Scientometrics
https://doi.org/10.1007/s11192-019-03278-7

A flexible approach for measuring author-level publishing performance

Nadia Simoes[1] · Nuno Crespo[1]

Received: 9 March 2019

Abstract
We propose a framework to evaluate, in relative terms, author-level publishing performance. To that end we introduce the publishing performance index (PPI) and the publishing performance box (PPB), and discuss the associated publishing profiles. We illustrate our approach conducting an extensive empirical application covering 472 top economists and developing several robustness tests. Instead of using a pre-designed measure without flexibility to adjust to the circumstances of each specific case, our approach accommodates alternative evaluation criteria, as defined by the evaluators. Beyond this key characteristic, our approach has some other important advantages: (1) it is easy to apply; (2) it is sensitive to the full list of publications and citations; (3) it is able to include additional dimensions of scientific performance beyond papers and citations; (4) it is a high granularity measure, providing a complete ranking of the authors under analysis.

Keywords Research evaluation · Scientific outputs · Bibliometrics · Selection criteria · Evaluators

JEL Classification A11 · A14 · I23 · M51

Introduction

Bibliometric analysis is gradually extending to the full spectrum of disciplines (Wildgaard et al. 2014) and gaining widespread use as support for critical decisions concerning research funding allocation, academic promotion, hiring, awards, and academic rankings (Alonso et al. 2009; Perry and Reny 2016; Hamermesh 2018; Osório 2018; Schreiber 2018). Getting fair decisions in all these aspects requires precise answers to difficult questions: Who is the best author? Who reaches a given level of performance? What is the profile of a given author?

Either as main criterion or as support for a qualitative evaluation, bibliometric indicators provide objective measures to rank authors according to the dimensions considered

✉ Nuno Crespo
nuno.crespo@iscte-iul.pt

1 ISCTE Business School Economics Department, BRU - IUL (Business Research Unit), Instituto Universitário de Lisboa (ISCTE - IUL), Av. das Forças Armadas, 1649-026 Lisbon, Portugal

relevant in each evaluation. The validity of the answers—with strong financial and career impacts—critically depends on the rigor of the measures (Bornmann and Marx 2011; Schreiber 2018) and on their adequacy to the dimensions under analysis, which are defined by the evaluators in each specific case. The *h*-index, suggested by Hirsch (2005), is the best-known author-level scientific performance measure (Todeschini and Baccini 2016). Despite its great popularity, it has several well-known drawbacks (Alonso et al. 2009; Egghe 2010). These limitations have led to the emergence of a myriad of alternative measures, each of them trying to solve one of the deficiencies of the original index.

The aim of this paper is to contribute to this area of research not by introducing a new variant of an existing index, but instead by bringing a comprehensive framework for the measurement of author's publishing performance. Our approach is particularly suitable when the evaluation occurs in the context of a specific competition (e.g., hiring, promotions, awards, or grants). It is based on two core ideas which are the key to successfully address this issue. First, we assume that the context matters. If we consider an author with 15 papers and 600 citations, is this a good or a bad performance? We cannot answer in absolute terms, only by comparison with the competing authors. The second idea is that there are several criteria that can be used to identify the best author. Who is the best: the author with more citations, more citations per year, more citations per paper, more papers published, more papers in top journals, or the author of the most cited paper? All these possibilities are valid to some extent and can be set as evaluation criteria if the evaluators decide to do so. Therefore, this second assumption implies that: (1) the criteria used in each case to select one author or group of authors should be explicitly defined; (2) we need a flexible approach, able to accommodate alternative criteria.

We develop a framework that incorporates these ideas. To that end, we propose an index, defined in relative terms (publishing performance index—PPI), develop a visual tool (publishing performance box—PPB), and analyze the publishing performance profiles emerging from the analysis.

The paper is organized as follows. Section two reviews the key methodological issues inherent to an author-level bibliometric analysis. Section three presents the simplest version of our methodology. Section four discusses the data used in the empirical analysis. Section five applies the methodology to evaluate the publishing performance of the members of economics departments of world top universities. Section six provides robustness tests to alternative evaluation criteria. Section seven proposes solutions for the comparison of authors in different stages of their careers. Section eight extends the methodological approach introducing new dimensions of performance. Section nine concludes.

Background: methodological options and measures

The debate on the measurement of author's publishing performance is focused on the properties of the indicators and their inherent methodological options. The *h*-index, proposed by Hirsch (2005), is the most commonly used metric. Until 2019, it was cited 3971 times (Web of Science, in 2 October).[1] An author has an *h*-index h if h of his/her papers have at least h citations each, while the remaining papers have no more than h citations each. Its great impact and popularity result from two appealing characteristics (Alonso et al. 2010):

[1] Google Scholar reports 8,943 citations on the same date.

(1) it joins in one single measure two critical dimensions of performance—number of papers (production) and number of citations (impact); (2) its simplicity. These factors have led both Scopus and Web of Science to include it as an indicator less than 2 years after its introduction (Egghe 2010). Nevertheless, this measure has important drawbacks. Two general problems can be highlighted. First, it is a low granularity index, thereby implying the existence of many ties. Second, it usually does not consider the full list of publications (and citations) of the author (Perry and Reny 2016; Fenner et al. 2018). These limitations derive from the methodological options assumed. The objective of this section is to provide an overview of five main aspects that must be addressed before undertaking a specific empirical exercise: dimensions of performance, citations, fields of research, academic age, and number and sequence of authors.

Dimensions of performance

The evaluation of publishing performance requires identifying the dimensions that should be taken into account. While some traditional measures confine their focus to citations, one of the advantages of the *h*-index is the joint consideration of both number of papers and number of citations (Alonso et al. 2009; Bornmann and Haunschild 2018). Some authors argue, however, that even in this case we are neglecting two important aspects—the quality of the journals and the quality of the citations—since all papers and all citations are assumed to be equal.

The exclusion of the quality of the journals as a relevant criterion can be seen as a shortcoming since there are significant differences among them regarding the quality filters applied by editors and referees (Van Raan 2006; Bornmann and Haunschild 2018). In fact, if the quality of the journal is not relevant for performance, what is it that justifies selective publication strategies, and why do top universities preferentially hire authors with papers published (or accepted) in a selected group of top journals? As mentioned by Bornmann et al. (2018, p. 659), “scientists with publications in high-ranked journals have a higher probability of getting tenure, research funding, and/or reputation”. One argument to exclude the quality of the journals from the evaluation is that it is reflected in the number of citations (redundancy argument). However, many papers with high impact are published in journals with low impact factor and there are many papers published in top journals that do not receive citations (Hamermesh 2018; Kosmulski 2018).

Another characteristic of most measures is that they treat all citations as equal, ignoring their quality (Gao et al. 2016; Dunnick 2017). The quality of a citation can be analyzed at three distinct levels: the relevance of the citing authors (Ding 2011), the quality of the citing journals (Bergstrom et al. 2008), and the quality of the citing papers (Schreiber 2018).

Citations

The papers that contribute to the *h*-index compose the *h*-core. Once a paper belongs to the h-core, additional citations do not receive any credit (excess citations). This means that an influential paper, with hundreds of citations, contributes the same as another one with h citations. This is commonly seen as a shortcoming (Dunnick 2017; Fenner et al. 2018; Schreiber 2018) and has led to the emergence of a vast set of alternative measures. The *g*-index emphasizes the highly cited papers, aiming to capture their disproportionate impact. An author obtains a *g*-index g when g is the highest rank such that the top g papers (in terms of citations) have, together, at least g^2 citations, while the top $(g + 1)$ papers have together fewer than $(g+1)^2$ citations (Egghe 2006). Since the distribution of citations among the g-core papers does not matter for the g-index, $g_i \geq h_i$, for any author i. The hg-index aims to minimize the limitations of h and g while retaining their most appealing characteristics. Proposed by Alonso et al. (2010), it is defined as the geometric mean of h and g ($hg_i = \sqrt{h_i g_i}$). Obviously, $h_i \leq hg_i \leq g_i$. Other measures suggested to overcome this shortcoming of the h-index include, among many others (Todeschini and Baccini 2016): (1) the m-index, which calculates the median of the ordered citations of the papers that compose the h-core (Bornmann et al. 2008); (2) the e-index advanced by Zhang (2009) and defined as $e_i = \sqrt{\sum_{p=1}^{h} C_{ip} - h_i^2}$, in which p represents papers and C means citations; (3) the w-index (Wu 2010), corresponding to the highest number of papers (w) receiving at least $10w$ citations each while the remaining papers receive at most $10(w+1)$. Recently, Perry and Reny (2016) have proposed the Euclidean index, defined as the Euclidean norm of the citation vector (considering the full list of citations of the author).

Beyond the issue of how to take citations into account, we should keep in mind that the number of citations is an imperfect impact measure. First, as highlighted by Ball (2005), citations need context. A paper can be cited as a core element for subsequent work, merely as a peripheral contribution or even in negative sense. Second, there are human problems in citations. These include the possibility that authors prefer to cite papers from friends and close colleagues or from the editor of the journal that he/she is submitting the paper to, and papers published in the same journal (Dunnick 2017). Finally, we need to address the topic of self-citations. The argument favoring exclusion of self-citations points out that they do not reflect the impact of the paper and that they can be used by authors to inflate their perceived performance (Van Raan 2006). Nevertheless, there are also valid reasons supporting the inclusion of self-citations (Glänzel et al. 2006). First, they avoid the repetition of previous material. Second, it is reasonable to assume that previous works by the author are key inputs for further developments in his/her research. If this is the case, self-citations are not only acceptable, but even necessary.

