



ELSEVIER

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: [www.elsevier.com/locate/joi](http://www.elsevier.com/locate/joi)

Regular article

# Self-Citations and scientific evaluation: Leadership, influence, and performance



Nadia Simoes, Nuno Crespo

Instituto Universitário De Lisboa (ISCTE - IUL), ISCTE Business School Economics Department, BRU - IUL (Business Research Unit), Av. Das Forças Armadas, 1649-026, Lisboa, Portugal

## ARTICLE INFO

### Article history:

Received 12 August 2019

Received in revised form 1 November 2019

Accepted 25 November 2019

### Keywords:

Scientific performance

Scientific leadership

Scientific influence

Self-citations

Credit

## ABSTRACT

The  $h_\alpha$  index was recently proposed by Hirsch (2019) to measure the degree of scientific leadership. However, as discussed in recent literature, this measure has important shortcomings. We introduce an alternative approach that does not suffer from these limitations. It uses self-citations as a source of information, makes the evaluation at the paper-level, and centers the analysis on the new and broader concept of scientific influence. In each specific paper, the level of scientific influence of an author ranges between 0 and 1 and corresponds to his/her share in the total number of self-citations in that paper. Moreover, we show how this concept can be used to produce a more accurate co-authorship weighting scheme that allows for adjusting the standard measures of scientific performance. This is particularly important in the case of areas in which the alphabetical order of the names is the rule commonly followed, since it provides a way to differentiate the authors according to their role in each paper. We illustrate our method with an empirical example, comparing male and female economists of world top universities. The evidence highlights the existence of substantial gender differences in terms of scientific leadership and scientific influence.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Bibliometric assessment is increasingly important to support a wide range of very important decisions in science, including hiring, promotions, allocation of research funds, awards, and rankings (Hamermesh, 2018; Osório, 2018). The importance of these decisions puts a high priority on the definition of rigorous and fair measures (Schreiber, 2018).

The  $h$ -index, introduced by Jorge Hirsch in 2005 (Hirsch, 2005), is unequivocally the most widely used author-level bibliometric measure of scientific performance and, as stated by Wildgaard, Schneider, and Larsen (2014), probably the only one that most researchers know. An author has an  $h$ -index  $h$  when  $h$  of his/her papers have at least  $h$  citations each and the remaining papers have no more than  $h$  citations each. Among the widely discussed shortcomings of this index (see, for instance, Egghe, 2010), the most important one concerns the fact that it does not take into consideration the number of authors of each paper and their different role/importance in the conception and production of the study (Hirsch, 2010, 2019). As mentioned by Vavryčuk (2018 p. 1), “so far, the single-author or multi-author publications contribute to the publication record of a researcher equally. This full counting scheme is apparently unfair and causes unjust disproportions, in particular, if ranked researchers have distinctly different collaboration profiles”. Without a clear identification of the

E-mail address: [nadia.simoes@iscte-iul.pt](mailto:nadia.simoes@iscte-iul.pt) (N. Simoes).

individual contributions, it is impossible to attribute the correct credit to each author (Tschardtke, Hochberg, Rand, Resh, & Krauss, 2007).

The importance of this issue is now greater than ever since the proportion of multi-author papers is rapidly increasing (Fang, 2018; Fernandes & Monteiro, 2017; Frandsen & Nicolaisen, 2010; Kosmulski, 2012). This is true even in the social sciences, where the average number of authors per paper is not as high as in natural and life sciences (see Marušić, Bošnjak, & Jerončić, 2011, for a discussion on the differences in co-authorship patterns in different sciences).

A comprehensive solution to this problem necessarily involves the reconciliation of the standard measures of performance (e.g.,  $h$ -index or  $g$ -index) with an adjustment that takes into account both the number of authors of each paper and the role they have in its elaboration. While the question related to the number of authors is easy to solve and several co-authorship weighting schemes are available (for a survey see Todeschini & Baccini, 2016), the second dimension is a much more complex challenge (Tschardtke et al., 2007), namely in scientific areas where the alphabetical order of the authors is the tradition. When this occurs, the use of the paper byline to identify the contributions of the several authors is not possible.

The concept of scientific leadership and the  $h_\alpha$  index recently proposed by Hirsch (2019) are interesting contributions. However, we point out that both the concept and the measure need to be revised. First, we need a measurement procedure that is able to solve the shortcomings of the  $h_\alpha$  index (Bornmann, Ganser, Tekles, & Leydesdorff, 2019; Leydesdorff, Bornmann, & Opthof, 2019). Second, we should move our attention from the concept of scientific leadership to the new and broader concept of scientific influence. With these two modifications, we are able to design an alternative co-authorship weighting scheme that can be used to allocate the credit for a paper among its several authors.

The present study makes three methodological contributions. First, we propose a measure of scientific leadership that avoids the limitations of the  $h_\alpha$  index. At the paper level, we evaluate scientific leadership through the analysis of self-citations, taken as proxy for specific human-capital. The scientific leader of a paper is the author with the highest number of self-citations in that paper. Second, we introduce the concept of scientific influence. In this context, the credit is divided by all authors as a function of their share in the total number of self-citations. Third, putting together the traditional  $h$ -index and the measure of scientific influence proposed in this study, we suggest a procedure to build an index of scientific performance that takes into account the number of authors and the role of each author in the paper. At the empirical level, we compare male and female economists working in world top universities, providing evidence concerning their levels of scientific leadership, scientific influence, and scientific performance.

The remainder of the paper is structured as follows. In Section 2 we discuss in detail the methodological issues under analysis and introduce our approach. In Section 3 we present the data and show preliminary descriptive evidence. In Section 4 we conduct our empirical application. Section 5 presents some final remarks.

## 2. Methodological approach: scientific leadership, scientific influence, and scientific performance

### 2.1. On the measurement of scientific leadership – the $h_\alpha$ index

Hirsch (2019) introduces the  $h_\alpha$  index to capture the degree of scientific leadership. Its rationale and measurement procedure are easy to understand. Considering the papers that compose the  $h$ -core of the author (i.e., those having at least  $h$  citations each), the  $h_\alpha$  index measures the number of papers in which the author was the scientific leader. Since leadership is a non-observable dimension, Hirsch (2019) suggests capturing it through the  $h$ -index, with the author having the highest  $h$ -index being classified as the leader of that paper. Obviously,  $h_\alpha \leq h$ . Additionally, Hirsch (2019) introduces the  $r_\alpha$  index, which aims to complement the first measure and can be obtained as  $r_\alpha = \frac{h_\alpha}{h}$ . It expresses the percentage of papers of the  $h$ -core of the author in which he/she was the scientific leader. While the concept of scientific leadership is interesting and can have important applications, the  $h_\alpha$  index suffers from evident shortcomings.

First, it requires the values of the  $h$ -index of all co-authors at the time the paper is produced. Since this information is not available in the standard bibliometric databases, the index is difficult to obtain (Leydesdorff et al., 2019). In order to overcome this problem, Hirsch (2019) considers as proxy the current values of the  $h$ -index of each author. This is problematic because the ranking of the authors in terms of their  $h$ -index can change – sometimes dramatically – between the moment the paper is published and the moment the leadership assessment is undertaken.

Associated with this first shortcoming, another problem of the  $h_\alpha$  index is its potential instability (Leydesdorff et al., 2019). The  $h$ -index of the various authors can change as a result of new papers produced and new citations of them.

Third, the  $h_\alpha$  index evaluates an overall type of scientific leadership (captured through the  $h$ -index). However, in our perspective, the concept of scientific leadership must be linked to the topic studied in each paper.

Fourth, the  $h_\alpha$  index identifies only one leader in each paper. Nevertheless, in many cases the influence of the authors is approximately the same, making it unreasonable to assume that there is only one leader. Hirsch (2019), himself, mentions the possibility of extending his concept of scientific leadership in order to take this feature into account, but does not advance a concrete way to do so.

