphasized that C and C' are fully nonparametric confidence set and interval, respectively. Indeed, their implementation does not demand the specification of a distribution, but solely requires the validity of condition (10), which is likely to hold for almost all the distributions of interest in the area of scientometrics. Moreover, a large simulation study carried out by Pratelli et al. (2012) show that the actual coverage of C and C' is appropriate even for quite small n.

In the case that k scholars have to be jointly compared, a suitable procedure should be applied in order to achieve simultaneous inference. Let us suppose that the k scholars act independently and that they have published $n_1,\ldots,n_k$ papers each, while their corresponding k theoretical h-indexes are given by $h_1,\ldots,h_k$. Accordingly, let $\widehat{H}_1,\ldots,\widehat{H}_k$ be the empirical h-indexes of these scholars and let $\widehat{V}_1,\ldots,\widehat{V}_k$ be the variance estimators. Thus, on the basis of the inequality suggested by

Šidák (1967), $k^* = k(k-1)/2$ large-sample conservative simultaneous confidence sets for the differences $(h_j - h_l)$ $(l > j = 1, \ldots, k)$ at the $(1-\gamma)$ confidence level are given by

$$C_{jl} = \{[\![\widehat{H}_j - \widehat{H}_l - z_{\gamma^*}\sqrt{\widehat{V}_{jl}}]\!], \ldots, [\![\widehat{H}_j - \widehat{H}_l + z_{\gamma^*}\sqrt{\widehat{V}_{jl}}]\!]\}, \tag{18}$$

where $\widehat{V}_{jl} = \widehat{V}_j + \widehat{V}_l$, while $\gamma^* = (1 + (1-\gamma)^{1/k^*})/2$. Equivalently, k^* large-sample conservative simultaneous confidence intervals for the differences $(\mathrm{E}[\widehat{H}_j] - \mathrm{E}[\widehat{H}_l])$ $(l > j = 1, \ldots, k)$ at the $(1-\gamma)$ confidence level are given by

$$C'_{jl} = (\widehat{H}_j - \widehat{H}_l - z_{\gamma^*}\sqrt{\widehat{V}_{jl}}, \widehat{H}_j - \widehat{H}_l + z_{\gamma^*}\sqrt{\widehat{V}_{jl}}). \tag{19}$$

Obviously, even if similar large-sample pairwise simultaneous procedures based on Šidák inequality are commonly adopted (see *e.g.* Drton & Perlman, 2004), more refined simultaneous proposal could be implemented, such as the bootstrap techniques recently suggested by Mandel and Betensky (2008) or by Xiong (2011).

3. Results and analysis

In order to exemplify the discussed statistical tools, we have considered the citation datasets of the Nobel Laureates in the last five years and of the Fields medallists from 2002 onward. Citation performances of these authors are drawn from an author search on Scopus carried out during February 2012. Tables 1 and 2 present the analyzed scholars, who are ordered according to their empirical h-indexes for each discipline. Moreover, in these Tables the number of papers, the empirical h-index and the large-sample confidence set at the 95% confidence level are given for each scholar.

As a specific example for the statistical interpretation of Tables 1 and 2, Adrian Fert – winner of the Nobel Prize for Physics in 2007 – displays an empirical h-index equal to 52, which is the highest for physicists. Once the inferential paradigm is assumed, $\widehat{H} = 52$ constitutes a point estimate of the (unknown) theoretical h-index that should be coupled with an estimate of the sampling variability. Loosely speaking, the set estimate – *i.e.* the corresponding confidence set $\{46, \ldots, 58\}$ – allows for jointly assessing the two aspects. The point estimate and the confidence set of Adrian Fert are not really different from those of Andre Geim, the second in this ranking. In contrast, Konstantin Novoselov displays an empirical h-index equal to 42, even if the corresponding confidence set, *i.e.* $\{35, \ldots, 49\}$, overlaps the confidence sets of the previous physicists.

A simple analysis of Tables 1 and 2 leads to three main conclusions. The first argument is well-known, in the sense that top scientists of different disciplines have different scientometric indexes. These differences mainly depend on the specific pattern of productivity and on the citation habits of the discipline. The second conclusion relies on the fact that in each discipline the use of the h-index flattens the performance of scholars. The choice of a unique value – synthesizing the individual productivity and the citations received – tends to equalize very different publication behaviors adopted by different scholars. As an example, in Physics Adrian Fert and Brian Schmidt have similar empirical h-indexes, even if Adrian Fert's papers double the papers published by Brian Schmidt. The third conclusion is the most striking one: in each discipline the majority of confidence sets intersects, so that a strict ranking of the considered scholars may not be feasible. This is a very important issue, since the common use of the h-index is to rank individuals, journals and so on. If these rankings fail to consider the sample variability, the differences between scholars in different positions may be not more than an optical illusion.

