Journal of Informetrics 12 (2018) 868–892

Contents lists available at ScienceDirect

Journal of Informetrics

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

Individual and field citation distributions in 29 broad scientific fields☆

Javier Ruiz-Castillo[a,*], Rodrigo Costas[b]

[a] Departamento de Economía, Universidad Carlos III of Madrid, Spain
[b] Centre for Science and Technology Studies, Leiden University, The Netherlands

ARTICLE INFO

Article history:
Received 8 March 2018
Received in revised form 3 July 2018
Accepted 3 July 2018
Available online 30 July 2018

Keywords:
Citation distributions
Individual authors
Scientific fields
Skewness of science

ABSTRACT

Using an initial dataset consisting of 18.5 million distinct authors and 15 million distinct articles published in the period 2000–2016, which are classified into 29 broad scientific fields, we search for regularities at the individual level for very productive authors with citation distributions of a certain size, and for the existence of a macro-micro relationship between the skewness of a scientific field citation distribution and the characteristics of the individual citation distributions of the authors belonging to the field. Our main results are the following three. Firstly, although the skewness of individual citation distributions varies greatly within each field, their average skewness is of a similar order of magnitude in all fields. Secondly, as in the previous literature, field citation distributions are highly skewed and the degree of skewness is very similar across fields. Thirdly, the skewness of field citation distributions is essentially explained in terms of the average skewness of individual authors, as well as individuals' differences in mean citation rates and the number of publications per author. These results have important conceptual and practical consequences: to understand the skewness of field citation distributions at any aggregate level we must simply explain the skewness of the individual citation distributions of their very productive authors.

1. Introduction

At any aggregation level, bibliometric studies using citation counts may reveal statistically significant macro-patterns in the communication process that cannot be seen from the limited perspective of the individual researcher in peer review exercises. Given a classification system of publications in the periodical literature into a set of scientific fields, field citation distributions consist of the citation counts for all publications in each field. Similarly, an individual citation distribution consists of the citation counts for an author's publications. In this paper, we search for regularities at the level of individual authors, and the nature of the macro-micro relationship between a field citation distribution and the individual citation distributions of the authors in the field.

Information available for large datasets for journal-based or publication-level classification systems indicates that field citation distributions are typically highly skewed in the sense that a large proportion of articles receives no or few citations while a small percentage of them account for a disproportionate amount of all citations. Furthermore, in spite of

☆ The peer review process of this paper was handled by Vincent Larivière, Associate Editor of Journal of Informetrics.

* Corresponding author.
E-mail address: javierruizc8@gmail.com (J. Ruiz-Castillo).

https://doi.org/10.1016/j.joi.2018.07.002
1751-1577/

wide differences in production and citation practices, the degree of skewness is very similar across fields at very different aggregation or granularity levels (Albarrán & Ruiz-Castillo, 2011; Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011; Glänzel, 2007; Li, Castellano, Radicchi, & Ruiz-Castillo, 2013; Radicchi, Fortunato, & Castellano, 2008; Radicchi & Castellano, 2012; Ruiz-Castillo & Waltman, 2015; Schubert, Glänzel, & Braun, 1987). In so far as field citation distributions consist of articles published by individual authors, the three following questions naturally arise.

Firstly, which are the basic characteristics of individual citation distributions? In particular, are they typically as skewed as field citation distributions, or are they normally, uniformly, or otherwise symmetrically distributed? Alternatively, are authors so different that it is impossible to assign them any systematic pattern at all? Clearly, independently of the conceptual interest of answering this first question, the existence of a typical pattern for individual citation distributions in a given field would facilitate the choice of reference standards for the citation-impact assessment of individual authors as recently suggested, for example, in Thijs, Debackere, and Glänzel (2017).