Fifth, the  $h_\alpha$  index reinforces the Matthew effect (Bornmann et al., 2019; Leydesdorff et al., 2019), according to which “the more reputable scientist receives more credit than the less reputable scientist for a scientific contribution, although the contribution is of the same scientific quality. Thus, the credit is not attributed fairly on the basis of the performed

contribution, but (unfairly) on the basis of previous contributions” (Bornmann et al., 2019, p. 12). This occurs in the case of the  $h_a$  index, since the  $h$ -index of each author depends exclusively on his/her previous scientific achievements.

## 2.2. A new approach: using self-citations to measure scientific leadership and scientific influence

In light of the limitations discussed above, we argue that a new approach is needed. A key contribution of this paper is to introduce such an approach, which involves two critical modifications. First, instead of considering the  $h$ -index of the authors to assess an overall type of scientific leadership, we propose an alternative metric, based on the use of self-citations. Second, we put the focus on the concept of scientific influence. In the present section we discuss these two features in detail.

In our perspective, the concept of scientific leadership (and that of scientific influence we will introduce below) should reflect the scientific authority of the author in the context of the topic discussed in the paper. Our approach therefore establishes a direct association between leadership and knowledge.<sup>1</sup> This specific knowledge is the critical asset to perform several key roles in collaborative research, as defined, for example, by Allen, Brand, Scott, Altman, and Hlava (2014). They propose the Contributor Roles Taxonomy (CRediT), which has been adopted by several publishers (including Elsevier, MIT Press, Oxford University Press, Springer, or Wiley VCH). An alternative taxonomy is introduced by Rahman, Regenstein, Kassim, and Haque (2017). Some of the most important roles that can be associated with scientific leadership include: formulation of the initial research proposal (concept/idea/hypothesis); literature review to support the different stages of paper preparation; research (methodology) design; and supervision of the project.

According to Lazear (2012), leadership requires skills, and the broader the scope of intervention of the leaders the more general their skills should be. Academic papers are usually focused on relatively confined areas of study. As mentioned by Leahy & Reikowsky (2008, p. 426), “specialization is a defining characteristic of modern science”.

Having published scientific research on the topic of the paper gives an author mastery and an in-depth understanding of the relevant literature, and means that he/she has made a contribution to that body of knowledge (Brybaert & Smyth, 2011; Glänzel, Debackere, Thijs, & Schubert, 2006; Hou, Li, & Niu, 2011). These elements are critical for developing further research. The optimal solution to capture this level of expertise would be to consider the previous works of each author in the area of the paper. Nevertheless, this can only be done by the authors of the paper and a few other researchers working on the same issue. In the present study we propose using the self-citations included in the paper as a proxy for this information. As pointed out by Pichappan & Sarasvady (2002, p. 285), “science is a continuum of experiences consisting of the author and objects of experiences. Authors’ experience is the integral part of their work that gets reflected through self-references. (...) Self-citations indicate a continuous research process of one author or a group. (...) Authors are engaged in constant and continuing inquiry into the nature of science through a series of investigations that include a dialogue between the past and present work including his/her own”.

The debate on the inclusion of self-citations as regular citations has been intense in recent years. On the one hand, we should take into account that the use of self-citations avoids the repetition of materials already published and makes the link to previous works that support the contributions developed in the papers under analysis. When such considerations are the genuine motive of the author(s), self-citations are not only valid but even necessary (for a general discussion, see Glänzel et al., 2006). One step further is taken in a recent study conducted by Zhao, Strotmann, and Cappello (2018). They conclude that “self-citations are less likely to function as nonessential citations than do external citations, and have a particularly high likelihood to be Applied citations, the type of citations that have influenced the citing article most strongly” (Zhao et al., 2018, p. 952). As a consequence, they suggest that “contrary to a common belief, self-citations should not be discounted in citation analysis (...) and should perhaps be given more weight than external citations in weighted citation analysis that attempts to assign higher weights to essential citations” (Zhao et al., 2018, p. 952). On the other hand, there is convincing evidence pointing to specific cases of strategic use of self-citations to inflate the perceived performance (Ioannidis, Baas, Klavans, & Boyack, 2019; Van Noorden & Chawla, 2019). In short, while the use of self-citations is valid and acceptable, the abuse of self-citations is a real problem that must be addressed. As summarized by Van Raan (2006) and Seeber, Cattaneo, Meoli, and Malighetti (2019), for example, bibliometric measures should therefore be conceived carefully in order to avoid this strategic use of self-citations. We present our approach below and explain how it avoids potential problems associated with the “inappropriate” use of self-citations.

We start by building a new measure of scientific leadership. At the paper level we obtain, for each author, a variable ( $l_{ip}$ ) with value 1 if author  $i$  is the scientific leader of the paper  $p$  and 0 otherwise. Following our reasoning mentioned above, the scientific leader is the author with the highest number of self-citations in that paper. When two or more authors have the same number of self-citations in that paper, both/all are classified as leaders and the credit associated with that leadership

<sup>1</sup> This concept of scientific leadership shares some features with the “expert leadership theory” developed in the organizational context. This theory argues that organizations perform more effectively when led by individuals who have a deep understanding of the core business of their organization (Goodall, 2012).

(1) is divided equally among them. Finally, we calculate an author-level measure of scientific leadership –  $L$  – summing up  $l_{ip}$  across the papers of the  $h$ -core of the author ( $p = 1, 2, \dots, P_h$ ):

$$L = \sum_{p=1}^{P_h} l_{ip} \quad (1)$$

This addresses the shortcomings of the  $h_\alpha$  index and captures, as discussed above, a more accurate concept of scientific leadership. We even suggest that we can enrich the analysis by taking an additional step toward a more comprehensive measurement of scientific performance. To do that we move from the more restricted concept of scientific leadership to the more inclusive concept of scientific influence. The level of scientific influence of each author in each paper corresponds to his/her share in the total number of self-citations of the paper ( $s_{ip}$ ). In this way, all authors receive part of the credit (ranging from 0 to 1) due to their role (scientific influence) in the paper. Then, we create an author-level measure of scientific influence ( $I$ ) summing up the credits received by the author in each of the papers that comprise his/her  $h$ -core<sup>2</sup> :

$$I = \sum_{p=1}^{P_h} s_{ip} \quad (2)$$

Taking into account the considerations mentioned above, and especially the need to avoid the strategic use of self-citations to inflate scientific performance scores, we must assure that the approach developed to measure scientific influence is immune to this problem. In this context, it is important to stress four important characteristics of our method. First, the credit given to scientific influence/leadership in a given paper is always 1. Second, the number of self-citations is irrelevant on its own. What really matters is the distribution of the self-citations among the authors of the paper, i.e., the share of each author in the total number of self-citations of the paper. Thus, if the number of self-citations of all authors is multiplied by a constant, the measure of scientific influence remains unchanged. Third, if one of the authors has additional self-citations, his/her share will increase but the share of all the remaining authors will decrease. This is critical since it operates as a mechanism of control according to which the authors have an incentive to restrict the use of self-citations of their co-authors to those that are actually important to the paper. This feature is an important differentiation element of our approach. Fourth, in the case of single-author papers, the credit in terms of scientific influence is the same independently of the number of self-citations.

The new approach introduced in this section avoids the shortcomings of the  $h_\alpha$  index discussed above. Four main aspects can be highlighted in this regard. First, the levels of scientific leadership and scientific influence are evaluated taken as reference the moment the paper is produced, and is therefore insensitive to what happens thereafter. Second, we assume concepts of scientific leadership and scientific influence that are specific to the topic analyzed in the paper. Third, our method allows the existence of several leaders. Moreover, regarding scientific influence specifically, the importance of all authors is taken into account and they receive a credit ranging between 0 and 1. Finally, there is a clear distinction between the best scientist (evaluated in general terms) and the best author in the specific issue that is the topic of the paper. It is this last author that will be defined as leader and will receive the greatest credit in terms of scientific influence.