Nobel Laureates for Economics were analyzed in order to show the practical implementation of simultaneous inference. Since in this group there are $k = 10$ scholars with theoretical h-indexes given by $h_1, \ldots, h_{10}$, $k^* = 45$ differences $(h_j - h_l)$ $(l > j = 1, \ldots, 10)$ must be considered. Table 3 reports the corresponding large-sample pairwise simultaneous confidence sets at the 95% confidence level. By analyzing Table 3, if the simultaneous confidence sets not containing the zero are considered, some orderings on the theoretical h-indexes may be statistically stated. More precisely, by considering the first nine confidence intervals, it follows that $h_1 > h_8, h_9, h_{10}$; the subsequent eight confidence intervals provide $h_2 > h_8, h_9, h_{10}$; the subsequent seven confidence intervals provide $h_3 > h_9, h_{10}$; the subsequent six confidence intervals provide $h_4 > h_{10}$; the subsequent five confidence intervals provide $h_5 > h_{10}$; the subsequent four confidence intervals provide $h_6 > h_{10}$; and the subsequent three confidence intervals provide $h_7 > h_{10}$. Hence, a strict statistical ranking of these scholars is not available. Indeed, in synthesis, it can be solely stated that $h_1, h_2 > h_8, h_9, h_{10}$ and $h_3 > h_9, h_{10}$ and $h_4, h_5, h_6, h_7 > h_{10}$ at the 95% confidence level.

4. Discussion

The analysis of the considered dataset emphasizes that bibliometrics and scientometrics should require the application of a correct statistical approach, since the adopted methods often appear pre-statistical and pre-inferential. Indeed, once a statistical model is assumed, the sampling variability involved in the estimation of the theoretical h-index may be assessed by means of the proposed confidence sets and proper statistical ranking of the scholars may be achieved on the basis of simultaneous techniques. As noticed by Peter Hall "... issues that are obvious to statisticians are often ignored in bibliometric analysis...", and for example "... many proponents of impact factors, and other aspects of citation analysis, have little concept of the problems caused by averaging very heavy tailed data..." (IMS Presidential Address, IMS Bulletin Online, September 2, 2011). On the other hand, Hall concludes that "... we should definitely take a greater interest in this area". Indeed, also in our

Table 1
Citation performance of the considered Nobel winners. n_j and $\widehat{H}_j$ are the number of papers and the empirical h-index of the j-th Nobel winner respectively, while C_j is the large-sample confidence set for the corresponding theoretical h-index h_j (95% confidence level).