Fields of research

Getting a fair evaluation of a group of authors is an intricate task because, in several dimensions, those researchers are not directly comparable. One critical aspect is the coexistence of different fields of research with distinct publishing patterns. Attempts have been made to advance methods allowing inter-field comparisons. Mazurek (2018) simply proposes to divide the h-index of the author by the maximum h in that specific field, obtaining an index ranging between 0 and 1 which expresses the "distance to the top of the field". This procedure is appealing but not without problems since it depends on a single term of reference. Alternatively, Iglesias and Pecharromán (2007) divide the h-index by the average number of citations per paper.

A different but connected issue was raised by Amjad and Daud (2017). They highlight that some authors work in two or even more areas. When, for example, we try to select a researcher for an economics position, what are we looking for: the best scientist (considering all publications), or the best economist (only considering the publications in economics)? Moreover, we can generalize this question considering intra-field differences. If the position is for a health economist, should we value the publications in energy or transport economics? There is no definitive answer to these questions because it depends on the objectives inherent to each selection process.

Academic age

One of the most intensively discussed aspects in bibliometric analysis when the purpose is to conduct an author-level performance evaluation regards the comparison of authors in different stages of their careers. Who is the better author: one with the best stock of publications or one with the highest ratio of publications per time period? The h-index and the majority of the measures proposed to solve its limitations assume the first perspective, favoring authors with longer careers. Seeing this as a potential shortcoming, three alternative approaches have been suggested. The first was proposed by Hirsch himself—the m quotient—and simply divides the h-index by the academic age of the author $\left(m_i = \frac{h_i}{AA_i}\right)$. Alternatives to m include for example the α-index (Abt 2012) measuring the academic age of i in decades $\left(\alpha_i = \frac{h_i}{\mathrm{dec}_i}, \text{with } \mathrm{dec}_i = \mathrm{int}\left(\frac{\mathrm{AA}_i}{10}\right)\right)$. Other studies, belonging to a second approach, calculate the h-index confining the period under analysis to the last n years (Schreiber 2018). A final group of studies considers the entire career but gives higher weights to outputs obtained in recent years. The discounted cumulated impact index (DCI index) is an example of this approach (Järvelin and Persson 2008).

Number and sequence of authors

In their *Handbook of Bibliometric Indicators*, Todeschini and Baccini (2016) dedicate 21 pages to an overview of a wide set of coauthorship-weighted indices. The importance of this topic derives from the well documented increase in the average number of authors per paper (Wuchty et al. 2007; Frandsen and Nicolaisen 2010) and opens the debate on the correct way to give each author the fair credit for the publication and the corresponding citations (Liu and Fang 2012).[2]

Defining the position of the author in the paper's list of authors as k $(k = 1, 2, \ldots, K)$, we need to assign a weight $p(k)$ to each author. An extensive discussion on this question is far beyond the purpose of the present section, namely because the co-authorship weighting schemes that take into account the position of the author in the byline are less useful in sciences, such as economics, where the alphabetical order of the names is largely dominant (Frandsen and Nicolaisen 2010). Then, we confine ourselves to some well-known examples. The extreme cases are the standard counting and the straight counting. While the first one—full credit to all authors, i.e., $p(k) = 1, \forall k$—is commonly applied (including by the h-index), the second gives the full credit of the paper to the first author $(p(1) = 1 \wedge p(k) = 0$, for $k = 2, 3, \ldots., K)$. Two intermediate schemes are also often considered: (1) the uniform counting, which assigns the same weight to all authors, as in the standard counting, but the sum of these weights is one—$p(k) = \frac{1}{K}, \forall k$; (2) the proportional counting, in which the position of each author is relevant: $p(k) = \frac{1}{k}$.

[2] For a discussion of the main reasons justifying the increasing role of collaborative work in research, see Leahey (2016). Henriksen (2018) focuses on the specific case of economics.

A new approach for the measurement of publishing performance

Publishing performance index

We propose quantifying the performance of each author i ($i = 1, 2, \dots, I$) through the publishing performance index (PPI_i). The number of authors (I) corresponds to the relevant candidates in a given evaluation. In its simplest version, two performance dimensions are considered: number of papers published and number of citations received.[3] NP_i and NC_i are the number of papers and the number of citations of author i, respectively. In order to make these dimensions comparable, we capture them through NP_i and $\sqrt{\mathrm{NC}_i}$. Then, we calculate the share of author i in the total of the group under scrutiny in terms of these two variables (p_i and c_i, respectively):

$$p_i = \frac{\mathrm{NP}_i}{\sum_{i=1}^{I} \mathrm{NP}_i} \tag{1}$$

$$c_i = \frac{\sqrt{\mathrm{NC}_i}}{\sum_{i=1}^{I} \sqrt{\mathrm{NC}_i}}. \tag{2}$$

Obviously, $0 \le p_i, c_i \le 1$.

PPI_i is expressed as follows:

$$\mathrm{PPI}_i = \alpha\left(p_i - \frac{1}{I}\right) + (1-\alpha)\left(c_i - \frac{1}{I}\right); \quad 0 \le \alpha \le 1. \tag{3}$$

The term $\left(p_i - \frac{1}{I}\right)$ captures, for author i, the difference between his/her share in the total number of papers of the group $\left(p_i\right)$ and the share corresponding to the equal distribution $\left(\frac{1}{I}\right)$. If the number of papers published by author i is greater than the average of the group ($\mathrm{NP}_i > \frac{\sum_{i=1}^{I} \mathrm{NP}_i}{I}$) then $\left(p_i - \frac{1}{I}\right) > 0$. Regarding citations, when $\frac{\sqrt{\mathrm{NC}_i}}{\sum_{i=1}^{I} \sqrt{\mathrm{NC}_i}} > \frac{1}{I}$ we obtain $\left(c_i - \frac{1}{I}\right) > 0$.

The parameter α captures the weight given to the dimension "papers" in the global evaluation of performance, while the importance of "citations" is given by $(1-\alpha)$. The value of α (as well as the remaining evaluation criteria discussed below) must be pre-defined by the evaluators according to the objectives of each evaluation process.

PPI_i is positive when the author is above the average of the group (in the combined analysis that, in this simplest version of the index, accounts for papers and citations) and negative when the author has a level of performance below the average of the group.

Publishing performance box and profiles

Based on PPI_i, we build a visual tool that helps to see the position of all authors—publishing performance box (*PPB*). In order to illustrate the application of the *PPI* and the

[3] While the number of papers is excluded from several author-level performance measures, there are valid reasons to include this dimension, as discussed by Hausken (2016).

Table 1 Publishing performance index—an example

Authors ($I = 10$)	NP_i	NC_i	Performance			Profile/ area of the box
			Papers	Citations	PPI_i	
1	20	30	Below average $p_1 = 0.041;\ p_1 - \frac{1}{I} = -0.059$	Below average $c_1 = 0.017; c_1 - \frac{1}{I} = -0.083$	−0.071 (9)	[E]
2	35	2500	Below average $p_2 = 0.072;\ p_2 - \frac{1}{I} = -0.028$	Above average $c_2 = 0.157; c_2 - \frac{1}{I} = 0.057$	0.014 (4)	[D]
3	2	2000	Below average $p_3 = 0.004;\ p_3 - \frac{1}{I} = -0.096$	Above average $c_3 = 0.140; c_3 - \frac{1}{I} = 0.040$	−0.028 (7)	[H]
4	60	100	Above average $p_4 = 0.123;\ p_4 - \frac{1}{I} = 0.023$	Below average $c_4 = 0.031; c_4 - \frac{1}{I} = -0.069$	-0.023 (6)	[G]
5	100	400	Above average $p_5 = 0.205;\ p_5 - \frac{1}{I} = 0.105$	Below average $c_5 = 0.063; c_5 - \frac{1}{I} = -0.037$	0.034 (3)	[C]
6	0	0	Below average $p_6 = 0;\ p_6 - \frac{1}{I} = -0.1$	Below average $c_6 = 0; c_6 - \frac{1}{I} = -0.100$	−0.100 (10)	[E]
7	30	1200	Below average $p_7 = 0.062;\ p_7 - \frac{1}{I} = -0.038$	Above average $c_7 = 0.109; c_7 - \frac{1}{I} = 0.009$	−0.015 (5)	[H]
8	120	6500	Above average $p_8 = 0.246;\ p_8 - \frac{1}{I} = 0.146$	Above average $c_8 = 0.253; c_8 - \frac{1}{I} = 0.153$	0.150 (1)	[A]
9	110	2000	Above average $p_9 = 0.226;\ p_9 - \frac{1}{I} = 0.126$	Above average $c_9 = 0.140; c_9 - \frac{1}{I} = 0.040$	0.083 (2)	[B]
10	10	800	Below average $p_{10} = 0.021;\ p_{10} - \frac{1}{I} = -0.079$	Below average $c_{10} = 0.089; c_{10} - \frac{1}{I} = -0.011$	−0.045 (8)	[F]
Total	487	15,530				