### 2.3. Building a co-authorship weighting scheme incorporating scientific influence

The strong increase in both the proportion of multi-author papers and the average number of authors per paper are remarkable trends of scientific production over the last decades. The importance of this phenomenon has led to a vast discussion on the reasons for that trend (Fang, 2018; Leahey, 2016; Stephan, 2012), on its effects (see, for example, Larivière, Gingras, Sugimoto, & Tsou, 2015), and above all, on the possible methods to give fair credit to each of the authors (co-authorship weighting schemes). How can we measure the actual contribution/merit of each author? As emphasized by Osório (2018, p. 2161), “the answer to this question is crucial because academics and researchers (as well as their institutions) are ranked, rewarded, financed, and promoted according to their publications. However, the more researchers are involved in a project, the more difficult [it] is for third parties (e.g., a panel or an evaluation committee) to quantify the contribution of each of them”<sup>3</sup>.

In a given paper the contribution of each author depends on the number of authors and on the role/importance of each of them in its production. While the number of authors is easily available, knowing the exact contribution of each author is much more complex. The ideal solution would be for the authors themselves to describe their individual contribution to the paper, as suggested by Lukovits and Vinkler (1995). Some journals already ask for qualitative information in this

<sup>2</sup> An alternative method to gauge the contribution of the co-authors to a given paper is proposed by Egghe, Guns, and Rousseau, (2013). In its simplest version, the credit of each author can be calculated as the ratio between his/her  $h$ -index and the sum of the  $h$ -indexes of all authors of the paper. The main difference between this method and our approach is the fact that while we try to capture the contribution of each author to the conception of the paper, in the analysis conducted by Egghe et al. (2013) the focus is on the contribution to the visibility of the paper, the critical argument being that this visibility critically depends on its authors.

<sup>3</sup> On this topic, see also Bornmann and Osório (2019).

regard. However, this procedure does not overcome the critical problem associated with the conversion of this qualitative information into a quantitative measure to use in the context of a more refined co-authorship weighting scheme.

Excluding this ideal solution, the usual procedure involves the use of the paper byline as a source of information (Berker, 2018). Nevertheless, this is complex, not always clear, and often impossible. Below we briefly discuss some of the most widely used weighting schemes (for surveys on this topic, see Osório, 2018; Todeschini & Baccini, 2016). The position of the author in the byline of the paper is represented by  $k$  ( $k = 1, 2, \dots, K$ ).

Two of the most commonly used weighting schemes ignore the question of the role, giving the same weight to all authors. First, the standard counting gives the full credit to all authors. This is the scheme used in some standard performance measures, such as the  $h$ -index (Hirsch, 2005) and the  $g$ -index (Egghe, 2006). Second, the uniform counting gives a credit equal to  $1/K$  to each author (Lindsey, 1980), representing the normalized version of the previous scheme (Osório, 2018). This last option is usually seen as preferable since it does not inflate the weight of some papers.

A second group of schemes assumes that the sequence of authors in the byline has informative value. The simplest version – the proportional counting – assigns weight  $1/k$  to the author in  $k$ -position.

A third approach ignores the number of authors and aims only to capture their role, inferring the importance of contributions from the positions in the paper byline (Liu & Fang, 2012). Following the tradition of some scientific areas, the first author, the last author, and the corresponding author are usually seen as more important contributors than authors in the remaining positions (Kosmulski, 2012). Therefore, some weighting schemes give higher weights to these authors. In some extreme cases, the full credit is given to one single author, ignoring the contribution of all the remaining ones. This is the case of the straight counting scheme (Cole & Cole, 1973), in which the first author receives all of the credit. In other cases, mixed solutions are proposed, such as those by Abramo, D'Angelo, and Rosati (2013) and Vavryčuk (2018). This kind of procedure has important implications (see for instance Frandsen and Nicolaisen (2010) for a discussion on the advantages of first authorship and Efthyvoulou (2008) for a more general analysis) and led to the emergence, in some fields, of the phenomenon of equally important authorships, namely the existence of several first-authors and several corresponding authors, each with equal contributions to the paper (Fang, 2018; Hu, 2009; Hu, Rousseau, & Chen, 2010). A fundamental problem of this third approach is the absence of a consensual rule concerning the positions in the author's list, which could lead to important mistakes when a common weighting scheme is applied. In short, "interpretation of author sequence can be like a lottery" (Tscharrntke et al., 2007).

The deficiencies of the methods mentioned above are tremendously amplified when, as is the case in some scientific areas, the alphabetical order of the names is a tradition followed by the majority of the authors. This is the case of economics, as documented by Frandsen and Nicolaisen (2010), Kadel and Walter (2015) and Kuld and O'Hagan (2018) and confirmed in the present study. In these circumstances it is impossible to use the paper byline to infer differentiated contributions (Berker, 2018). We should keep in mind, however, that in most cases this behavior is much more the consequence of a commonly adopted tradition and of the intention to avoid disharmony in the group by violating this tradition than the result of an effective situation of equal contributions of all authors (Engers, Gans, Grant, & King, 1999; Frandsen & Nicolaisen, 2010; Liu & Fang, 2012). Fang (2018, p. 1592) makes this explicit in saying that "collaboration requires solidarity and harmony rather than conflict and competition. Otherwise, the collaboration breaks down and fails to advance. Therefore, it is wise for authors to avoid exact comparisons of their importance if doing so will cause arguments among the collaborators". The problematic implication of this situation is that when alphabetical order is adopted, the traditional co-authorship weighting schemes can, at most, take into account the number of authors (Berker, 2018). As summarized by Kosmulski (2012, p. 640), "there is no evidence that the authors indeed contribute to the paper equally, but an assumption of equal contribution seems natural when explicit information about the contributions of particular authors is lacking".<sup>4</sup>

The inability to identify distinct contributions in the case of alphabetical order is a critical drawback in the literature. A main argument of the present paper is that the concept of scientific influence can provide an interesting solution, aiming to obtain an author-level measure of performance (evaluating the traditional performance dimensions, i.e., number of papers and number of citations) that is sensitive to the number of authors and to their role in the paper.

Starting at the paper level, we define the share of the overall credit of the paper that should be given to author  $i$  as the simple average between: (i) the share that corresponds to the equal distribution among the co-authors (i.e., applying the uniform counting scheme); (ii) the share of the author in the total number of self-citations,  $s_{ip}$ .<sup>5</sup> We define this average as  $z_{ip}$ :

$$z_{ip} = \frac{1}{K} + \frac{s_{ip}}{2} \quad (3)$$

Putting this new weighting scheme together with the traditional  $h$ -index (or some alternative measure of scientific performance), we can obtain a performance index – which we call  $h$  adjusted index ( $h^{adj}$ ) – that captures the number of papers and the number of citations while also taking into consideration the number of authors of each paper (belonging to the  $h$ -core) and their role in terms of scientific influence.

<sup>4</sup> This is not a neutral solution. In fact, there is convincing evidence that authors late in the alphabet react to this "alphabetical discrimination" reducing the number of multi-author papers in which they participate (Kadel & Walter, 2015; Laband & Tollison, 2000).

<sup>5</sup> While we adopt this procedure for the sake of simplicity, alternative methods to weight these dimensions can of course be considered.

In the literature there are some indicators that adjust the  $h$ -index according to a co-authorship weighting scheme. These measures do so by cutting the total number of papers (e.g., the paper fractional  $h$ -index, proposed by Egghe (2008), cutting the total number of citations (e.g., the weighted  $h$ -index suggested by Abbas (2011), or adjusting the  $h$ -index after its calculation. Three examples of this last option are the pure  $h$ -index (Wan, Hua, & Rousseau, 2007), the complementary  $h$ -index (Dorta-González & Dorta-González, 2011), and the  $h^*$  index (Crespo & Simoes, 2019). Following this last formulation, we obtain:

$$h^{adj} = \sum_{p=1}^{P_h} z_{ip} \quad (4)$$

It is easy to verify that  $h^{adj} \leq h$ . In fact, we obtain  $h^{adj} = h$  only when all the papers of the  $h$ -core of the author are single-author.