j	Physics	n_j	$\widehat{H}_j$	C_j	Medicine	n_j	$\widehat{H}_j$	C_j	Chemistry	n_j	$\widehat{H}_j$	C_j	Economics	n_j	$\widehat{H}_j$	C_j
1	Fert, A. (2007)	254	52	{46, . . ., 58}	Steinmann, R.M. (2011)	412	110	{102, . . ., 118}	Tsien, R.Y. (2008)	268	101	{92, . . ., 110}	Ostrom, E. (2009)	97	29	{24, . . ., 34}
2	Geim, A. (2010)	196	51	{43, . . ., 59}	Beutler, B.A. (2011)	308	77	{69, . . ., 85}	Steitz, T.A. (2009)	247	79	{69, . . ., 89}	Krugman, P. (2008)	71	29	{22, . . ., 36}
3	Schmidt, B.P. (2011)	126	45	{39, . . ., 51}	Hoffmann, J.A. (2011)	203	71	{61, . . ., 81}	Ertl, G. (2007)	545	72	{66, . . ., 78}	Sargent, T. (2011)	85	25	{21, . . ., 29}
4	Novoselov, K. (2010)	133	42	{35, . . ., 49}	Smithies, O. (2007)	297	66	{57, . . ., 75}	Negishi, E. (2010)	316	53	{48, . . ., 58}	Diamond, P.A. (2010)	49	19	{15, . . ., 23}
5	Perlmutter, S. (2011)	133	38	{32, . . ., 44}	Capecchi, M. (2007)	190	62	{54, . . ., 70}	Ramakrishnan, V.R. (2009)	100	47	{41, . . ., 53}	Maskin, E.S. (2007)	55	19	{14, . . ., 24}
6	Riess, A. (2011)	95	36	{30, . . ., 42}	Blackburn, E.H. (2009)	227	59	{51, . . ., 67}	Suzuki, A. (2010)	768	39	{33, . . ., 45}	Myerson, R.B. (2007)	47	19	{15, . . ., 23}
7	Kobayashi, M. (2008)	386	35	{32, . . ., 38}	Greider, C.W (2009)	97	58	{51, . . ., 65}	Shimomura, O. (2008)	259	39	{35, . . ., 43}	Pissarides, C.A. (2010)	37	17	{12, . . ., 22}
8	Grunberg, P.A. (2007)	132	21	{18, . . ., 24}	Szostak, J.W. (2009)	203	56	{48, . . ., 64}	Chalfie, M. (2008)	102	39	{34, . . ., 44}	Sims, C.A. (2011)	36	15	{10, . . ., 20}
9	Nambu, Y. (2008)	80	17	{13, . . ., 21}	zur Hauser, H. (2008)	338	54	{48, . . ., 60}	Heck, R.F. (2010)	106	30	{24, . . ., 36}	Mortensen, D.A. (2010)	28	12	{8, . . ., 16}
10	Boyle, W.S. (2009)	23	7	{4, . . ., 10}	Barré-Sinoussi, F. (2008)	251	45	{40, . . ., 50}	Yonath, A. (2009)	145	30	{25, . . ., 35}	Hurwicz, L. (2007)	20	7	{5, . . ., 9}
11	Smith, G.E. (2009)	30	5	{3, . . ., 7}	Evans, M.J. (2007)	141	44	{37, . . ., 51}	Shechtman, D. (2011)	81	16	{13, . . ., 19}				
12	Kao, C.K. (2009)	20	1	{0, . . ., 2}	Montagnier, L. (2008)	311	42	{36, . . ., 48}								
13	Maskawa, T. (2008)	5	1	{0, . . ., 3}	Edwards, R.G. (2010)	316	41	{35, . . ., 47}								

Table 2
Citation performance of the considered Field medallists. The notations are the same as in Table 1.

j	Field medallists	n_j	$\widehat{H}_j$	C_j
1	Tao, T. (2006)	164	29	{25, . . ., 33}
2	Villani, C. (2010)	55	21	{16, . . ., 26}
3	Okounkov, A. (2006)	48	18	{16, . . ., 20}
4	Werner, W. (2006)	39	16	{12, . . ., 20}
5	Lindenstrauss, E. (2010)	26	8	{5, . . ., 11}
6	Smirnov, S. (2010)	24	8	{6, . . ., 10}
7	Bao Chau, N. (2010)	9	7	{5, . . ., 9}
8	Voevodsky, V. (2002)	12	6	{3, . . ., 9}
9	Lafforgue, L. (2002)	5	2	{0, . . ., 4}
10	Perelman, G. (2006)	2	1	{0, . . ., 2}

opinion, scientometricians and statisticians should be more and more cooperative in order to achieve a proper development of the evaluation of the scientific performance.

As to the future directions of research, a critical point, as remarked by an anonymous referee, is to overcome the stringent requirements used in this paper. The inference on the theoretical h-index is carried out under the assumptions that citation counts be identically and independently distributed random variables and that the number of papers of a scientist be fixed. However, often scientists publish clustered papers on a special topic or containing the final results of a research project. Within these clusters citations might be more homogeneously distributed than between clusters. As a consequence data are not independently distributed anymore and the standard errors are actually higher than under the assumption of independently distributed data. Moreover, the number of papers may vary across years of publication (time period effects) and the number of papers of a scientist should be treated more properly as a further random variable. The assumed simplified model of this paper is solely a starting point towards proposing more adequate models. Indeed, in order to adhere more strictly to real frameworks, a scientist's production could be modeled in terms of a suitable stochastic process which takes into account dependence between citation counts - in such a way that large-sample results may be possibly obtained by means of the Martingale Central Limit Theorem, as well as the stochastic number of papers may be handled by using in addition the Anscombe–Doeblin Theorem. It is worth noting that under this improved model the variability of the empirical h-index is likely to increase – and in turn the size of the confidence sets.