In the column for PPI_i, between brackets is the ranking of author i

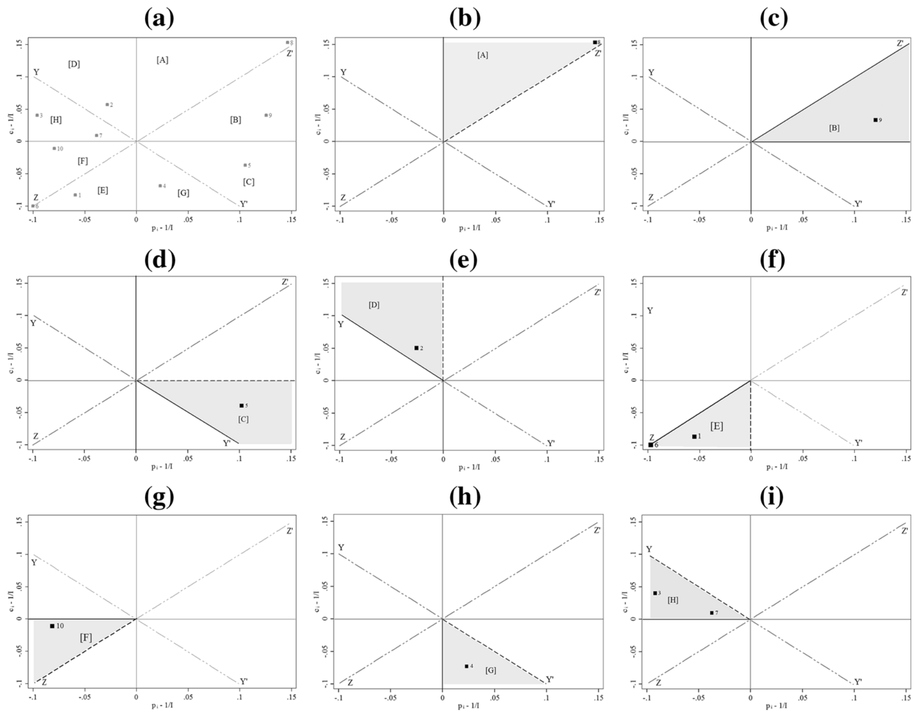

Fig. 1 Publishing performance box—an example

PPB, Table 1 and Fig. 1 consider a hypothetical example with ten authors. Figure 1, panel (a) presents the complete box while panels (b–i) show a more detailed visualization of each area of the box, including the definition of the borders between different areas.

The origin of the PPB corresponds to the case in which $p_i = c_i = \frac{1}{I}$, meaning that the author is equal to the average of the group in both dimensions (i.e., $\mathrm{NP}_i = \frac{\sum_{i=1}^{I}\mathrm{NP}_i}{I}$ and $\sqrt{\mathrm{NC}_i} = \frac{\sum_{i=1}^{I}\sqrt{\mathrm{NC}_i}}{I}$). At the right of the vertical axis are authors above the average in terms of papers while above the horizontal axis are authors above the average in terms of citations. Additionally, two reference lines are used. First, the line YY' separates the cases in which $\mathrm{PPI}_i \geq 0$ and those with $\mathrm{PPI}_i < 0$, assuming, in order to keep the discussion as simple as possible, $\alpha = 0.5$. This line crosses the origin and has slope -1. If only two authors are competing, that line connects the upper left and the lower right corners of the box. When the number of authors increases, the origin moves towards the lower left corner of the box since the equal distribution in terms of papers and citations corresponds to lower values in both axes. Second, the comparison between p_i and c_i allows the identification of the dimension (papers or citations) in which the author has advantage. This corresponds to the line ZZ' connecting the lower left and the upper right corners of the box. At the right (areas [B], [C], [E], and [G]) are authors with advantage in papers while at left (areas [A], [D], [F], and [H]) are authors with advantage in citations. Taking these elements together, we can distinguish eight different areas ([A] to [H]), each corresponding to a specific publishing profile.

Data

The group of authors under analysis is composed of those that are relevant in a given competition (e.g., the authors that apply for a position or those that are nominated for a prize). In the following sections we will illustrate the application of the framework introduced in "A new approach for the measurement of publishing performance" section. To that end, let us consider the context of a hypothetical prize for which faculty members of the top 10 world universities in the area of economics are eligible. In association with this prize, a complete ranking is divulged. We assume as starting point the QS World University Ranking, choosing "Economics & Econometrics" as subject. Considering the information provided by departmental websites in October 2018, tenure-stream or tenured faculty members with primary appointments in economics departments in the top 10 universities in this ranking were selected. This comprises 472 authors, distributed as follows: Harvard University (56); Massachusetts Institute of Technology (39); Stanford University (41); University of California, Berkeley (49); Princeton University (56); University of Chicago (34); London School of Economics and Political Science (49); University of Oxford (50); Yale University (47); and Columbia University (51).

The choice of the database to retrieve information about the authors in the sample is critical. As discussed by da Silva and Dobránszki (2018) and Martín-Martín et al. (2018), there are remarkable differences regarding coverage and reliability of the data contained in the three most frequently used options: Web of Science (Clarivate Analytics), Scopus (Elsevier), and Google Scholar. While the first two options include mainly papers published in scientific journals, Google Scholar has substantial additional coverage, including theses, books, book chapters, conference papers, and unpublished materials (Martín-Martín et al. 2018). This feature is an advantage for research evaluation in the area of humanities and social sciences, where, for example, books and book chapters tend to have greater importance than in natural and life sciences (Bornmann et al. 2016). Nevertheless, Google Scholar continues to have some important disadvantages that suggest that it should be used with some caution (e.g., the inclusion of duplicate entries, the fact that indicators can be easily manipulated, and that results are difficult to replicate; see Prins et al. 2016). Due to these reasons and given the time period covered, we opt to use Web of Science. Data on papers and citations were gathered from Web of Science (WoS) database (core collection) between 14 October and 19 November of 2018. Concerning the papers selected, only articles and reviews written in English were considered.

Having selected the source of data, we must pay attention to some well-known limitations of these databases: (1) the existence of several authors with the same name, notably in the case of common names; (2) spelling problems concerning the names of the authors; (3) the absence of the author in the list of authors of the paper; (4) authors that change the names used in their papers. All these problems occurred in the present case. They materialize a "precision problem" and imply, as mentioned by Schreiber (2018, p. 3), that "at present, an accurate citation database still requires to compare the publications in the citation record with a publication list of the author". Unfortunately, even this procedure is not a panacea for all problems because it assumes that the publications lists (provided by the authors) are available, up to date, complete, and correct (in terms of the number of authors and their names, title of the paper, year of publication, journal, etc.), which often does not happen.

To assure the validity of all data used, we executed a detailed comparison for each author between the list of papers given by WoS and the information on individual publications retrieved from departmental and personal websites. When necessary, complementary sources such as Econlit were also used. Once this process concluded, we obtain a database with 15,243 papers published between 1957 and 2018. These papers were published in 709 different journals and cited 1,288,803 times. Given the composition of the sample (authors belonging to world top universities), the strong level of concentration in high-ranked journals, with 50% of the papers published in only 14 top journals, is not surprising. Summing up the academic age of all authors, we reach a total of 10,076 years of work, giving therefore an average of 1.51 papers per year.

Publishing performance of world top economists

As discussed in "Background: methodological options and measures" section, any author-level performance measure (implicitly or explicitly) needs to assume a specific range of methodological options. We argue that these options should correspond to the evaluation criteria (EC) established for each specific selection process. Therefore, we start by assuming that the committee of the prize defined the simplest and most generic scenario, which includes three evaluation criteria concerning the objective of the selection process (EC1 to EC3), five evaluation criteria related to the operationalization of the measurement procedure (EC4 to EC8) and one criterion related to the scientific outputs considered (EC9). Subsequent sections will illustrate the impact of using different sets of criteria. The baseline scenario is defined as follows: (EC1—academic age): all career in accumulated terms; (EC2—dimensions of performance): number of papers and number of citations; (EC3—fields of research): all fields; (EC4—weights of papers and citations): $\alpha = 0.5$; (EC5—measurement of citations): all included, equal weights; (EC6—distribution of citations): ignored; (EC7—self-citations): included; (EC8—number and sequence of authors): standard counting. Finally, (EC9) is defined as: papers (articles and reviews) published in scientific journals included in the Web of Science (core collection). The results obtained assuming this scenario are shown in Table 2, Fig. 2 (complete PPB), and Fig. 3, providing, in this last case, a detailed visual perception of what happens in each of the areas of the PPB.[4]

[182 authors (out of 472) obtain $\mathrm{PPI}_i \geq 0$. 148 of them belong to the area [A + B] in the PPB, where $\left(p_i - \frac{1}{I}\right) \geq 0$ and $\left(c_i - \frac{1}{I}\right) \geq 0$. The top 20 comprises authors from 7 different American universities. Joseph Stiglitz occupies the 1st position due to his advantage in terms of papers. He has published until now 247 papers. Concerning citations, he reaches the 3rd position among the authors under analysis. His level of performance is summarized in the PPI_i score ($\mathrm{PPI}_i = 0.0109$). The equal distribution ($\frac{1}{I}$) corresponds, in our case ($I = 472$), to 0.0021. The PPI_i score is the weighted average between: (1) the difference between the share the author has in the total of the group in the dimension "papers" and the share that corresponds to the equal distribution; (2) a

[4] The example discussed in this Section includes mainly top researchers. When the analysis considers medium-to-low researchers, some differences may emerge. Since the differences among them in terms of scientific outputs are probably lower, the PPI index will produce more approximate values. This makes even more important the consideration of additional evaluation criteria. This procedure can be developed in two different ways. First, in the context of our framework, namely through additional rounds with new criteria for the group of authors with higher levels of scientific performance. Second, through the consideration of qualitative elements (peer review). It seems fair to say that the role of the evaluators is even more important when the group under analysis is more homogeneous.