### 3. Data and preliminary evidence

Aiming to illustrate the methods introduced in the previous section, we consider a database composed of the authors with primary appointments at the departments of economics of the world top 10 universities in the area of “Economics & Econometrics” according to the QS World University Ranking (QS, 2018): Harvard University; Massachusetts Institute of Technology; Stanford University; University of California, Berkeley; Princeton University; University of Chicago; London School of Economics and Political Science; University of Oxford; Yale University; and Columbia University. This criterion leads to the identification of 472 authors, corresponding to the full list of tenure-track or tenured faculty members, as defined by departmental websites. All data were gathered from the Web of Science (WoS) database (core collection). The papers selected include those that fulfil two conditions: (i) are classified as articles or reviews; (ii) are written in English. To assure the validity of the data we made a detailed comparison for each of the authors included in our database between the list of papers given by WoS and the information on individual publications retrieved from departmental and personal websites. When necessary, complementary sources were also used (e.g., Econlit). Table 1 shows some key characterization elements of our sample.

Given the composition our sample, the number of published papers grew significantly over the period. Regarding the distribution of the papers by quartiles (measuring the quality of the journals in which the papers are published), we see a clear predominance of papers published in Q1 journals. These represent 65.91 % of the total number of papers published until 2018 by the 472 authors under analysis. In turn, papers in Q2 journals correspond to almost 20 % of the total. Regarding the journals in which more papers are published, there is a predominance of the “big 5” journals in economics: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*.

The evidence presented in Table 2 clearly shows a strong trend in terms of scientific production: the skyrocketing increase in the average number of authors per paper. The proportion of single-author papers falls from a value above 65 % at the beginning of the period under analysis to less than 15 % at the end. It is interesting to note that while the share of papers with two authors increased rapidly until the 1990s (from a value of 30.15 % in the beginning to 48.55 % in the 1990s), after that moment there is a reversal of that trend, and the value for the current decade (2010s) is only 37.40 %. This recent trend is explained by a robust increase in the number of papers having three or more authors. In the current decade, the percentage of papers with three or four authors is more than 45 % of the total.

Analyzing the evidence for each of the authors in our sample, we conclude that eight authors have only single-author papers (with a maximum number of papers of 13). In turn, the maximum average value is 16.1 authors per paper (David Laibson, who is also one of the 255 authors of the paper with the most authors in our sample).

Considering the papers with two or more authors, we verify that in the whole period, 91.38 % of the papers list the authors in alphabetical order in the byline. This confirms the situation in the field of economics, as mentioned in the previous section. Moreover, this number remains roughly stable over the entire period. Two hundred and five out of 447 authors (since 17 authors have no papers and eight have only single-author papers) have all their papers with the authors listed in alphabetical order. In another 138 cases, this percentage is greater than or equal to 90 %.

### 4. Leadership, influence, and performance: the case of economists in world top universities

The evidence presented in the previous section makes it clear that multi-authorship is a remarkable trend of the last decades. This gives high priority to the analysis of adequate ways to share the credit associated with each paper. The most complex issue in this is identifying the roles of the authors when listing them in alphabetical order is the tradition. In the empirical example discussed in this section we illustrate the use of the concept of scientific influence for this purpose. Beyond this objective, we aim to test the existence of gender differences in this dimension. This is an important topic because the existence of a gender gap in terms of scientific performance is well documented concerning not only synthetic measures of performance such as the  $h$ -index (Cameron, White, & Gray, 2016; Geraci, Balsis, & Busch, 2015), but also number of publications (Larivière, Ni, Gingras, Cronin, & Sugimoto, 2013; Mayer & Rathmann, 2018), number of citations (Beaudry & Larivière, 2016; Larivière et al., 2013), quality of publications (Brooks, Fenton, & Walker, 2014; Mayer & Rathmann, 2018), hiring (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012), awards (Boyle, Smith, Cooper, Williams, & O'Connor,

**Table 1**  
Scientific publications of the 472 economists in the sample.

	1957–1970	1971–1980	1981–1990	1991–2000	2001–2010	2011–2018	Total
No. of papers	136	683	1,936	2,993	4,861	4,634	15,243
No. of citations	14,606	84,419	245,934	426,167	421,731	95,946	1,288,803
Distribution by quartiles of journals							
Q1	77.94 %	72.47 %	64.10 %	62.15 %	66.55 %	67.11 %	65.91 %
Q2	13.97 %	16.84 %	21.44 %	21.42 %	18.54 %	20.39 %	19.92 %
Q3	6.62 %	4.69 %	7.23 %	10.32 %	9.75 %	8.14 %	8.80 %
Q4	1.47 %	6.00 %	7.23 %	6.11 %	5.16 %	4.36 %	5.37 %
Journals with more papers							
1st	Review of Economic Studies (13.97 %)	American Economic Review (10.83 %)	American Economic Review (11.78 %)	American Economic Review (12.23 %)	American Economic Review (12.90 %)	American Economic Review (15.15 %)	American Economic Review (13.20 %)
2nd	American Economic Review (11.03 %)	Econometrica (8.78 %)	Econometrica (6.92 %)	Quarterly Journal of Economics (6.31 %)	Econometrica (5.45 %)	Econometrica (5.05 %)	Econometrica (5.60 %)
3rd	Journal of Political Economy and Quarterly Journal of Economics (10.29 %)	Journal of Political Economy (8.78 %)	Journal of Political Economy (5.32 %)	Econometrica (4.98 %)	Quarterly Journal of Economics (5.33 %)	Quarterly Journal of Economics (4.45 %)	Quarterly Journal of Economics (5.25 %)
4th	–	Review of Economic Studies (7.61 %)	Quarterly Journal of Economics (4.70 %)	European Economic Review (4.58 %)	Journal of the European Economic Association (3.39 %)	Journal of Political Economy (2.96 %)	Journal of Political Economy (3.94 %)
5th	Economic Journal (9.56 %)	Quarterly Journal of Economic (6.00 %)	Review of Economic Studies (4.29 %)	Journal of Political Economy (4.41 %)	Journal of Political Economy (3.19 %)	Journal of Econometrics (2.78 %)	Review of Economic Studies (3.35 %)

**Table 2**  
Distribution of papers according to the number of authors and use of alphabetical order (%).

No. of authors	1957–1970	1971–1980	1981–1990	1991–2000	2001–2010	2011–2018	Total
1	65.43	58.41	42.93	32.54	21.36	14.74	26.34
2	30.15	37.19	46.28	48.55	46.82	37.40	43.65
3	3.68	4.10	8.83	15.27	22.71	29.18	20.45
4	0.74	0.15	1.91	2.87	5.78	11.18	6.06
5	–	–	–	0.37	2.20	4.40	2.11
6–10	–	0.15	0.05	0.27	0.80	2.31	1.02
11–100	–	–	–	0.13	0.33	0.73	0.35
>100	–	–	–	–	–	0.06	0.02
Alphabetical order (multi-author papers)	86.96	87.70	92.80	92.90	91.30	90.60	91.38

**Table 3**  
Scientific leadership of a given author ( $h$ -index = 12).

Paper	No. of authors	Position of the author	Number of self-citations							Leader(s)	$l_{ip}(0-1)$
			Author 1	Author 2	Author 3	Author 4	Author 5	Max	Total		
Paper 1	5	5	3	3	7	3	4	7	20	Author 3	0
Paper 2	2	2	4	3	–	–	–	4	7	Author 1	0
Paper 3	2	2	4	3	–	–	–	4	7	Author 1	0
Paper 4	2	2	1	1	–	–	–	1	2	Authors 1 and 2	0.5
Paper 5	5	5	0	0	0	0	2	2	2	Author 5	1
Paper 6	2	2	3	3	–	–	–	3	6	Authors 1 and 2	0.5
Paper 7	2	2	2	5	–	–	–	5	7	Author 2	1
Paper 8	2	2	2	2	–	–	–	2	4	Authors 1 and 2	0.5
Paper 9	2	2	3	3	–	–	–	3	6	Authors 1 and 2	0.5
Paper 10	4	4	1	1	1	3	–	3	6	Author 4	1
Paper 11	2	2	0	0	–	–	–	0	0	Authors 1 and 2	0.5
Paper 12	3	3	0	2	2	–	–	2	4	Authors 2 and 3	0.5
$\Sigma$											6

Note: The sequence of authors identified in the columns related to the number of self-citations follows the order of the byline.