Secondly, is there any relationship between the characteristics of individual citation distributions in a given field and the characteristics of the field citation distribution of the articles they publish? In the context of science as a system of highly interconnected entities at different levels (individual researchers, research groups, university departments, research institutes, universities), Costas, Bordons, van Leeuwen, and van Raan (2009) have emphasized the importance in large networked systems of the relations between large-scale attributes and local patterns (i.e. between field and individual citation distributions in our case). More generally, Katz (2016) views the global research system as a complex innovation system exhibiting a variety of scale-invariant properties that are statistically similar at many levels of observation. Costas et al. (2009) study the scaling relationship between the number of citations and the number of scientific publications. Specifically, they investigate whether the scaling behavior identified at the research group level (van Raan, 2006a, 2006b, 2008) is also observed at the individual level. As for Katz (2016), he studies scale-invariant correlations between the growth of impact and size over time, and between impact and size across fields and sub-fields at a point in time. In this paper, we investigate the possibility of explaining the skewness of field citation distributions in terms of the characteristics of individual citation distributions. For assessing the skewness of citation distributions, we use the Characteristic Scores and Scales (CSS hereafter) technique for grouping ranked observations into ranked-specific categories (Glänzel & Schubert, 1988; Schubert et al., 1987).

Thirdly, is the macro-micro relationship between the characteristics of field and individual citation distributions common to all sciences, or is the authors' research experience quite different in more basic or more applied fields, in fields with a high or a low citation-density, or in the natural, the engineering and the social sciences?

These are key questions for understanding the communication process in any science. However, the systematic study of the characteristics of individual citation distributions has traditionally been hampered by the lack of appropriate information. In this paper, we have largely overcome this difficulty by constructing a large dataset along the lines initiated in Ruiz-Castillo and Costas (2014a) –RCC hereafter. Our initial dataset consists of 15 million distinct articles indexed by Clarivate Analytics, formerly the IP & Science business of Thomson Reuters, and published by 18.5 million distinct authors in the period 2000-2016. Citations of articles published in a given year are recorded up to the year 2016 in a variable citation window.

It should be noted that, in order to study the skewness of entire citation distributions at the individual level, we must ignore authors with few publications. Thus, as in Thijs et al. (2017), we must restrict our attention to researchers with a citation distribution of a certain size. Specifically, we focus on *very productive* authors with a number of publications above a certain relative benchmark that takes into account that the average number of articles per author varies widely across fields. We also consider *merely productive* authors, defined as those who publish at least five articles during our 16-year period. On average over all fields, these two types of productive authors only represent 5.2% and 9.4% of the population but are responsible for 38.0% and 47.9% of all publications.

The remainder of the paper is organized into five Sections and an Appendix A. Section 2 presents the data, the notation, some descriptive statistics, and a brief description of the CSS method. Section 3 contains the within- and between-field results concerning individual citation distributions among very productive authors. Section 4 presents the within- and between-field results concerning the macro-micro relationship between field and individual citation distributions with the help of some illustrative examples presented in the Appendix A. Section 5 discusses the main findings of the paper, while Section 6 offers some concluding comments.

In order to facilitate the reading of the paper, three issues have been relegated to a Supplementary Material section. In Part 1 of the Supplementary Material, we assess the reliability of our dataset by comparing some of its key characteristics with those of the RCC dataset. Part 2 of the Supplementary Material is devoted to the following problem. We solve the assignment of individual responsibility in cases of co-authorship in a multiplicative manner. However, previous research on field citation distributions need not contend with this problem. This means that the size of field citation distributions in the two cases are very different. Fortunately, in Part 2 of the Supplementary Material we establish that the characteristics of field citation distributions are independent of the co-authorship problem. Finally, in Part 3 of the Supplementary Material we study the robustness of our results in Section 3 for individual citation distributions when we consider merely productive authors.

2. Data, descriptive statistics, and methods

2.1. The construction of the dataset

Since we wish to address a homogeneous population, we only study research articles published in academic journals or, simply, *articles*.[1] We begin with a large sample, consisting of 15,047,087 distinct articles published in the period 2000-2016.

To pursue our study, we must confront the following four methodological problems: the classification of articles into scientific fields; the identification of the author(s) of each article; the allocation of authors to fields, and the attribution of individual responsibility in cases of multiple authorship. Since we solve these problems exactly as we did in RCC, in this Sub-section we briefly discuss the issues involved. A more detailed justification of the solutions in each case can be found in our previous contribution.