Table 2 Publishing performace—top 20

Ranking	Author	PPI_i	$\frac{PPI_i}{Max\ PPI_i} * 100$	NP_i	NC_i	*h*-index	PPB area
1	Joseph Stiglitz	0.01,090	100	247 (1)	32,136 (3)	75 (3)	[B]
2	Andrei Shleifer	0.00995	91.351	159 (7)	62,611 (1)	91 (1)	[A]
3	Peter Phillips	0.00988	90.663	241 (2)	22,296 (5)	54 (7)	[B]
4	James Heckman	0.00897	82.332	183 (4)	34,457 (2)	82 (2)	[B]
5	Daron Acemoglu	0.00705	64.677	165 (6)	18,761 (7)	61 (4)	[B]
6	Martin Feldstein	0.00656	60.184	186 (3)	8830 (42)	51 (12)	[B]
7	John List	0.00646	59.285	178 (5)	9994 (29)	53 (9)	[B]
8	Robert Barro	0.00537	49.298	100 (27)	23,604 (4)	54 (7)	[A]
9	David Cutler	0.00525	48.178	142 (8)	9781 (31)	52 (10)	[B]
10	Edward Glaeser	0.00516	47.376	115 (14)	16,393 (9)	56 (5)	[B]
11	Alberto Alesina	0.00456	41.816	97 (30)	16,250 (10)	56 (5)	[A]
12	Dale Jorgenson	0.00452	41.475	126 (10)	8354 (44)	40 (36)	[B]
13	Lawrence Summers	0.00448	41.146	108 (18)	12,464 (17)	46 (21)	[B]
14	James Stock	0.00444	40.718	85 (36)	18,899 (6)	44 (24)	[A]
15	Amartya Sen	0.00431	39.599	109 (16)	10,877 (23)	51 (12)	[B]
16	Elhanan Helpman	0.00431	39.534	98 (29)	13,737 (14)	42 (28)	[B]
17	Jeffrey Sachs	0.00430	39.450	108 (18)	11,004 (22)	47 (17)	[B]
18	Philippe Aghion	0.00425	39.037	101 (24)	12,463 (18)	49 (15)	[B]
19	Drew Fudenberg	0.00423	38.781	110 (15)	9969 (30)	50 (14)	[B]
20	Alan Krueger	0.00414	37.984	91 (32)	14,262 (12)	44 (24)	[A]
Whole sample							
Average values		0		32.3	2730.5		
Correlation between PPI_i and other measures				0.9671	0.8576	0.9756	

In the columns for the number of papers, the number of citations, and the *h*-index, between brackets are the rankings of author *i* in each of these dimensions

similar difference in terms of citations. Joseph Stiglitz has, on average, 1.09 p.p. more than the value that corresponds to the equal distribution (obtained from the simple average between 0.0141—the term (1) above—and 0.0077—the term (2) above). The following positions in the ranking are occupied by Andrei Shleifer (7th position in papers and 1st in citations), Peter Phillips (2nd in papers and 5th in citations), James Heckman (4th in papers and 2nd in citations), and Daron Acemoglu (6th in papers and 7th in citations). Due to these different sources of performance, 2 authors of the top 10 are in area [A] in the PPB, while 8 are in [B]. The 182 authors with $PPI_i \geq 0$ are distributed as follows: 67 authors in [A], 81 in [B], 10 in [C], and 24 in [D].

The evidence presented in Table 2 emphasizes the low granularity problem of the *h*-index. Analyzing the full sample, we verify that 324 authors have an h-index between 0 and 20. The most frequent value is $h = 1$ (31 authors), followed by $h = 3$ (26 authors), while only 13 authors have an *h*-index higher than 50. The elimination of this problem is an important advantage of the PPI.

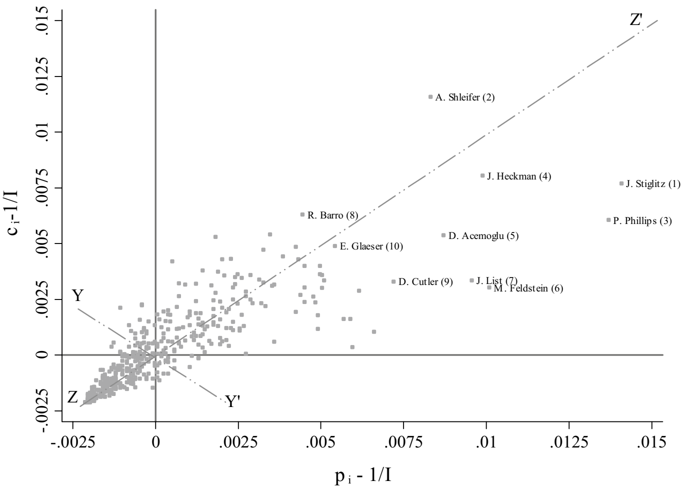

Fig. 2 Publishing performance box—overall sample. *Note* Between brackets are the rankings of the top 10 authors

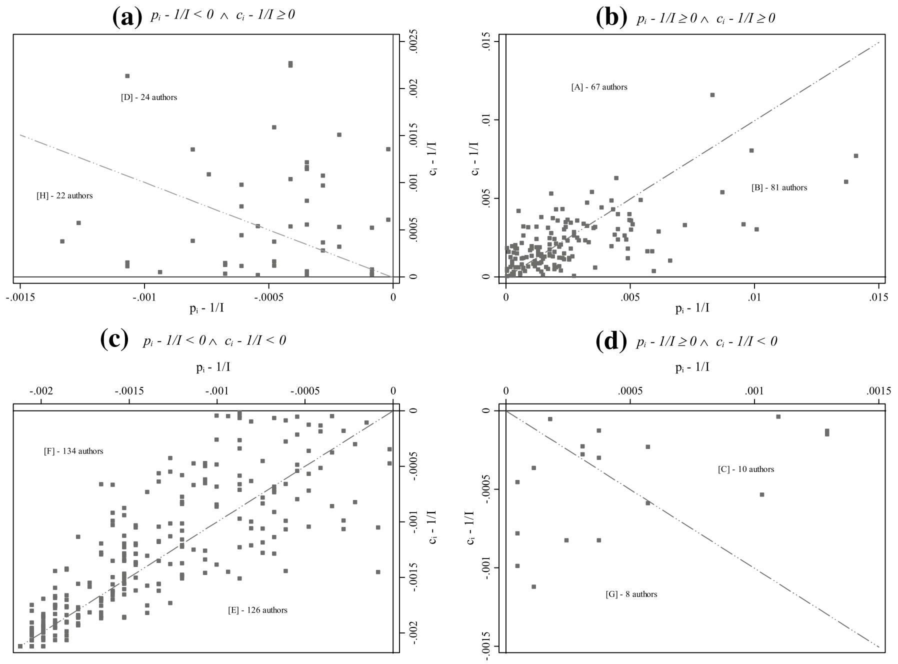

Fig. 3 Publishing performance box—an analysis by area

When bibliometric measures are used to evaluate author's performance, as important as the value of the index that ranks the competing authors is the definition of the author profile. The analysis of the PPB allows us to take some steps forward concerning the definition of such profiles. Table 3 shows additional information on this topic.

Table 3 Publishing profiles

Profiles	No. authors	NP	NC	Academic age		PPI rank			% women
				Average	Interval	Average	Interval	Best author	
[A]	67 (14.2%)	54.9	8259.6	31.8	14–52	84.1	2–177	Andrei Shleifer	7.5
[B]	81 (17.2%)	85.0	6288.1	36.6	15–62	69.5	1–171	Joseph Stiglitz	6.2
[C]	10 (2.1%)	42.7	1224.7	28.0	14–47	164.5	139–182	Benjamin Friedman	10.0
[D]	24 (5.1%)	26.4	3524.2	28.4	14–44	156.0	121–179	Xavier Sala-i-Martin	16.7
[E]	126 (26.7%)	11.2	183.1	10.8	0–52	357.2	197–456	Fernando Alvarez	15.1
[F]	134 (28.4%)	10.7	460.0	15.0	2–56	326.0	192–444	Mark Duggan	23.1
[G]	8 (1.7%)	35.3	649.0	25.3	13–45	208.6	186–234	Stephen Broadberry	0
[H]	22 (4.7%)	23.0	1788.6	20.5	12–39	202.4	183–232	Muriel Niederle	9.1
Whole sample		32.3	2730.5	21.4	–	–	–	Joseph Stiglitz	14.2

Four interesting conclusions can be highlighted. First, as expected, there is on average a clear advantage of the authors represented in [A] and [B], as shown by several distinct indicators: more papers, more citations, higher average ranking position, and higher ranking of the authors with best and poorest performance. Second, authors in these areas have long careers. This is not surprising since we are evaluating performance considering the full extension of those careers, as defined in (EC1). While for several purposes this is the adequate perspective, there are other valid ways to address the question "who is the best author in terms of publishing performance?" We will explore other perspectives in "Academic age and performance" section. Third, in area [E] is a vast group of young authors. Therefore, it is not surprising that they have the poorest average ranking position (357.2 in a total of 472 authors). Interestingly, Raj Chetty, Yuriy Gorodmichenko, Benjamin Olken, and Parag Pathak are the youngest authors (AA = 14) with $PPI_i \geq 0$. It is noteworthy that we can use the approach developed in this study to detect "emerging stars". We can do that by confining the analysis to a given academic age interval. For example, concerning authors with academic age between 0 and 10, the highest PPI_i score is obtained by Pascaline Dupas. Fourth, regarding the distribution by gender, females represent only 14.2% of the total (67 authors). 19 of them are in [E] and 31 in [F]. Only 15 female authors exhibit a positive performance score (8.2% of 182 authors in the same situation), with the highest rankings obtained by Janet Currie (area [B], 28th position), Esther Duflo (area [A], 61st position), Serena Ng (area [A], 84th position), Claudia Goldin (area [B], 92nd position), and Graciela Chichilnisky (area [B], 96th position). This evidence is partially explained by the significant difference between men and women in terms of academic age (22.2 for men and 16.5 for women) but confirms previous results pointing to the existence of a gender gap in terms of scientific performance as measured by the number of papers, the number of citations, the h-index, and the quality of the journals (e.g., Larivière et al. 2013; Mayer and Rathmann 2018).