2015; Lincoln, Pincus, Koster, & Leboy, 2012), and promotions (De Paola, Ponzo, & Scoppa, 2018; Shen, 2013). Additionally, empirical studies on this topic have shown that women have a lower probability of occupying the first or last position in the paper byline (Larivière et al., 2013) and that they exhibit different co-authorship patterns (Araújo, Araújo, Moreira, Herrmann, & Andrade, 2017; Bozeman & Gaughan, 2011). In this study we complement this body of literature discussing gender differences in terms of scientific influence and leadership, namely concerning authors working in world top universities. To that end, we select the top 25 male and the top 25 female authors in terms of their  $h$ -index. In the case of males, this group includes authors with an  $h$ -index ranging between 44 and 91 (Andrei Shleifer) while for females the values range between 12 and 43 (Janet Curie).

We start the analysis presenting a simple example to illustrate the application of our method by considering the case of one author with  $h = 12$ . Table 3 shows the results concerning the analysis of scientific leadership.

Paper 1 was produced by 5 authors and has a total of 20 self-citations (the highest value among the 12 papers considered). The author in the third position in the byline has the maximum number of self-citations (7) among the authors of the paper, and is therefore identified as the scientific leader of this paper, while the author we are analyzing (author in the fifth position in the byline) has 4 self-citations. Authors in the positions 1, 2, 4, and 5 of the byline receive 0 in this dimension since they were not the leaders of the paper. The last two columns of Table 3 show that in six of the 12 papers considered there is more than one leader. While in the case of paper 11 this occurs because none of the authors have self-citations in that paper, in the other five there are self-citations in the paper but two authors are tied at the maximum. Taking together the evidence in Table 3, we see that this author was the single leader in three papers (with a number of self-citations ranging between 2 and 5), shared the leadership in six papers (with a number of self-citations between 0 and 3), and was not the leader in the remaining three papers (with a number of self-citations between 3 and 4). These results lead to an overall leadership score of  $L = 6$ .

As mentioned in Section 2.2, we argue that the empirical analysis gains new insights when we move from the concept of scientific leadership to the more inclusive concept of scientific influence. Table 4 illustrates, for the same author, the measurement procedure applied to capture the concept of scientific influence and its impact on scientific performance.

Considering Paper 1 again, the 4 self-citations from the author under evaluation correspond to 20 % of the total number of self-citations in that paper. Author 3 has 7 self-citations, corresponding to 35 % of the total, and each of the remaining authors have a level of scientific influence of 15 %. If only the number of authors is taken into account and the uniform counting scheme is applied, the credit received by each author is 0.2, since this paper is co-authored by 5 authors. Nevertheless, this procedure neglects differences among the authors concerning their role in the paper. Incorporating both dimensions



**Table 4**  
Scientific influence of a given author ( $h$ -index = 12).

Paper	$\frac{1}{k}$	Share of the author in the total number of self-citations					Overall credit of the paper					$Z_{ip}$
		Author 1	Author 2	Author 3	Author 4	Author 5	Author 1	Author 2	Author 3	Author 4	Author 5	
Paper 1	0.200	0.150	0.150	0.350	0.150	0.200	0.175	0.175	0.275	0.175	0.200	0.200
Paper 2	0.500	0.571	0.429				0.536	0.464				0.464
Paper 3	0.500	0.571	0.429				0.536	0.464				0.464
Paper 4	0.500	0.500	0.500				0.500	0.500				0.500
Paper 5	0.200	0.000	0.000	0.000	0.000	1.000	0.100	0.100	0.100	0.100	0.600	0.600
Paper 6	0.500	0.500	0.500				0.500	0.500				0.500
Paper 7	0.500	0.286	0.714				0.393	0.607				0.607
Paper 8	0.500	0.500	0.500				0.500	0.500				0.500
Paper 9	0.500	0.500	0.500				0.500	0.500				0.500
Paper 10	0.250	0.167	0.167	0.167	0.500		0.208	0.208	0.208	0.375		0.375
Paper 11	0.500	0.500	0.500				0.500	0.500				0.500
Paper 12	0.333	0.000	0.500	0.500			0.167	0.417	0.417			0.417
$\Sigma$	<b>4.983</b>											<b>5.627</b>

Note: The sequence of authors identified in the columns related to the number of self-citations follows the order of the byline.

(number of authors and role, evaluated through scientific influence) in the same weighting scheme, authors 1, 2, and 4 receive a credit of 0.175 each (the simple average between 0.2 and 0.15), while author 3 receives 0.275 (the simple average between 0.2 and 0.35), and the author under scrutiny receives 0.2.

As stressed in the methodological discussion above, the evaluation of scientific influence is based on the distribution of self-citations among the authors of each paper and not on the absolute number of self-citations. This idea is clear in the evidence presented in Table 4. For example, the author under analysis receives a credit of 0.2 (i.e., 20 % of the total credit) for his/her scientific influence in paper 1 due to his/her 4 self-citations. However, he/she receives a credit of 0.5 in papers 4 and 11, with only 1 and 0 self-citations, respectively.

This author has  $h = 12$ . However, it is well-known that this measure does not fractionalize the credit of each paper, which in a context of increasing number of authors per paper is hard to accept as valid. When we take into account only the number of authors, the adjusted  $h$ -index falls to 4.983. However, this author has an above-average profile of scientific influence, allowing the final performance score ( $h$ -index adjusted by number of authors and role –  $h^{adj}$ ) to increase *vis-à-vis* the score that incorporates only the number of authors.  $h^{adj}$  is therefore 5.627, i.e., we can say that the author is responsible for 46.9 % of the total credit due to the 12 papers that comprise the  $h$ -core.

Tables 5 and 6 present our main evidence for the two groups (males and females), including three dimensions: leadership, influence, and performance.

Let us consider first the results for scientific leadership. As expected, given the composition of the sample (economists working in top universities), the level of scientific leadership is very high. In the case of the men, there are 5 authors (Amartya Sen, Martin Feldstein, Peter Phillips, Alvin Roth, and James Heckman) who were leaders in more than 90 % of the papers they participated in ( $\frac{l}{h} > 0.9$ ) and only 1 with a value below 50 % (David Cutler, with  $\frac{l}{h} = 0.3846$ ). Although lower than for men, the leadership profile is still very high in the case of women. Twenty-one of the 25 female authors in Table 5 have  $\frac{l}{h} \geq 0.5$ . However, only Graciela Chichilnisky reaches a value greater than 0.9. Comparing the average values for men and women, we detect the advantage of the former (0.7102 vs. 0.6073).

How does this evidence change when we move the analysis from scientific leadership to scientific influence? As expected, the distribution of credits among the authors of a given paper is more balanced in terms of scientific influence, since all the authors with at least 1 self-citation receive part of the credit associated with the production of the paper.<sup>6</sup> Comparing  $l$  and  $L$ , we verify that  $l$  is lower in the case of 22 male authors and 14 female authors. Taking into account the above-mentioned profile of strong leadership of these authors, this is not surprising. In fact, for all authors classified as leaders, the analysis conducted in terms of scientific influence implies receiving a lower or, at most, equal credit (which occurs only when all of the non-leaders have no self-citations). The average value of  $\frac{l}{h}$  is 0.6421 for men and 0.5776 for women. Only 1 male author and 6 female authors have an overall level of scientific influence lower than 50 %, which is remarkable, due not only to the average number of authors per paper (2.31 for men and 2.27 for women), but also given the high scientific level of their co-authors.