1 There are two main approaches to tackling the problem created by the assignment of publications to two or more journal subject categories, or simply categories, in WoS datasets. The first is a fractional strategy, where each publication is fractioned into as many equal pieces as necessary with each piece assigned to its corresponding category. The second approach follows a multiplicative strategy in which each paper is counted as many times as necessary in the several categories to which it is assigned. In this way, the space of articles is expanded as much as necessary beyond the initial size in what we call the *extended count*. Fortunately, previous results indicate that for many purposes, journals assigned to a single or several subject categories share similar characteristics, so that the choice between the two strategies is not that crucial (see RCC for references). In this paper we follow a multiplicative approach. Consequently, the number of articles in the extended count, denoted by N, is 21,202,678, or 34.1% larger than the number of distinct articles. We adopt the classification system used in RCC, consisting of 30 broad fields, which is based in a partition of scientific activity into 35 fields introduced by Tijssen, Hollanders and van Steen (2010) and used in other publications (see RCC for references). However, in contrast with RCC, here we remove the heterogeneous 'Multidisciplinary journals' category by proportionally classifying these publications in the fields of the cited references. Therefore, we distinguish between 29 fields.[2]

2 For the assignment of articles to individual authors, we use the author disambiguation algorithm generated by Caron and van Eck (2014) for large bibliometric databases, whose main features are discussed in RCC. Overall, there are 18,526,987 distinct researchers associated to the 15 million distinct articles of the dataset.

3 For the purpose of analyzing the characteristics of individual citation distributions in a given field, as we do in this paper, researchers who write articles in several fields should be treated as independent, different authors in their respective fields. Therefore, the number of authors, denoted by I, goes up to 35,057,987 individuals, an 89.2% increase relative to the original number of distinct authors.

4 A fundamental difficulty in the study of scientists' productivity is the definition of the individual contribution to an article in a world dominated by co-authorship in all fields (see the references in RCC, as well as the recent contributions by Perianes-Rodríguez & Ruiz-Castillo, 2015a; Waltman & Van Eck, 2015). In this paper, we use a multiplicative strategy in which any article co-authored by two or more scholars is wholly assigned as many times as necessary to each of them. Of course, this means that the set of articles actually studied increases quite dramatically: the total number of articles in what we call the *double extended count*, denoted by N^D, becomes 105,289,384, or seven times larger than the number of distinct articles.[3] The total number of citations in the double extended count is 2.102 million, or nine times larger than the initial number of citations for the 15 million distinct articles.

2.2. Descriptive statistics

We denote by N_f and $N_f{}^D$ the number of articles in each field in the extended and the double extended count, so that $\Sigma_f N_f = N = 21.1$ and $\Sigma_f N_f{}^D = N^D = 105.3$ million articles. Similarly, we denote by I_f the number of authors in each field, so that $\Sigma_f I_f = N = 35.1$ million authors. Table 1 presents the distribution of articles by field in the extended and double extended counts, as well as the distribution of authors by field, whereas Table 2 includes some evidence on the variability of co-authorship patterns within and between fields.

In this paper, the within- and between-field variation for all magnitudes is measured by the coefficient of variation (CV hereafter) over the 29 fields. The CV is defined as the ratio of the standard deviation over the mean. There is no generally agreed upon criterion in statistics concerning when a CV is "large" or "small", possibly because this distinction is context dependent. Although any reader is free to apply a different criterion, in this paper we will use the following convention. We say that the within- or between-field variability of any characteristic is

[1] Following Waltman and Van Eck (2013a, 2013b), we exclude publications in local journals, as well as magazine and trade journals.

[2] It is not claimed that this scheme provides the best possible representation of the structure of science. It is rather a convenient simplification for the discussion of field comparability issues in this paper.

[3] In comparison, this dataset is approximately twice as large as the one used in RRC. Specifically, RCC begin with 7.7 million distinct articles published in the period 2003–2011 by 9.6 million distinct authors and end up with 17.2 million authors and 48.2 million articles.

Table 1
Number of articles, and number of authors by scientific field, and average and coefficient of variation of these magnitudes over the 29 fields.