Testing alternative evaluation criteria

In the previous section we assumed a baseline scenario. Now, we perform a battery of tests concerning the evaluation criteria related to the measurement procedure (EC4–EC8), aiming to show the influence of those options on the final results (Table 4).

The parameter α, defining the weight of the dimension "papers", is a key element of the analysis conducted so far. (EC4) assumes $\alpha = 0.5$. When other values are assumed, some effects are evident. For example, with $\alpha = 0.75$, 14 authors change their position in the PPB, namely from [G] to [C] (2 authors) and from [D] to [H] (12 authors), leading to changes of sign in PPI_i. The sum of the variations, in absolute value, in the ranking positions reaches 4.1% of the possible maximum ($\frac{I^2}{2} = 111,392$, corresponding to the complete inversion of the ranking order). Marc Melitz is the most penalized author with this modification because he has a negative score in $\left(p_i - \frac{1}{I}\right)$, resulting from the comparison between his 16 papers and the average value of 32.3 in the whole sample, and a significant positive score in $\left(c_i - \frac{1}{I}\right)$.

As discussed in "Citations" section, the measurement of citations has been a critical issue in the literature. In the analysis conducted in "Publishing performance of world top economists" section we assumed that all citations are valued equally (EC5) and ignored their distribution among the papers of the author (EC6). The results in Table 4 allow us to evaluate alternative assumptions. Starting with the question of top papers, defined, as proposed by Fu et al. (2012), as papers with at least 500 citations, we verify that 422 papers in our sample reach that threshold (2.8% of the total), with the maximum value being obtained by James Heckman with his paper published in *Econometrica* in 1979, which until now has received 9748 citations. In order to test the influence of this element, we create a new scenario in which the number of citations of papers with 500 or more citations is doubled. Due fundamentally to his paper about heterogeneous firms published in 2003 (with 2969 citations), Marc Melitz is the author with the highest increase in the corresponding ranking position when we follow this criterion (28 positions). In overall terms, 58 authors change their position in the PPB. The second element that must be considered is the distribution of citations among the different papers of the author. This aspect is critical in the h-index and in several alternative measures. Aiming to address this issue, we introduce the number of citations adjusted by distribution, which is obtained as:

$$\mathrm{NC}_i^a = \mathrm{NC}_i E_i \tag{4}$$

with

$$E_i = 1 - \frac{1}{\beta} \sum_{p=1}^{P} \left| s_{pi} - \frac{1}{P} \right|; \quad \beta \geq 1; \quad E_i \in \left[\frac{1}{P}, 1 \right] \tag{5}$$

where s_{pi} is the share of paper p in the total in terms of the number of citations received.

When all papers have the same number of citations (perfect distribution), $\mathrm{NC}_i^a = \mathrm{NC}_i$. The existence of inequality in the distribution of citations introduces an adjustment in the number of citations. The higher the value of β, the lower the adjustment considered. In this

Table 4 The impact of alternative evaluation criteria

Evaluation criterion		Highest rise in the ranking	Highest fall in the ranking	Sum of changes in ranking positions	No. of area changes in the PPB	No. of changes to $PPI_i \geq 0$	No. of changes to $PPI_i < 0$
Code	Scenario						
(EC4)	$\alpha = 0.25$	Alwyn Young (40)	Graciela Chichilnisky (43)	4644 (4.2%)	15	9	6
(EC4)	$\alpha = 0.75$	Taisuke Otsu (44)	Marc Melitz (56)	4566 (4.1%)	14	2	12
(EC5)	Higher weight to papers with citations ≥ 500	Marc Melitz (28)	Richard Freeman and Henry Farber (11)	1424 (1.3%)	58	2	7
(EC6)	Considering the distribution of citations ($\beta = 2$)	Donald Davis (18)	Richard Clarida (29)	1813 (1.6%)	40	4	2
(EC7)	Excluding self-citations	Jesse Rothstein and John Asker (4)	Fuhito Kojima (7)	289 (0.3%)	9	1	0
(EC8)	Uniform counting	Han Hong – Stanford (128)	Parag Pathak (62)	7361 (6.6%)	112	9	12

In columns "Highest rise in the ranking" and "Highest fall in the ranking", the numbers between brackets correspond to the number of ranking positions (absolute value) by which the author went up or down in the ranking after taking into account the new evaluation criterion. From the comparison of the rankings obtained with the baseline and alternative scenarios, we calculate the indicator "Sum of changes in ranking positions" which gives the sum of the changes in ranking positions involving the 472 authors of the sample (absolute values). Between brackets, we identify the percentage of these changes in comparison to the possible maximum ($\frac{I^2}{2} = 111,392$). Finally, the column "No. of changes to $\mathrm{PPI}_i \geq 0$" reports the number of authors that in the baseline scenario had $\mathrm{PPI}_i < 0$ while in the new scenario have $\mathrm{PPI}_i \geq 0$. "No. of changes to $\mathrm{PPI}_i < 0$" accounts for the opposite shifts study we assume $\beta = 2$. In this scenario, Richard Clarida is the author with the greatest fall in the ranking (29 positions). This occurs because 65% of his citations are obtained by only two papers.

A final issue concerning citations is the inclusion/exclusion of self-citations (EC7). The importance of this question is small in our sample since the average share of self-citations in total citations is only 1.8%. As a consequence, the impact of this modification is limited, with only 9 authors changing position in the box, and only 1 changing his/her sign in the PPI_i.

On the contrary, the impact resulting from a different evaluation criterion concerning the number of authors is the most notable among all evaluated in Table 4. This is not surprising since the incapacity of some performance measures, including the h-index, to account for different co-authorship patterns is one of their most remarkable shortcomings (Hirsch 2019). The average number of authors per paper in our sample is 2.28. Comparing the scenario that considers the uniform counting method with the one defined by (EC8), 21 authors register changes of sign in PPI_i and 112 change their position in the PPB.[5] The modifications in the ranking correspond to 6.6% of the maximum possible value.

Taking the above discussion into account, two main lessons should be retained. First, the methodological options inherent to the different indices matter for final results, namely when we need to make decisions concerning individual authors. In these cases, any change in ranking positions is important. If the ranking is defined using measures that do not fully correspond to the requirements established by the evaluators, unfair decisions may be made. Second, since the adequate methodological options are case-specific and must correspond to the objectives at play in each selection process, rigid measures in terms of their assumptions are not the optimal solution. Instead, we argue that we need a flexible approach, able to accommodate different evaluation criteria without changing the framework of analysis. This is a remarkable advantage of the framework proposed in this study.

Academic age and performance

Who can be considered the best author: (1) author A with 20 papers, 200 citations, and $\text{AA}_A = 40$; or (2) author B with 10 papers, 100 citations, and $\text{AA}_B = 10$? In the baseline scenario for (EC1), the answer is A since we are measuring the accumulated performance. However, analyzing the same data per unit of time, the answer is B. In this section, we explore the impact of academic age on performance.

Following the logic of the m quotient, (EC1—academic age) is now defined as follows: outputs (papers and citations) per year. This new scenario is a different way of assessing publishing performance and can easily be incorporated in the methodology presented in "A new approach for the measurement of publishing performance" section. First, for each author, we calculate $\frac{\text{NP}_i}{\text{AA}_i}$ and $\frac{\sqrt{\text{NC}_i}}{\text{AA}_i}$. Then, we obtain the new values of p_i and c_i. Finally, we calculate PPI_i. Table 5 shows some key results.

[5] We test the uniform counting because the position of the authors in the byline is irrelevant when the alphabetical order of the names is the rule followed by a vast majority of authors and papers. This is the case of economics, in which the byline of around 90% of the multi-authored papers follow the alphabetical order (Kadel and Walter 2015). We conduct a similar analysis for our sample (11,230 multi-authored papers) and find a roughly similar value (91.38%).