A critical methodological contribution of our method is the inclusion of the credit associated with the level of scientific influence in a co-authorship weighting scheme that, beyond the number of authors, also accounts for the role of the authors in each paper. The last columns of Tables 5 and 6 allow a direct comparison between the performance scores of the authors calculated through: (i) the basic version of the  $h$ -index (i.e., disregarding the number of authors of each paper and their role in its elaboration); (ii) the  $h$ -index adjusted by the number of authors (Crespo & Simoes, 2019), which is traditionally seen as the best we can do to obtain a fair division of the credit inherent to a paper when the byline is not informative about the

<sup>6</sup> When the total number of self-citations of the paper is zero, all receive the same (positive) credit.

**Table 5**  
Scientific Leadership, Scientific Influence, and Scientific Performance (top 25 – males).

Authors	Leadership		Influence		Performance		
	$L$	$\frac{L}{h}$	$I$	$\frac{I}{h}$	$h$	$h$ - number of authors	$h$ - number of authors and influence
Andrei Shleifer	67.67	74.36 %	48.57	53.37 %	91 (1)	35.85 (6)	42.21 (5)
James Heckman	75.33	91.87 %	62.51	76.23 %	82 (2)	44.50 (3)	53.51 (2)
Joseph Stiglitz	59.50	79.33 %	55.49	73.99 %	75 (3)	51.65 (1)	53.57 (1)
Daron Acemoglu	43.50	71.31 %	36.24	59.40 %	61 (4)	33.70 (8)	34.97 (8)
Edward Glaeser	36.00	64.29 %	30.83	55.05 %	56 (5)	25.50 (16)	28.16 (15)
Alberto Alesina	38.00	69.09 %	31.36	57.01 %	55 (6)	25.28 (17)	28.32 (14)
Peter Phillips	52.75	97.69 %	44.36	82.14 %	54 (7)	33.75 (7)	39.05 (7)
Robert Barro	44.50	82.41 %	42.15	78.05 %	54 (7)	39.83 (5)	40.99 (6)
John List	36.33	68.55 %	33.49	63.19 %	53 (9)	28.12 (12)	30.81 (11)
David Cutler	20.00	38.46 %	19.42	37.35 %	52 (10)	18.17 (24)	18.80 (25)
John Campbell	40.67	78.21 %	34.60	66.54 %	52 (10)	29.50 (10)	32.05 (10)
Martin Feldstein	50.50	99.02 %	48.95	95.98 %	51 (12)	43.67 (4)	46.31 (4)
Amartya Sen	50.50	99.02 %	49.56	97.19 %	51 (12)	48.33 (2)	48.95 (3)
Drew Fudenberg	28.50	57.00 %	23.64	47.27 %	50 (14)	22.45 (21)	23.04 (23)
Philippe Aghion	25.92	52.89 %	21.10	43.06 %	49 (15)	17.37 (25)	19.23 (24)
Mark Rosenzweig	35.83	74.65 %	30.77	64.11 %	48 (16)	24.67 (18)	27.72 (16)
Gregory Mankiw	29.83	63.48 %	29.38	62.52 %	47 (17)	28.25 (11)	28.82 (13)
Alvin Roth	44.00	93.62 %	38.09	81.05 %	47 (17)	28.02 (13)	33.05 (9)
Jeffrey Sachs	27.33	58.16 %	26.27	55.90 %	47 (17)	20.81 (23)	23.54 (22)
David Card	25.50	54.26 %	27.09	57.64 %	47 (17)	27.75 (14)	27.42 (17)
Tim Besley	25.33	55.07 %	26.30	57.17 %	46 (21)	24.42 (19)	25.36 (19)
Lawrence Summers	30.58	66.49 %	26.26	57.09 %	46 (21)	22.37 (22)	24.32 (20)
James Poterba	31.00	68.89 %	30.89	68.65 %	45 (23)	29.92 (9)	30.40 (12)
Jeremy Stein	27.83	63.26 %	26.23	59.61 %	44 (24)	25.53 (15)	25.88 (18)
Alan Krueger	23.83	54.17 %	24.49	55.65 %	44 (24)	24.07 (20)	24.28 (21)

Note: Numbers in parentheses correspond to the position of the author in the ranking using each performance measure.

**Table 6**  
Scientific Leadership, Scientific Influence, and Scientific Performance (top 25 – females).

Authors	Leadership		Influence		Performance		
	$L$	$\frac{L}{h}$	$I$	$\frac{I}{h}$	$h$	$h$ - number of authors	$h$ - number of authors and influence
Janet Currie	28.00	65.12 %	24.60	57.22 %	43 (1)	20.53 (1)	22.57 (1)
Esther Duflo	20.42	55.18 %	17.10	46.21 %	37 (2)	15.77 (4)	16.43 (4)
Claudia Goldin	21.50	69.35 %	20.97	67.65 %	31 (3)	20.50 (2)	20.74 (2)
Serena Ng	12.00	42.86 %	13.47	48.11 %	28 (4)	14.50 (6)	13.99 (6)
Amy Finkelstein	18.83	67.26 %	14.55	51.95 %	28 (4)	12.79 (8)	13.67 (7)
Pinelopi Goldberg	10.33	46.97 %	11.21	50.93 %	22 (6)	10.75 (11)	10.98 (11)
Cecilia Rouse	13.00	59.09 %	12.46	56.65 %	22 (6)	12.08 (9)	12.27 (9)
Xiaohong Chen	14.50	65.91 %	12.06	54.81 %	22 (6)	9.53 (14)	10.80 (13)
Christina Romer	16.00	76.19 %	15.67	74.60 %	21 (9)	15.50 (5)	15.58 (5)
Graciela Chichilnisky	19.33	92.06 %	18.48	88.00 %	21 (9)	17.17 (3)	17.82 (3)
Stephanie Schmitt-Grohé	9.83	49.17 %	10.51	52.57 %	20 (11)	10.50 (12)	10.51 (14)
Ulrike Malmendier	11.33	59.65 %	9.33	49.09 %	19 (12)	8.92 (18)	9.12 (17)
Hilary Hoynes	11.00	57.89 %	10.20	53.69 %	19 (12)	9.25 (16)	9.73 (15)
Beata Javorcik	13.00	68.42 %	11.97	63.00 %	19 (12)	9.92 (13)	10.94 (12)
Nancy Stokey	11.00	61.11 %	11.25	62.50 %	18 (15)	13.50 (7)	12.38 (8)
Muriel Niederle	6.50	38.24 %	7.14	42.00 %	17 (16)	7.17 (21)	7.15 (21)
Oriana Bandiera	8.17	51.04 %	7.50	46.89 %	16 (17)	6.75 (22)	7.13 (22)
Nancy Rose	7.50	50.00 %	8.14	54.28 %	15 (18)	9.00 (17)	8.57 (18)
Caroline Hoxby	11.00	78.57 %	11.40	81.43 %	14 (19)	11.03 (10)	11.22 (10)
Silvana Tenreyro	7.00	50.00 %	7.50	53.55 %	14 (19)	7.20 (20)	7.35 (20)
Pascaline Dupas	7.50	57.69 %	7.54	58.00 %	13 (21)	7.23 (19)	7.39 (19)
Alessandra Casella	9.50	73.08 %	9.88	76.03 %	13 (21)	9.33 (15)	9.61 (16)
Leeat Yariv	6.00	50.00 %	6.27	52.26 %	12 (23)	4.98 (25)	5.63 (24)
Emi Nakamura	8.00	66.67 %	6.79	56.60 %	12 (23)	6.42 (23)	6.60 (23)
Gita Gopinath	8.00	66.67 %	5.51	45.96 %	12 (23)	5.58 (24)	5.55 (25)

Note: Numbers in parentheses correspond to the position of the author in the ranking using each performance measure.