	N_f (1)	% (2)	I_f (3)	% (4)	$N^D{}_f$ (5)	% (6)
AGRICULTURE & FOOD SCIENCE	570,298	2.7	1,144,697	3.3	2,712,680	2.6
ASTRONOMY & ASTROPHYSICS	259,174	1.2	296,611	0.8	3,136,494	3.0
BASIC LIFE SCIENCES	1,918,845	9.1	3,827,339	10.9	11,123,668	10.6
BASIC MEDICAL SCIENCES	342,397	1.6	950,157	2.7	1,961,761	1.9
BIOLOGICAL SCIENCES	1,025,326	4.8	1,815,327	5.2	4,610,162	4.4
BIOMEDICAL SCIENCES	1,914,173	9.0	3,898,839	11.1	11,232,151	10.7
CHEMISTRY & CHEMICAL ENG.	2,337,898	11.0	3,466,432	9.9	10,521,685	10.0
CIVIL ENG. & CONSTRUCTION	155,484	0.7	259,045	0.7	501,431	0.5
CLINICAL MEDICINE	3,592,283	16.9	6,378,554	18.2	21,348,154	20.3
COMPUTER SCIENCES	575,536	2.7	828,925	2.4	1,812,869	1.7
EARTH SCIENCES & TECHNOLOGY	622,877	2.9	818,997	2.3	2,541,572	2.4
ECONOMICS & BUSINESS	262,412	1.2	240,653	0.7	588,016	0.6
EDUCATIONAL SCIENCES	146,675	0.7	245,618	0.7	422,657	0.4
ELECTRICAL ENG. & TELECOMM.	680,973	3.2	1,027,701	2.9	2,471,759	2.3
ENERGY SC. & TECHNOLOGY	330,577	1.6	676,827	1.9	1,582,464	1.5
ENVIRONMENTAL SCS & TECH.	867,052	4.1	1,368,240	3.9	3,461,632	3.3
GENERAL & INDUSTRIAL ENG.	190,700	0.9	337,448	1.0	592,217	0.6
HEALTH SCIENCES	455,500	2.1	891,728	2.5	1,983,543	1.9
INFORMATION & COMM. SCS.	63,516	0.3	94,965	0.3	156,287	0.1
INSTS. & INSTRUMENTATION	180,699	0.9	463,082	1.3	1,085,429	1.0
LAW & CRIMINOLOGY	75,480	0.4	107,371	0.3	193,701	0.2
MANAGEMENT & PLANNING	117,093	0.6	146,102	0.4	270,554	0.3
MATHEMATICS	536,382	2.5	410,133	1.2	1,204,809	1.1
MECHANICAL ENG. & AEROSPACE	433,086	2.0	625,175	1.8	1,383,929	1.3
PHYSICS & MATERIALS SCIENCE	2,696,102	12.7	3,514,815	10.0	15,699,147	14.9
PSYCHOLOGY	391,568	1.8	528,432	1.5	1,388,570	1.3
SOCIAL & BEHAVIORAL SCIENCES	93,762	0.4	167,312	0.5	278,710	0.3
SOCIOLOGY & ANTHROPOLOGY	136,174	0.6	208,419	0.6	367,223	0.3
STATISTICAL SCIENCES	230,636	1.1	319,043	0.9	656,110	0.6
TOTAL	21,202,678	100.0	35,057,987	100.0	105,289,384	100.0
Average over the 29 fields	731,127		1,208,896		3,630,668	
Coefficient of variation	1.2		1.3		1.4	

$\boldsymbol{N_f}$ = Number of articles in field *f* in the extended count according to the multiplicative approach, where each article is counted as many times as the number of fields to which it is assigned in the Web of Science.
$\boldsymbol{I_f}$ = Number of authors in field *f* when researchers with articles in tow or more fields are treated as different authors.
$\boldsymbol{ND_f}$ = Number of articles in field *f* in the double extended count according to the multiplicative approach, where each article in the extended count is counted as many times as the number of its authors.

- "Small", if CV ≤ 0.10, meaning that the dispersion of this characteristic measured by the standard deviation is smaller than or equal to 10% of the mean.
- "Intermediate", if 0.10 < CV ≤ 0.30.
- "Large", if 0.30 < CV ≤ 0.60.
- "Very large", if CV > 0.60.