Table 5 Publishing performace—top 20 (intensity)

Ranking	Author	PPI_i	$\frac{PPI_i}{Max\ PPI_i} * 100$	NP_i	NC_i	*PPB* area	Ranking in the baseline scenario
1	John List	0.00780	100	8.5 (1)	475.9 (9)	[B]	7
2	Andrei Shleifer	0.00644	82.583	4.7 (5)	1841.5 (1)	[A]	2
3	Daron Acemoglu	0.00642	82.301	6.3 (2)	721.6 (3)	[B]	5
4	Edward Glaeser	0.00440	56.393	4.3 (7)	607.1 (5)	[B]	10
5	Peter Phillips	0.00407	52.251	5.1 (3)	474.4 (10)	[B]	3
6	Joseph Stiglitz	0.00395	50.622	4.8 (4)	618.0 (4)	[B]	1
7	James Heckman	0.00384	49.250	4.1 (8)	765.7 (2)	[B]	4
8	David Cutler	0.00364	46.723	4.6 (6)	315.5 (24)	[B]	9
9	Emmanuel Saez	0.00350	44.940	3.5 (14)	334.1 (20)	[B]	68
10	Jonathan Gruber	0.00306	39.259	3.9 (9)	259.0 (42)	[B]	31
11	Alberto Alesina	0.00292	37.386	3.0 (24)	507.8 (7)	[A]	11
12	Victor Chernozhukov	0.00277	35.552	3.7 (11)	146.2 (92)	[B]	102
13	Alan Krueger	0.00275	35.308	2.9 (29)	460.1 (11)	[A]	20
14	Esther Duflo	0.00273	35.013	2.7 (31)	365.2 (18)	[A]	61
15	Tim Besley	0.00268	34.321	3.5 (13)	282.6 (33)	[B]	21
16	Raj Chetty	0.00258	33.141	2.6 (37)	228.6 (52)	[A]	138
17	Steve Levitt	0.00252	32.320	3.0 (23)	283.4 (32)	[B]	54
18	Janet Currie	0.00250	32.102	3.5 (12)	224.1 (54)	[B]	28
19	Guido Imbens	0.00250	32.089	2.6 (35)	408.6 (15)	[A]	42
20	John Van Reenen	0.00242	31.076	3.1 (19)	253.8 (44)	[B]	51
Whole sample—average		0		1.5	90.5	–	–

In the columns for the number of papers and the number of citations, between brackets are the rankings of author *i* in each of these dimensions

John List occupies the 1st position in this ranking due to his advantage in terms of papers. His 8.5 papers per year correspond to 134.9% of the number of papers per year obtained by Daron Acemoglu (2nd position in this dimension). Regarding the number of citations, he occupies the 9th position with 475.9 citations per year. Comparing this new evidence with that obtained in "Publishing performance of world top economists" section, we detect, as expected, significant differences. 61 authors improve their position in the ranking by more than 100 positions, with the highest gains realized by Stefanie Stantcheva (256 positions; AA = 5) and Nathaniel Hendren (248 positions; AA = 6). Considering the top 20, six cases are particularly interesting since they are below position 50 in accumulated terms but appear in the top 20 when the analysis is done per unit of time: Emmanuel Saez, Victor Chernozhukov, Esther Duflo, Raj Chetty, Steve Levitt, and John Van Reenen.

Figure 4 establishes the link between the PPI obtained in terms of intensity (outputs per year) and the academic age. Figure 5 does the same for the baseline scenario (accumulated terms). In both cases, we represent two lines: (1) the overall average of PPI (which is 0, by definition); (2) the average of the specific academic age category.

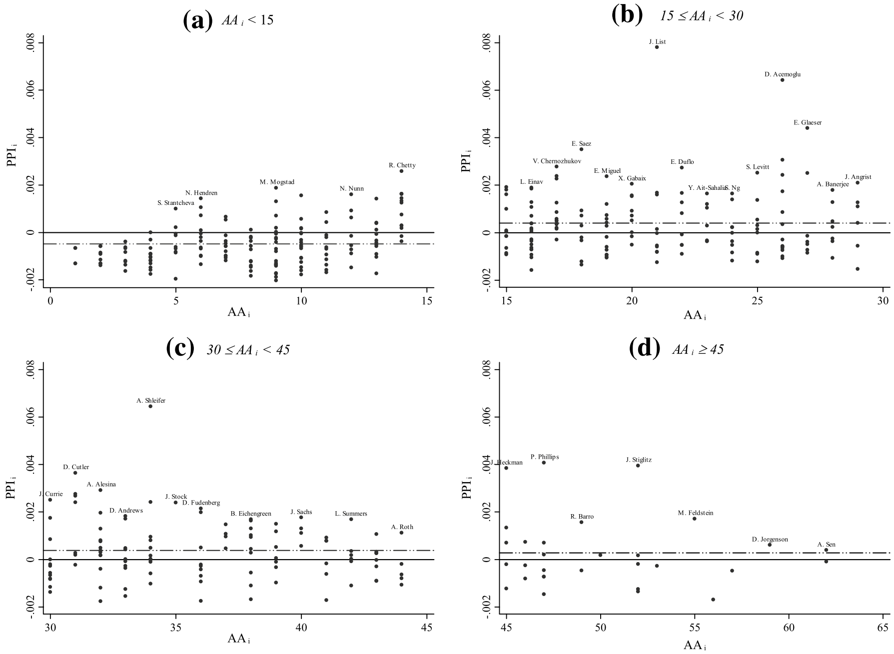

Fig. 4 PPI (internsity terms) and academic age. *Note* The dotted line in each of the graphs is the average PPI_i for that group

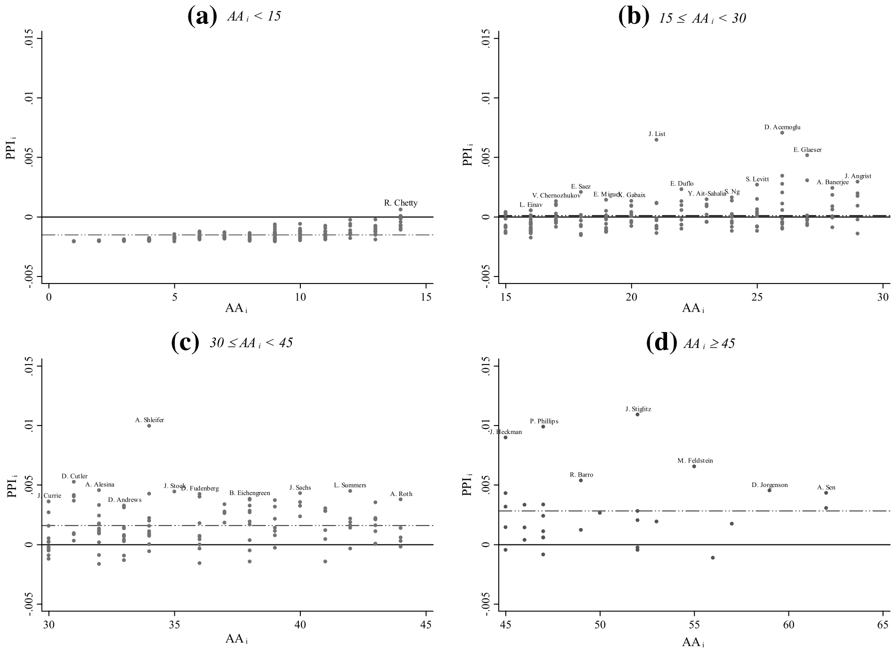

Fig. 5 PPI (accumulated terms) and academic age. *Note* The dotted line in each of the graphs is the average PPI_i for that group

Considering the results in Fig. 4, the youngest authors with PPI ≥ 0 are Stefanie Stantcheva and Matteo Maggiori with AA = 5. Despite this result in terms of intensity, their short academic age do not allow them to obtain PPI ≥ 0 in accumulated terms. As already mentioned in "Publishing performance of world top economists" section, the youngest authors obtaining a PPI ≥ 0 in the baseline scenario have AA = 14. In the next academic age category (15–29), there are several authors with high levels of performance per year. These results are in line with the idea that the second and third decades of work are the ones with the highest number of papers per year. The evidence regarding this indicator is clear: 1.41 papers/year in the first decade of work, 1.69 in the second, 1.52 in the third, 1.50 in the fourth, 1.41 in the fifth, and 1.12 in the sixth. Several authors classified in top ranking positions belong to the academic age category 15-29, namely Emmanuel Saez (AA = 18), John List (AA = 21), Daron Acemoglu (AA = 26), and Edward Glaeser (AA = 27). The last three of these are also in the top 10 in the baseline scenario. In the third academic age category, it is possible to highlight the performance of David Cutler, Alberto Alesina, Alan Krueger, Tim Besley, Janet Curie, and above all Andrei Shleifer, all with AA $\leq$ 35. As age increases, it would be expectable to see decreases in the level of performance per year. Therefore, the persistence of high-level performance by, for example, Jeffrey Sachs, Lawrence Summers, and Alvin Roth is remarkable. Even more noteworthy – given their academic age and the values of PPI obtained—are the results for James Heckman, Peter Phillips, and Joseph Stiglitz, all of them with AA $\geq$ 45. They are in the top 10 in both accumulated terms and outputs per year.

The evidence in Fig. 5 allows us to identify who is the best among those starting their careers in a given year. For example, Robert Barro is the best among those starting in 1970 (AA = 49). Extending this evaluation for decades, the best authors in terms of publishing performance (among those considered in the sample) starting in each decade are: Dale Jorgenson (1951–1960), Joseph Stiglitz (1961–1970), Peter Phillips (1971–1980), Andrei Shleifer (1981–1990), Daron Acemoglu (1991–2000), Emmanuel Saez (2001–2010), and Nathaniel Hendren (2011–2018).