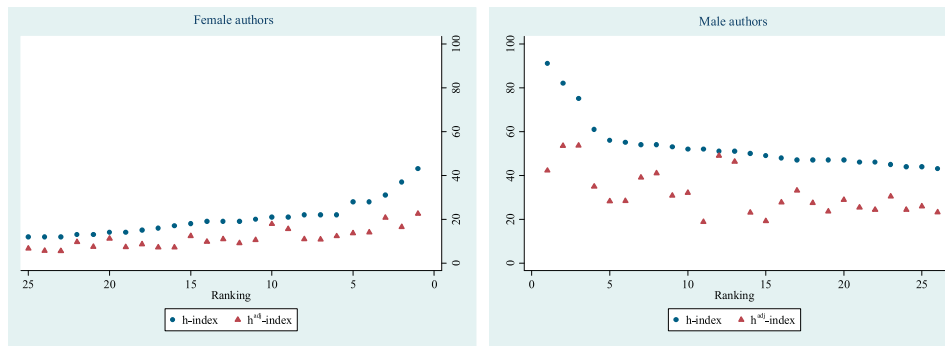


Fig. 1. Distribution of the h-index and  $h^{adj}$  index.

different role of the authors; and (iii) the measure proposed in the present study ( $h^{adj}$ ) to account for both the number of authors and their role, captured through the respective level of scientific influence. Let us emphasize some key results.

First, the consideration of the number of authors introduces very substantial decreases in the performance scores. At this level, the cases of Amartya Sen and Martin Feldstein are the most remarkable ones, as these are the authors with the lowest decreases in relative terms.

Second, the index considering the number of authors and their role is, for the vast majority of the authors, greater than when the role is ignored. The exceptions are David Card and 5 female authors (Serena Ng, Nancy Stokey, Muriel Niederle, Nancy Rose, and Gita Gopinath). This means that, in general, the authors in our sample have a level of scientific influence above the average (including themselves and their co-authors).

Third, for both genders, there are important modifications in terms of the ranking of the authors according to the different metrics, highlighting the importance of a more detailed and rigorous method to allocate the credit derived from the production of a given scientific output. Fig. 1 shows the comparison between the metric introduced in this paper and the traditional  $h$ -index.

The existence of impactful changes in the case of male authors is evident, expressed in a high number of ranking modifications. On the contrary, in the case of female authors there is more stability in terms of ranking.

## 5. Final remarks

There are deep differences between areas of science concerning the use of alphabetical order in the paper byline. While in the majority of cases, especially in natural and life sciences, the list of authors does not follow the alphabetical order, and instead expresses the differing contributions of the authors to the paper, in other fields, the “alphabetical rule” is widely adopted. The most studied example of this last case is economics, with many studies finding robust evidence pointing to around 90 % of the papers listing the authors according to this “alphabetical rule”. A critical implication of this practice is that it precludes the building of a weighting scheme that simultaneously takes into account the number of authors of the paper and their roles.

The inability to identify different contributions according to the role of the authors in the paper is an important drawback. In this study we propose a solution to this problem, involving the concept of scientific influence. Given the important shortcomings of the measure of scientific leadership suggested by Hirsch (2019), we propose a different approach based on the analysis of self-citations, which enlarges the evaluation to the broader concept of scientific influence.

It is well known that the use of self-citations is not immune to criticism, namely because they can be strategically used by the authors to inflate their perceived performance. A key advantage of our approach is the fact that this type of criticism does not apply to it, since what matters in our approach is the distribution of self-citations among the authors of the paper and not the number of self-citations itself. Moreover, the method proposed in this study to account for scientific influence incorporates a mechanism of control according to which the “policing of self-citations” (Van Noorden & Chawla, 2019) can be made by the research team of the paper.

We conduct an empirical application of our method. From that exercise four main conclusions can be drawn. First, confirming previous evidence, the use of alphabetical order is very common in the field of economics (91.38 % in our sample, considering the period 1957–2018). Moreover, this figure remains quite stable throughout that period. Second, there is a tremendous growth in the average number of authors per paper, due to (among other factors) a remarkable decrease of the number of single-author papers. Third, comparing the performance by gender, we see a clear advantage of males, illustrated for example in the fact that the first woman appears in the 37th position of the ranking in terms of  $h$ -index. Finally, the levels of scientific leadership and scientific influence are also stronger in the case of males. While our main contribution is methodological, these empirical results are important additions to the body of literature that evaluates gender differences in scientific activity.

An important final remark concerns the use of bibliometric indicators to conduct author-level evaluations. There are many examples of the application of the most diffused bibliometric measures (e.g., the *h*-index) to rank authors in the context of different scientific evaluation processes, including, for example, hiring, promotions, or awards (see, for instance, Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015; De Paola et al., 2018; Tietze & Hofmann, 2019). However, the existence of important shortcomings in such a purely quantitative approach is also clear. The importance of complementing this quantitative approach with qualitative assessment (peer review) is today widely recognized since it can mitigate the weaknesses of both approaches (Hammarfelt & Rushforth, 2017; Moed, 2007). The basic idea in this regard is clearly expressed by the first principle of the Leiden Manifesto: “Quantitative evaluation should support qualitative, expert assessment: Quantitative metrics can challenge bias tendencies in peer review and facilitate deliberation. This should strengthen peer review, because making judgements about colleagues is difficult without a range of relevant information. However, assessors must not be tempted to cede decision-making to the numbers. Indicators must not substitute for informed judgement. Everyone retains responsibility for their assessments” (Hicks et al., 2015, p. 430). The importance of peer review does not mean, however, that less of an effort should be made to develop rigorous bibliometric metrics. On the contrary, in the words of Hammarfelt and Rushforth (2017), they represent critical “judgment devices” supporting fairer evaluation processes in science.

## Author contributions

Nadia Simoes: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper.

Nuno Crespo: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper.

## Acknowledgement

This work was supported by the Fundação para a Ciência e a Tecnologia under Grant UID/GES/00315/2019. We are grateful to the two anonymous referees for their very useful comments. The usual disclaimer applies.

## References

- Abbas, A. (2011). *Weighted indices for evaluating the quality of research with multiple authorship*. *Scientometrics*, 88, 107–131.
- Abramo, G., D'Angelo, C., & Rosati, F. (2013). *The importance of accounting for the number of co-authors and their order when assessing research performance at the individual level in the life sciences*. *Journal of Informetrics*, 7, 198–208.
- Allen, L., Brand, A., Scott, J., Altman, M., & Hlava, M. (2014). *Credit where credit is due*. *Nature*, 508, 312–313.
- Araújo, E., Araújo, N., Moreira, A., Herrmann, H., & Andrade, J. (2017). *Gender differences in scientific collaborations: Women are more egalitarian than men*. *PLoS One*, 12, e0176791.
- Beaudry, C., & Larivière, V. (2016). *Which gender gap? Factors affecting researchers' scientific impact in science and medicine*. *Research Policy*, 45, 1790–1817.
- Berker, Y. (2018). *Golden-ratio as a substitute to geometric and harmonic counting to determine multi-author publication credit*. *Scientometrics*, 114, 839–857.
- Bornmann, L., & Osório, A. (2019). *The value and credits of n-authors publications*. *Journal of Informetrics*, 13, 540–554.
- Bornmann, L., Ganser, C., Tekles, A., & Leydesdorff, L. (2019). *Does the H<sub>a</sub> index reinforce the Matthew Effect in science? Agent-based simulations using Stata and R* Accessed at: <https://arxiv.org/abs/1905.11052>
- Boyle, P., Smith, L., Cooper, N., Williams, K., & O'Connor, H. (2015). *Gender balance: Women are funded more fairly in social science*. *Nature News*, 525, 181.
- Bozeman, B., & Gaughan, M. (2011). *How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers*. *Research Policy*, 40, 1393–1402.
- Brooks, C., Fenton, E., & Walker, J. (2014). *Gender and the evaluation of research*. *Research Policy*, 43, 990–1001.
- Brybaert, M., & Smyth, S. (2011). *Self-enhancement in scientific research: The self-citation bias*. *Psychologica Belgica*, 51, 129–137.
- Cameron, E., White, A., & Gray, M. (2016). *Solving the productivity and impact puzzle: Do men outperform women, or are metrics biased?* *BioScience*, 66, 245–252.
- Cole, J., & Cole, S. (1973). *Social stratification in science*. Chicago: University of Chicago Press.
- Crespo, N., & Simoes, N. (2019). *Publication performance and number of authors – Evidence for world top economists* MPRA Working Paper No. 92581.
- De Paola, M., Ponzio, M., & Scoppa, V. (2018). *Are men given priority for top jobs? Investigating the glass ceiling in Italian Academia*. *Journal of Human Capital*, 12, 475–503.
- Dorta-González, P., & Dorta-González, M. (2011). *Central indexes to the citation distribution: A complement to the h-index*. *Scientometrics*, 88, 729.
- Efthymiou, G. (2008). *Alphabet economics: The link between names and reputation*. *The Journal of Socio-economics*, 37, 1266–1285.
- Egghe, L. (2006). *Theory and practise of the g-index*. *Scientometrics*, 69, 131–152.
- Egghe, L. (2008). *Mathematical theory of the h- and g-index in case of fractional counting of authorship*. *Journal of the American Society for Information Science and Technology*, 59, 1608–1616.
- Egghe, L. (2010). *The Hirsch index and related impact measures*. *Annual Review of Information Science and Technology*, 44, 65–114.
- Egghe, L., Guns, R., & Rousseau, R. (2013). *Measuring co-authors' contribution to an article's visibility*. *Scientometrics*, 95, 55–67.
- Engers, M., Gans, J., Grant, S., & King, S. (1999). *First-author conditions*. *The Journal of Political Economy*, 107, 859–883.
- Fang, H. (2018). *Normalized paper credit assignment: A solution for the ethical dilemma induced by multiple important authors*. *Science and Engineering Ethics*, 24, 1589–1601.
- Fernandes, J., & Monteiro, M. (2017). *Evolution in the number of authors of computer science publications*. *Scientometrics*, 110, 529–539.
- Frandsen, T., & Nicolaisen, J. (2010). *What is in a name? Credit assignment practices in different disciplines*. *Journal of Informetrics*, 4, 608–617.
- Geraci, L., Balsis, S., & Busch, A. (2015). *Gender and the h index in psychology*. *Scientometrics*, 105, 2023–2034.
- Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). *A concise review on the role of author self-citations in information science, bibliometrics and science policy*. *Scientometrics*, 67, 263–277.
- Goodall, A. (2012). *A theory of expert leadership* IZA Discussion Papers 6566.
- Hamermesh, D. (2018). *Citations in economics: Measurement, uses, and impacts*. *Journal of Economic Literature*, 56, 115–156.