The following three points should be noted. Firstly, according to the number of authors, fields can be classified into three groups (see column 3 in Table 1). (i) There are five fields with more than three million authors with at least 9.9% of the total number of authors (Clinical Medicine; Biomedical Sciences; Basic Life Sciences; Physics & Materials Science, and Chemistry & Chemical Engineering). The largest is Clinical Medicine that has 6.4 million authors and 18.2% of the total. (ii) There are eleven intermediate fields with 528,000–1,815,000 authors, or 1.5%–5.2% of the total. (iii) The remaining fifteen fields have fewer than 364,000 authors or 1.3% of the total. The smallest is Information & Communication Sciences with 94,965 authors, or 0.3% of the total. In view of this partition, the dispersion of field sizes is very large: the CV over the 29 fields is 1.3.[4]

Secondly, the average number of authors per article is 4.2 (column 1 in Table 2). However, the between-field variation is quite large: the coefficient of variation over the 29 fields is 0.46, and the range of variation goes from 2.2 and 2.3 authors per article in Mathematics and Management & Planning, up to 6.0 and 12.1 in Instruments & Instrumentation and Astronomy & Astrophysics. On the other hand, the within-field variation is very large indeed (column 2 in Table 2), ranging from a coefficient of variation of 0.51 in General & Industrial Engineering up to 8.50 and 8.54 in Physics & Materials Science and Astronomy & Astrophysics. Finally, the maximum number of authors per article (column 3 in Table 2) exhibits a phenomenal range of variation from 57 and 73 in General & Industrial Engineering and Management & Planning, up to 3195 and 5109 in Astronomy & Astrophysics and Physics & Materials Science.

[4] Between-field variation when size is measured as the number of articles is also very high indeed: in these cases, the coefficients of variation over the 29 fields are 1.2 in the extended count and 1.4 in the double extended count.

Table 2
Average, coefficient of variation, and maximum number of authors per article, and average and coefficient of variation of these magnitudes over the 29 fields.

	Number of authors per article:		
	Average (1)	CV (2)	Maximum (3)
AGRICULTURE & FOOD SCIENCE	4.8	0.67	479
ASTRONOMY & ASTROPHYSICS	12.1	8.50	3195
BASIC LIFE SCIENCES	5.8	0.82	1014
BASIC MEDICAL SCIENCES	5.7	0.64	404
BIOLOGICAL SCIENCES	4.5	0.91	769
BIOMEDICAL SCIENCES	5.9	0.69	1010
CHEMISTRY & CHEMICAL ENG.	4.5	0.54	341
CIVIL ENG. & CONSTRUCTION	3.2	0.54	133
CLINICAL MEDICINE	5.9	0.86	2458
COMPUTER SCIENCES	3.1	1.89	3035
EARTH SCIENCES & TECHNOLOGY	4.1	0.84	496
ECONOMICS & BUSINESS	2.2	0.77	479
EDUCATIONAL SCIENCES	2.9	0.75	192
ELECTRICAL ENG. & TELECOMM.	3.6	0.66	385
ENERGY SC. & TECHNOLOGY	4.8	2.41	3035
ENVIRONMENTAL SCS & TECH.	4.0	0.74	415
GENERAL & INDUSTRIAL ENG.	3.1	0.51	57
HEALTH SCIENCES	4.4	0.69	283
INFORMATION & COMM. SCS.	2.5	0.77	86
INSTS. & INSTRUMENTATION	6.0	7.01	3035
LAW & CRIMINOLOGY	2.6	0.97	163
MANAGEMENT & PLANNING	2.3	0.63	73
MATHEMATICS	2.2	0.62	287
MECHANICAL ENG. & AEROSPACE	3.2	0.55	174
PHYSICS & MATERIALS SCIENCE	5.8	8.54	5109
PSYCHOLOGY	3.5	0.80	479
SOCIAL & BEHAVIORAL SCIENCES	3.0	0.89	198
SOCIOLOGY & ANTHROPOLOGY	2.7	0.94	201
STATISTICAL SCIENCES	2.8	1.33	598
Average over the 29 fields	4.2	1.6	985.6
Coefficient of variation	0.46	1.39	1.29

Thirdly, the increase in the total number of articles in the double extended count varies a lot across fields. Comparing columns 2 and 6 in Table 1, we observe that the percentage of the number of articles in the double extended count is greater than in the original count in only seven fields whose mean number of authors per article (column 1 in Table 2) is well above the average for all fields (Astronomy & Astrophysics; Basic Life Sciences; Basic Medical Sciences; Biomedical Sciences; Clinical Medicine; Instruments & Instrumentation, and Physics & Materials Science).