Table 3
Pairwise simultaneous confidence sets for the dataset of the Nobel Laureates for Economics. C_{jl} represents the large-sample confidence set for the difference $(h_j - h_l)$ (95% simultaneous confidence level), where h_j is defined as in Table 1.

j	l	C_{jl}	j	l	C_{jl}
1	2	{−14, . . ., 14}	4	5	{−11, . . ., 11}
1	3	{−7, . . ., 15}	4	6	{−10, . . ., 10}
1	4	{−1, . . ., 21}	4	7	{−9, . . ., 13}
1	5	{−2, . . ., 22}	4	8	{−6, . . ., 14}
1	6	{−1, . . ., 21}	4	9	{−3, . . ., 17}
1	7	{0, . . ., 24}	4	10	{4, . . ., 20}
1	8	{2, . . ., 26}			
1	9	{6, . . ., 28}	5	6	{−10, . . ., 10}
1	10	{13, . . ., 31}	5	7	{−9, . . ., 13}
			5	8	{−7, . . ., 15}
2	3	{−9, . . ., 17}	5	9	{−4, . . ., 18}
2	4	{−3, . . ., 23}	5	10	{3, . . ., 21}
2	5	{−4, . . ., 24}			
2	6	{−3, . . ., 23}	6	7	{−8, . . ., 12}
2	7	{−2, . . ., 26}	6	8	{−6, . . ., 14}
2	8	{1, . . ., 27}	6	9	{−3, . . ., 17}
2	9	{4, . . ., 30}	6	10	{4, . . ., 20}
2	10	{10, . . ., 34}			
			7	8	{−9, . . ., 13}
3	4	{−4, . . ., 16}	7	9	{−6, . . ., 16}
3	5	{−4, . . ., 16}	7	10	{2, . . ., 18}
3	6	{−3, . . ., 15}			
3	7	{−2, . . ., 18}	8	9	{−7, . . ., 13}
3	8	{0, . . ., 20}	8	10	{0, . . ., 16}
3	9	{3, . . ., 23}			
3	10	{11, . . ., 25}	9	10	{−3, . . ., 13}

Acknowledgements

We would like to thank the two anonymous referees for the kind and helpful comments on the paper. We would also like to thank Prof. Lorenzo Fattorini for a careful reading of the early version of the paper.

References

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). *h*-index: A review focused in its variants computation and standardization for different scientific fields. *Journal of Informetrics*, *3*, 273–289.
Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, *436*, 900.
Barcza, K., & Telcs, A. (2009). Paretian publication patterns imply Paretian Hirsch index. *Scientometrics*, *81*, 513–519.
Beirlant, J., & Einmahl, J. H. J. (2010). Asymptotics for the Hirsch index. *Scandinavian Journal of Statistics*, *37*, 355–364.
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*, 193–203.
Drton, M., & Perlman, M. D. (2004). Model selection for Gaussian concentration graphs. *Biometrika*, *91*, 591–602.
Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Oxford: Elsevier.
Egghe, L. (2006). Theory and practise of the *g*-index. *Scientometrics*, *69*, 131–152.
Egghe, L. (2010). The Hirsch index and related impact measures. *Annual Review of Information Science and Technology*, *44*, 65–114.
Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, *69*, 121–129.
Glänzel, W. (2006). On the *h*-index – A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, *67*, 315–321.
Glänzel, W. (2010). On reliability and robustness of scientometrics indicators based on stochastic models. An evidence-based opinion paper. *Journal of Informetrics*, *4*, 313–319.
Hirsch, J. E. (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.
Mandel, M., & Betensky, R. A. (2008). Simultaneous confidence intervals based on the percentile bootstrap approach. *Computational Statistics & Data Analysis*, *52*, 2158–2165.
Marcheselli, M., Baccini, A., & Barabesi, L. (2008). Parameter estimation for the discrete stable family. *Communications in Statistics – Theory and methods*, *37*, 815–830.
Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, *81*, 635–670.
Pratelli, L., Baccini, A., Barabesi, L., Marcheselli, M., Statistical analysis of the Hirsch index, Scandinavian Journal of Statistics, http://dx.doi.org/10.1111/j.1467-9469.2011.00782.x, in press.
Rousseau, R. (2008). Reflections on recent developments of the h-index and h-type indices. In H. Kretschmer, & F. Havemann (Eds.), *Proceedings of WIS 2008, fourth international conference on webometrics, informetrics and scientometrics & ninth COLLNET meeting* Berlin,
Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate distributions. *Journal of the American Statistical Association*, *74*, 626–633.
Xiong, S. (2011). An asymptotics look at the generalized inference. *Journal of Multivariate Analysis*, *102*, 336–348.
Ye, F. Y. (2011). A unification of three models for the *h*-index. *Journal of the American Society for Information Science and Technology*, *62*, 205–207.
Zhu, R., & Joe, H. (2009). Modelling heavy-tailed count data using a generalized Poisson-inverse Gaussian family. *Statistics & Probability Letters*, *79*, 1695–1703.