- Hammarfelt, B., & Rushforth, A. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26, 169–180.
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometric: The Leiden manifesto for research metrics. *Nature*, 520, 429–431.
- Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569–16572.
- Hirsch, J. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, 85, 741–754.
- Hirsch, J. (2019). H<sub>ox</sub>: An index to quantify an individual's scientific leadership. *Scientometrics*, 118, 673–686.
- Hou, W., Li, M., & Niu, D. (2011). Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution. *BioEssays*, 33, 724–727.
- Hu, X. (2009). Loads of special authorship functions: Linear growth in the percentage of 'equal first authors' and corresponding authors. *Journal of the American Society for Information Science and Technology*, 60, 2378–2381.
- Hu, X., Rousseau, R., & Chen, J. (2010). In those fields where multiple authorship is the rule, the h-index should be supplemented by role-based h-indices. *Journal of Information Science*, 36, 73–85.
- Ioannidis, J., Baas, J., Klavans, R., & Boyack, K. (2019). A standardized citation metrics author database annotated for scientific field. *PLoS Biology*, 17, e3000384.
- Kadel, A., & Walter, A. (2015). Do scholars in economics and finance react to alphabetical discrimination? *Finance Research Letters*, 14, 64–68.
- Kosmulski, M. (2012). The order in the lists of authors in multi-author papers revisited. *Journal of Informetrics*, 6, 639–644.
- Kuld, L., & O'Hagan, J. (2018). Rise of multi-authored papers in economics: Demise of the 'lone star' and why? *Scientometrics*, 114, 1207–1225.
- Laband, D., & Tollison, R. (2000). Intellectual collaboration. *The Journal of Political Economy*, 108, 632–662.
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. (2013). Global gender disparities in science. *Nature*, 504, 211–213.
- Larivière, V., Gingras, Y., Sugimoto, C., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66, 1323–1332.
- Lazear, E. (2012). Leadership: A personnel economics approach. *Labour Economics*, 19, 92–101.
- Leahey, E. (2016). From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annual Review of Sociology*, 42, 81–100.
- Leahey, E., & Reikowsky, R. (2008). Research specialization and collaboration patterns in Sociology. *Social Studies of Science*, 38, 425–440.
- Leydesdorff, L., Bornmann, L., & Opthof, T. (2019). H<sub>ox</sub>: The scientist as chimpanzee or bonobo. *Scientometrics*, 118, 1163–1166.
- Lincoln, A., Pincus, S., Koster, J., & Leboy, P. (2012). The Matilda effect in science: Awards and prizes in the United States, 1990s and 2000s. *Social Studies of Science*, 42, 341–343.
- Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science*, 10, 145–162.
- Liu, X., & Fang, H. (2012). Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution. *Journal of Informetrics*, 6, 557–565.
- Lukovits, I., & Vinkler, P. (1995). Correct credit distribution: A model for sharing credit among coauthors. *Social Indicators Research*, 36, 91–98.
- Marušić, A., Bošnjak, L., & Jerončić, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS One*, 6, e23477.
- Mayer, S., & Rathmann, J. (2018). How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. *Scientometrics*, 117, 1663–1693.
- Moed, H. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science & Public Policy*, 34, 575–583.
- Moss-Racusin, C., Dovidio, J., Brescoll, V., Graham, M., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109, 16474–16479.
- Osório, A. (2018). On the impossibility of a perfect counting method to allocate the credits of multi-authored publications. *Scientometrics*, 116, 2161–2173.
- Pichappan, P., & Sarasvady, S. (2002). The other side of the coin: The intricacies of author self-citations. *Scientometrics*, 54, 285–290.
- QS. (2018). *QS world university rankings by subject*. Assessed 1 October. <https://www.topuniversities.com/university-rankings/university-subject-rankings/2018/economics-econometrics>
- Rahman, M., Regenstein, J., Kassim, N., & Haque, N. (2017). The need to quantify authors' relative intellectual contributions in a multi-author paper. *Journal of Informetrics*, 11, 275–281.
- Schreiber, M. (2018). A skeptical view on the Hirsch index and its predictive power. *Physica Scripta*, 93, 102501.
- Seeber, M., Cattaneo, M., Meoli, M., & Malighetti, P. (2019). Self-citations as strategic response to the use of metrics for career decisions. *Research Policy*, 48, 478–491.
- Shen, H. (2013). Mind the gender gap. *Nature*, 495, 22–24.
- Stephan, P. (2012). *How economics shapes science* (Vol. 1) Cambridge, MA: Harvard University Press.
- Tietze, A., & Hofmann, P. (2019). The h-index and multi-author hm-index for individual researchers in condensed matter physics. *Scientometrics*, 119, 171–185.
- Todeschini, R., & Baccini, A. (2016). *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*. Weinheim: John Wiley & Sons.
- Tscharntke, T., Hochberg, M., Rand, T., Resh, V., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. *PLoS Biology*, 5, 18.
- Van Noorden, R., & Chawla, D. (2019). Policing self-citations. *Nature*, 572, 578–579.
- 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.
- Vavryčuk, V. (2018). Fair ranking of researchers and research teams. *PLoS One*, 13, e0195509.
- Wan, J., Hua, P., & Rousseau, R. (2007). The pure h-index: Calculating an author's h-index by taking co-authors into account. *Collnet Journal of Scientometrics and Information Management*, 1, 1–5.
- Wildgaard, L., Schneider, J., & Larsen, B. (2014). A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*, 101, 125–158.
- Zhao, D., Strotmann, A., & Cappello, A. (2018). In-text function of author self-citations: Implications for research evaluation practice. *Journal of the Association for Information Science and Technology*, 69, 949–952.