Papers and citations adjusted by quality

In the baseline approach ("A new approach for the measurement of publishing performance" section), PPI_i includes only two dimensions of performance: number of papers and number of citations. Let us consider now that the evaluators redefine (EC2) assuming that overall performance also depends on the quality of the journals in which the papers are published and the quality of the citations. Let us define NP_i' as the number of papers adjusted by quality (of the journals) published by author i. This first adjustment can be made using one of the metrics available to rank journals (impact factor, quartiles, several rankings). The share of author i in the total is:

$$p_i' = \frac{\text{NP}_i'}{\sum_{i=1}^{I} \text{NP}_i'}. \tag{6}$$

The second adjustment is introduced in the measurement of citations. As mentioned in "Background: methodological options and measures" section, the basic idea is that citations are not all equivalent, some of them give more credit to the author of the paper. We capture the quality of the citing papers of i through the average number of citations (per paper) obtained by them (indirect citations).[6] Based on that, we obtain the number of citations adjusted by quality ($NC_i^{'}$). Then, we use this information to calculate the weight of i in the total:

$$c_i^{'} = \frac{\sqrt{NC_i'}}{\sum_{i=1}^{I} \sqrt{NC_i'}}. \tag{7}$$

The publishing performance index adjusted by quality ($PPI_i^{'}$) is obtained replacing p_i by $p_i^{'}$ and c_i by $c_i^{'}$ in Eq. (3). The same adaptation can be introduced in the PPB.

As a consequence of these quality adjustments, two new evaluation criteria are needed. To keep the empirical application as simple as possible, we consider that the evaluators use the quartile of the journal as the relevant metric to capture the quality of the papers published in that journal: a paper published in a Q1 journal counts 1, in a Q2 journal counts 0.75, in a Q3 journal counts 0.5, and in a Q4 journal counts 0.25 (EC10—quality of papers). Regarding the quality of citations, the evaluators also consider quartiles, splitting the authors according to the number of indirect citations per citing document and applying the same scale (1, 0.75, 0.5, 0.25) to the number of citations (EC11—quality of citations).

Focusing the analysis on the 25% of authors with the most papers, Amy Finkelstein is the author with the highest ratio between the number of papers adjusted by quality and the total number of papers ($\frac{NP_i^{'}}{NP_i} = 0.99$) corresponding to 49 papers in Q1 journals and two in Q2 journals. Also interesting are the cases of David Cutler, Andrei Shleifer, and Daron Acemoglu. All of them combine a very high number of papers (top 8) with an average quality of those papers clearly above the average. The case of Andrei Shleifer is even more impressive because he manages to obtain high performance levels in the four dimensions considered.

In the context of this new scenario, we calculate $PPI^{'}$ and obtain the new PPB. When compared with the baseline scenario, we find that: (1) there are no changes in the top 10; (2) Fuhito Kojima is the author showing the highest increase in his position (54 rank positions); (3) the modifications in the ranking correspond to 3.1% of the maximum possible value; (4) 107 authors change their position in the PPB.

We conclude the empirical analysis with some further discussion on publishing profiles. Taking into account the four dimensions of performance, we are able to define a total of 16 profiles.[7] Each of them will be defined by a four letter code in the following sequence: NP_i, $\sqrt{NC_i}$, quality of papers (assessed through $\frac{NP_i^{'}}{NP_i}$), and quality of citations (evaluated through the number of citations per citing document). The letter A means equal or above average while B means below average. Thus, for example, an author classified with profile ABBB has a number of papers equal to or higher than average while registering a value below average in the remaining dimensions. Table 6 presents the key characterization elements of the authors belonging to each of these profiles.

Sixty-nine authors, including 12 of the top 20, have profile AAAA. They exhibit a large advantage in terms of citations (9224.6, on average), occupy the 2nd position in terms of papers published, and obtain a remarkable result in terms of the average quality of the

[6] Our sample comprises a total of more than 22 million indirect citations.

[7] This procedure can, of course, be extended to more than four dimensions.

Table 6 Publishing profiles (4 dimensions of performance)

Profiles	NP	NC	$\frac{NP'}{NP}$	Citations per citing document	No. of authors	PPI rank			AA
						Average	Interval	Best author	
AAAA	70.3	9224.6	0.92	24.7	69	66.6	2-165	Andrei Shleifer	34.2
AAAB	45.5	3119.9	0.93	12.0	12	126.4	68-177	Emmanuel Saez	18.3
AABA	81.1	6407.0	0.81	25.5	55	69.9	1-154	Joseph Stiglitz	39.9
AABB	58.9	3033.6	0.79	13.0	12	109.0	39-166	Rick Van der Ploeg	26.4
ABAA	52.0	1320.0	0.88	27.1	1	139	139-139	Benjamin Friedman	47.0
ABAB	40.3	1145.6	0.91	11.1	7	172.9	145-191	Liran Einav	22.0
ABBA	38.5	876.7	0.78	16.7	6	192.3	141-234	Frank Cowell	33.3
ABBB	36.0	710.0	0.76	9.3	4	202.8	182-218	Philip Reny	20.3
BAAA	23.3	2909.1	0.93	23.7	22	178.2	121-231	Xavier Sala-i-Martin	25.7
BAAB	25.3	2219.1	0.96	12.5	14	187.0	144-232	Marc Melitz	16.5
BABA	27.2	2886.2	0.84	27.5	10	166.0	135-209	Richard Clarida	33.5
BABB	–	–	–	–	–	–	–	–	–
BBAA	17.7	914.4	0.92	22.0	21	260.2	192-384	Mark Duggan	24.9
BBAB	9.8	289.6	0.95	5.9	115	346.3	202-452	Oriana Bandiera	9.4
BBBA	18.7	696.9	0.77	24.4	24	270.5	203-419	Wouter Den Haan	30.5
BBBB	10.7	186.4	0.75	6.1	83	351.3	237-452	Han Hong (Stanford)	12.5

Columns NP, NC, $\frac{NP'}{NP}$, and AA present the average values for each group

papers, with $\frac{NP_i'}{NP_i}=0.92$. The two profiles with more authors are BBAB (115 authors) and BBBB (83 authors), corresponding to the lowest values of academic age (9.4 and 12.5, respectively). They differ in the average quality of the papers and, probably as a consequence, in the number of citations. A final highlight goes to the performance of authors classified in the profile BAAB. Despite having a number of papers that are below average, which is a consequence of their low academic age, they have an advantage vis-à-vis the average in terms of citations, and the quality of the journals in which these papers are published is the highest in the sample.

Final remarks

The main contribution of this paper is the proposal of a new approach for the measurement of author-level publishing performance. We developed a toolkit including the publishing performance index and the publishing performance box and discussed publishing performance profiles.

The logic behind the approach developed in this study is different from that inherent to the dominant measures used to assess author-level scientific performance. A first characteristic of our approach is that it is conducted in relative terms. More specifically, it confines the analysis to the relevant candidates for each specific selection process, making a direct comparison between them. This characteristic makes this approach especially suitable for the cases in which the group of authors under analysis is pre-defined (i.e., the assessment process is conducted after the definition of the sample),[8] which is what occurs in many critical decisions in science, including, for example, hiring processes, promotions, and awards.

Obviously, as with any other method that aims to evaluate and rank authors, we need to define the relevant criteria. An advantage of our approach is that instead of a rigid measure with the necessary methodological options already assumed, we introduce a flexible framework in which the evaluators can define the criteria they wish to apply in that specific selection process. This is important since, as mentioned by Abramo and D'Angelo (2014, p. 1130), "performance (…) should be evaluated with respect to the specific goals and objectives to be achieved. Because objectives may vary across research institutions and along time, recommending a sole indicator of performance would be inappropriate". The same key idea is summarized in the second principle of the Leiden Manifesto: "no single evaluation model applies to all contexts" (Hicks et al. 2015, p. 430).

We are aware that this more active participation of the evaluators in the definition of the criteria that are used to rank authors requires an extra effort from the evaluators. More specifically, they must have more informed knowledge about the methodological elements associated with the criteria applied. Since the evaluators are usually not bibliometric experts but rather experts in their respective fields, probably they are not fully aware of the specialized discussion on the advantages and shortcomings of the measures used. However, in the approach proposed here the evaluators are active actors instead of passive users. They should define what dimensions are relevant for that selection process and their relative importance. Despite this additional effort required, we should bear in mind that many of the selection processes have profound implications for individuals and institutions, giving high priority to the fairness of the decisions and to their adequacy to the specific objectives to be achieved. Additionally, by making the evaluation criteria explicit, the degree of transparency of the selection process increases considerably, this way guaranteeing a more open and responsible use of metrics.

Beyond these elements, the approach proposed in this Section has some important advantages. First, in terms of calculation, it is extremely simple to obtain the final score (PPI) of each author. Second, the PPI is a high granularity measure, making it possible to generate a full ranking of the authors under scrutiny. Third, for $0 < \alpha < 1$, the full list of papers and citations of each author is taken into account.

Following the traditional bibliometric approach, the simplest version of our methodology confines the analysis to papers and citations. This can be extended in four different ways. First, we can consider additional dimensions of publishing performance that can be evaluated through adjustments in the dimensions already included. This was done in "Papers and citations adjusted by quality" section, in which we build a measure that also includes the quality of journals and the quality of citations. Second, other scientific outputs can be added to the evaluation (e.g., books, book chapters, reports, and conference papers) in order to widen the range of publications that are covered. Third, we can include new dimensions to capture other forms of impact of the scientific contributions. To that end, including some alternative metrics accounting for the number of times a publication has been tweeted, downloaded, bookmarked, or shared, for example, offers that possibility of

[8] The criteria used for the selection process should, of course, be defined a priori.

extension (Waltman and Costas 2014). The fourth possibility goes beyond the scope of assessing publishing performance and aims to capture other dimensions of scholarly activity such as the amount of intellectual property produced (e.g., patents, licenses, spin-offs), production of software and hardware, or the number of research awards.