The construction of large datasets for the study of the research performance of individual authors is a daunting empirical exercise. Since we do not know of any other large dataset in the literature with which we can compare our own, it seems convenient to assess its reliability by comparing some of its characteristics with those of the dataset used in the RCC contribution, which, as indicated in note 3, is half as large as ours. To facilitate the reading of the paper, this exercise is included in **Part 1. The reliability of our dataset** in the Supplementary Material. The high degree of consistency observed for all characteristics demonstrates the reliability of the present construction: having followed the same criteria in both cases, the two datasets seem to reflect the same world.

2.3. Methods: the CSS approach

As indicated in the Introduction, in this paper we focus on a key characteristic emphasized since the inception of scientometrics by Price (1965) and Seglen (1992), namely, the skewness of citation distributions. At the field level, we should take into account that wide differences in production and citation practices across fields greatly affect the size and the mean of field citation distributions. Similarly, differences in individual productivity and citation impact among authors in a given field give rise to wide differences in the size and mean of individual citation distributions. Therefore, it seems convenient to evaluate the skewness of citation distributions abstracting from size and mean differences across fields and individual authors. For that purpose, we use CSS technique, which is size- and scale-independent, and can be briefly described as follows.

Let N be the number of elements in any citation distribution X, indexed by $k = 1, \ldots, N$, so that $X = (x_1, \ldots, x_k, \ldots, x_N)$ where x_k is the number of citations received by publication k. For later reference, let $\Gamma(X)$ be the total number of citations in X, i.e. $\Gamma(X) = \Sigma_k x_k$. Two *characteristic scores* will be used: m_1, the mean of X, and m_2, the second mean of X, or the mean of all publications in X with a number of citations greater than m_1. Using m_1 and m_2, we define the following three categories:

Table 3
First and second means of the field individual productivity (number of articles per author) distributions for all authors in the double extended count, and average and coefficient of variation of these magnitudes over the 29 fields\.

	First. mean (1)	CV (2)	Second mean (3)	CV (4)
AGRICULTURE & FOOD SCIENCE	2.4	2.22	8.3	1.32
ASTRONOMY & ASTROPHYSICS	10.6	3.37	65.4	1.15
BASIC LIFE SCIENCES	2.9	2.34	9.2	1.37
BASIC MEDICAL SCIENCES	2.1	1.97	6.9	1.25
BIOLOGICAL SCIENCES	2.5	2.08	8.3	1.22
BIOMEDICAL SCIENCES	2.9	2.38	9.4	1.38
CHEMISTRY & CHEMICAL ENG.	3.0	3.14	14.0	1.60
CIVIL ENG. & CONSTRUCTION	1.9	1.85	4.7	1.36
CLINICAL MEDICINE	3.3	3.08	15.2	1.52
COMPUTER SCIENCES	2.2	1.97	7.5	1.18
EARTH SCIENCES & TECHNOLOGY	3.1	2.39	13.0	1.17
ECONOMICS & BUSINESS	2.4	1.71	7.6	0.96
EDUCATIONAL SCIENCES	1.7	1.63	4.0	1.26
ELECTRICAL ENG. & TELECOMM.	2.4	2.33	8.4	1.39
ENERGY SC. & TECHNOLOGY	2.3	1.99	7.7	1.18
ENVIRONMENTAL SCS & TECH.	2.5	2.18	8.5	1.27
GENERAL & INDUSTRIAL ENG.	1.8	1.62	4.2	1.22
HEALTH SCIENCES	2.2	2.12	7.9	1.27
INFORMATION & COMM. SCS.	1.6	1.50	3.9	1.16
INSTS. & INSTRUMENTATION	2.3	1.96	8.1	1.11
LAW & CRIMINOLOGY	1.8	1.79	4.4	1.34
MANAGEMENT & PLANNING	1.9	1.39	4.2	0.99
MATHEMATICS	2.9	2.27	9.5	1.30
MECHANICAL ENG. & AEROSPACE	2.2	2.08	7.8	1.24
PHYSICS & MATERIALS SCIENCE	4.5	5.41	27.3	2.37
PSYCHOLOGY	2.6	2.27	9.0	1.31
SOCIAL & BEHAVIORAL SCIENCES	1.7	1.30	3.8	0.98
SOCIOLOGY & ANTHROPOLOGY	1.8	1.37	4.0	1.02
STATISTICAL SCIENCES	2.1	1.98	7.5	1.19
Average over the 29 fields	2.7	2.19	10.3	1.28
Coefficient of variation	0.60	0.35	1.10	0.20