The approach introduced in this study accommodates all of these potential extensions. While this is an advantage of the method, some aspects should be highlighted. First, it is important to stress that only quantitative dimensions can be included in the PPI. Second, the visual representation (PPB) is obviously restricted to only two dimensions of performance. In the present study we discussed not only the baseline scenario with papers and citations as relevant dimensions of performance but also an extended version with papers and citations adjusted by quality. This last case corresponds to the first kind of extension identified above. Third, the inclusion of additional dimensions requires the definition, by the evaluators, of the corresponding weights in the overall assessment. Fourth, a fair evaluation requires rigorous information and adequate metrics to measure the reality under analysis. Obviously, taking into account more dimensions of performance in order to enlarge the scope of what is captured in the PPI is a positive development. Nevertheless, there are issues to address that recommend prudency when taking steps in this direction. While the debate on the measurement of publishing performance through papers and citations is in a mature stage, the same often does not occur regarding these more recent measures. As mentioned by Wouters et al. (2015), in recent years several alternative metrics have been proposed but they have yet to prove themselves as credible tools. Improvements in this area are indeed one of the most promising avenues for further research.

A final remark should be made to emphasize that despite the merits of a (one-dimensional or multi-dimensional) quantitative analysis such as, for example, the one developed in the current study, it is probably insufficient on its own and should be complemented by qualitative assessment (peer review). The advantage of this mixed approach is clearly identified by Hammarfelt and Rushforth (2017). They highlight that it can operate as an intermediate equilibrium between pure quantitative and qualitative approaches, mitigating the weakness inherent to both of them. In a pioneering paper on this topic, published in 2007, Henk Moed said that “the future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review” (Moed 2007, p. 576). All these years later, this statement seems truer than ever.

Acknowledgements This work was supported by the Fundação para a Ciência e a Tecnologia under Grant UID/GES/00,315/2019. We are grateful to the two anonymous referees for their very useful comments. The usual disclaimer applies.

References

Abramo, G., & D’Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics, 101,* 1129–1144.

Abt, H. (2012). A publication index that is independent of age. *Scientometrics, 91,* 863–868.

Alonso, Sergio, Cabrerizo, Francisco, Herrera-Viedmac, Enrique, & Herrera, Francisco. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics, 3,* 273–289.

Alonso, Sergio, Cabrerizo, Francisco, Herrera-Viedmac, Enrique, & Herrera, Francisco. (2010). hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices. *Scientometrics, 82,* 391–400.

Amjad, Tehmina, & Daud, Ali. (2017). Indexing of authors according to their domain of expertise. *Malaysian Journal of Library & Information Science, 22,* 69–82.

Ball, Philip. (2005). Index aims for fair ranking of scientists. *Nature, 436,* 900.

Bergstrom, Carl, West, Jevin, & Wiseman, Marc. (2008). The Eigenfactor™ metrics. *Journal of Neuroscience, 28,* 11433–11434.

Bornmann, Lutz, Butz, Alexander, & Wohlrabe, Klaus. (2018). What are the top five journals in economics? A new meta-ranking. *Applied Economics, 50,* 659–675.

Bornmann, Lutz, & Haunschild, Robin. (2018). Plots for visualizing paper impact and journal impact of single researchers in a single graph. *Scientometrics, 115,* 385–394.

Bornmann, Lutz, & Marx, Werner. (2011). The h index as a research performance indicator. *European Science Editing, 37,* 77–80.

Bornmann, Lutz, Mutz, Rüdiger, & Daniel, Hans-Dieter. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology, 59,* 830–837.

Bornmann, Lutz, Thor, Andreas, Marx, Werner, & Schier, Hermann. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized google scholar data for the publications of a research institute. *Journal of the Association for Information Science and Technology, 67,* 2778–2789.

da Silva, J., & Dobránszki, J. (2018). Multiple versions of the h-index: Cautionary use for formal academic purposes. *Scientometrics, 115,* 1107–1113.

Ding, Ying. (2011). Applying weighted pagerank to author citation networks. *Journal of the American Society for Information Science and Technology, 62,* 236–245.

Dunnick, Nicholas. (2017). The h index in perspective. *Academic Radiology, 24,* 117–118.

Egghe, Leo. (2006). Theory and practise of the g-index. *Scientometrics, 69,* 131–152.

Egghe, Leo. (2010). The Hirsch index and related impact measures. *Annual Review of Information Science and Technology, 44,* 65–114.

Fenner, Trevor, Harris, Martyn, Levene, Mark, & Bar-Ilan, Judit. (2018). A novel bibliometric index with a simple geometric interpretation. *PLoS ONE, 13,* e0200098.

Frandsen, Tove, & Nicolaisen, Jeppe. (2010). What is in a name? Credit assignment practices in different disciplines. *Journal of Informetrics, 4,* 608–617.

Fu, Hui-Zhen, Wang, Ming-Huang, & Ho, Yuh-Shan. (2012). The most frequently cited adsorption research articles in the science citation index (expanded). *Journal of Colloid and Interface Science, 379,* 148–156.

Gao, Chao, Wang, Zhen, Li, Xianghua, Zhang, Zili, & Zeng, Wei. (2016). PR-index: using the h-index and pagerank for determining true impact. *PLoS ONE, 11,* e0161755.

Glänzel, Wolfgang, Debackere, Koenraad, Thijs, Bart, & Schubert, András. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics, 67,* 263–277.

Hamermesh, Daniel. (2018). Citations in economics: Measurement, uses, and impacts. *Journal of Economic Literature, 56,* 115–156.

Hammarfelt, Björn, & Rushforth, Alexander. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation, 26,* 169–180.

Hausken, Kjell. (2016). The ranking of researchers by publications and citations. *Journal of Economics Bibliography, 3,* 530–558.

Henriksen, Dorte. (2018). What factors are associated with increasing co-authorship in the social sciences? A Case study of Danish economics and political science. *Scientometrics, 114,* 1395–1421.

Hicks, Diana, Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometric: The Leiden Manifesto for research metrics. *Nature News, 520,* 429.

Hirsch, Jorge. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102,* 16569–16572.

Hirsch, Jorge. (2019). hα: An index to quantify an individual's scientific leadership. *Scientometrics, 118,* 673–686.

Iglesias, Juan, & Pecharromán, Carlos. (2007). Scaling the h-index for different scientific ISI fields. *Scientometrics, 73,* 303–320.

Järvelin, Kalervo, & Persson, Olle. (2008). The DCI index: Discounted cumulated impact-based research evaluation. *Journal of the American Society for Information Science and Technology, 59,* 1433–1440.

Kadel, Annke, & Walter, Andreas. (2015). Do scholars in economics and finance react to alphabetical discrimination? *Finance Research Letters, 14,* 64–68.

Kosmulski, Marek. (2018). Are you in top 1% (1‰)? *Scientometrics, 114,* 557–565.

Larivière, Vincent, Ni, Chaoqun, Gingras, Yves, Cronin, Blaise, & Sugimoto, Cassidy. (2013). Global gender disparities in science. *Nature, 504,* 211–213.

Leahey, Erin. (2016). From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annual Review of Sociology, 42,* 81–100.

Liu, Xuan, & Fang, Hui. (2012). Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution. *Journal of Informetrics, 6,* 557–565.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google scholar, web of science, and scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12,* 1160–1177.

Mayer, Sabrina, & Rathmann, Justus. (2018). How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. *Scientometrics, 117,* 1663–1693.

Mazurek, Jirí. (2018). A modification to Hirsch index allowing comparisons across different scientific fields. *Current Science, 114,* 2238–2239.

Moed, Henk. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy, 34,* 575–583.

Osório, António. (2018). On the impossibility of a perfect counting method to allocate the credits of multi-authored publications. *Scientometrics, 116,* 2161–2173.

Perry, Motty, & Reny, Philip. (2016). How to count citations if you must. *American Economic Review, 106,* 2722–2741.

Prins, A., Costas, R., van Leeuwen, T., & Wouters, P. (2016). Using google scholar in research evaluation of humanities and social science programs: A comparison with web of science data. *Research Evaluation, 25,* 264–270.

Schreiber, Michael. (2018). A skeptical view on the Hirsch index and its predictive power. *Physica Scripta, 93,* 102501.

Todeschini, Roberto, & Baccini, Alberto. (2016). *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*. Weinheim: Wiley.

Van Raan, A. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics, 67,* 491–502.

Waltman, Ludo, & Costas, Rodrigo. (2014). F 1000 recommendations as a potential new data source for research evaluation: A comparison with citations. *Journal of the Association for Information Science and Technology, 65,* 433–445.

Wildgaard, Lorna, Schneider, Jesper, & Larsen, Birger. (2014). A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics, 101,* 125–158.

Wouters, P., Thelwall, M., Kousha, K., Waltman, L, de Rijcke, S., Rushforth, A., & Franssen, T. (2015). *The metric tide: Literature review (supplementary Report I to the independent review of the role of metrics in research assessment and management)* Bristol: HEFCE.

Wu, Qiang. (2010). The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology, 61,* 609–614.

Wuchty, Stefan, Jones, Benjamin, & Uzzi, Brian. (2007). The increasing dominance of teams in production of knowledge. *Science, 316,* 1036–1039.

Zhang, Chun-Ting. (2009). The e-index, complementing the h-index for excess citations. *PLoS ONE, 4,* e5429.