category I consists of the proportion of *poorly cited* publications in X with x_k smaller than or equal to m_1; category II consists of the proportion of *fairly cited* publications in X with x_k greater than m_1 and smaller or equal to m_2, and category III consists of the proportion of *remarkably* or *outstandingly cited* publications in X with x_k greater than m_2. CSS results consists of six numbers, (p_1, p_2, p_3) and (s_1, s_2, s_3), where p_j, j = 1, 2, 3 is the proportion of publications in X in categories I, II, and III, and s_j, j = 1, 2, 3 is the share of $\Gamma(X)$ accounted by categories I, II, and III. In many cases, we will typically have CSS results at the field level, say (p_{f1}, p_{f2}, p_{f3}) and (s_{f1}, s_{f2}, s_{f3}), for f = 1,. . ., 29. We denote the average of the CSS results over the 29 fields by capital letters, i.e. (P_1, P_2, P_3) and (S_1, S_2, S_3). As before, the between-field variation of these magnitudes is measured by means of the CV over the 29 fields.

3. Within- and between-field results concerning individual citation distributions

3.1. Very productive authors

In each field f = 1,. . ., 29, let $c_f(i)$ be the citation distribution of author i with i = 1,. . ., I_f, where I_f is the number of authors in field f. For each i and f, let $n_f(i)$ be the size of $c_f(i)$, i.e. the number of articles of author i in field f. For each f, the first and second means of distribution $\{n_f(i), i = 1,..., I_f\}$ are presented in Table 3. Note that in all fields, the average number of articles per author is very low indeed (column 1 in Table 3). This is in agreement with Lotka (1926)'s Law, according to which a majority of scholars contribute rarely to the body of knowledge and the number of contributions per author decreases exponentially. In particular, on average, authors with a single publication in our 16-year publication period represent 71.3% of the total, whereas more than 90% of all authors have less than five publications (Table 4). Possibly, the decision to treat researchers with publications in two or more fields as different authors increases the percentage of individuals with few publications in

Table 4
Percentage distribution of authors classified by the number of articles they have published, and average and coefficient of variation of these magnitudes over the 29 fields.

	Number of publications per author:					
	One (1)	Two (2)	Three or four (3)	Five or more (4)	Total (5)	Percentage of total articles published by authors in columns (4) (6)
AGRICULTURE & FOOD SCIENCE	70.7	12.0	8.1	9.1	100.0	48.5
ASTRONOMY & ASTROPHYSICS	65.1	7.7	6.4	20.7	100.0	90.3
BASIC LIFE SCIENCES	66.4	11.8	9.3	12.5	100.0	58.2
BASIC MEDICAL SCIENCES	71.8	12.3	8.3	7.6	100.0	39.9
BIOLOGICAL SCIENCES	68.0	12.5	8.9	10.6	100.0	51.5
BIOMEDICAL SCIENCES	67.5	11.5	8.9	12.1	100.0	58.2
CHEMISTRY & CHEMICAL ENG.	70.6	10.3	8.0	11.1	100.0	61.1
CIVIL ENG. & CONSTRUCTION	74.4	11.8	7.3	6.5	100.0	36.8
CLINICAL MEDICINE	69.5	10.5	7.7	12.3	100.0	65.2
COMPUTER SCIENCES	71.3	12.4	8.1	8.3	100.0	43.7
EARTH SCIENCES & TECHNOLOGY	67.9	11.1	8.3	12.7	100.0	62.0
ECONOMICS & BUSINESS	68.0	11.9	8.8	11.4	100.0	50.3
EDUCATIONAL SCIENCES	76.3	11.7	6.7	5.2	100.0	28.9
ELECTRICAL ENG. & TELECOMM.	70.7	11.9	8.2	9.2	100.0	49.2
ENERGY SC. & TECHNOLOGY	69.8	12.0	8.6	9.6	100.0	47.5
ENVIRONMENTAL SCS & TECH.	69.3	11.9	8.5	10.4	100.0	52.0
GENERAL & INDUSTRIAL ENG.	76.3	11.4	6.8	5.4	100.0	30.6
HEALTH SCIENCES	71.9	12.0	7.7	8.3	100.0	45.2
INFORMATION & COMM. SCS.	78.0	11.0	6.3	4.7	100.0	26.4
INSTS. & INSTRUMENTATION	70.6	12.3	7.9	9.2	100.0	48.0
LAW & CRIMINOLOGY	76.5	11.1	6.6	5.8	100.0	33.0
MANAGEMENT & PLANNING	73.4	12.0	7.7	6.8	100.0	33.4
MATHEMATICS	66.7	11.8	8.6	12.9	100.0	59.4
MECHANICAL ENG. & AEROSPACE	72.0	11.8	7.9	8.3	100.0	44.8
PHYSICS & MATERIALS SCIENCE	71.8	9.0	7.0	12.2	100.0	74.6
PSYCHOLOGY	68.9	12.2	8.4	10.5	100.0	53.8
SOCIAL & BEHAVIORAL SCIENCES	75.9	12.1	7.0	4.9	100.0	25.8
SOCIOLOGY & ANTHROPOLOGY	74.3	12.3	7.6	5.8	100.0	29.5
STATISTICAL SCIENCES	73.9	11.7	7.1	7.3	100.0	41.1
Average over the 29 fields	71.3	11.5	7.8	9.4	100.0	47.9
Coefficient of variation	0.05	0.09	0.10	0.36	0.00	0.31

their minority field(s).[5] Nevertheless, the low CV in columns 1–3 in Table 4, representing authors with less than five articles, indicates the existence of a surprising similarity across fields as far as low publication rates are concerned.[6]

This poses a problem for the analysis of individual citation distributions: we are bound to restricting our attention to a very small percentage of authors with citation distributions of a certain minimum size. At any rate, how should we determine such a minimum size in each field? Note that differences in production practices at high publication rates give rise to considerable between-field variation in mean individual productivity: the CV over the 29 fields in column 1 in Table 3 is 0.61. Thus, for example, mean individual productivity is equal to 1.6 and 1.7 articles per author in Information & Communication Sciences and Social & Behavioral Sciences, while this magnitude is 3.3, 4.5, and 10.6 in Clinical Medicine, Physics & Materials Science, and Astronomy & Astrophysics. Therefore, it is natural to search for a benchmark that varies across fields.

In this vein, we define *very productive* authors in each field as those with a number of articles greater than the second mean in the distribution $\{n_f(i), i = 1,\ldots, I_f\}$ (column 3 in Table 3). We denote by $I_f{}^*$ the number of very productive authors in field $f = 1,\ldots, 29$. Although the percentage of very productive authors is generally very small, they typically account for a relatively large percentage of all articles in the double extended count. In Agriculture and Food Science, for example, only 4.3% of all authors with a number of publications equal to or greater than nine will be considered very productive. However, this small percentage is responsible for 36.0% of all articles in the field. On average, very productive authors are 5.3% of the total, publish eleven or more articles *per capita*, and are responsible for 38.0% of all articles (for details, see columns 3 and 6 in Table SM2 in Part 1 of the Supplementary Material).

[5] Moreover, as indicated in RCC, the Caron and van Eck (2014) name disambiguation algorithm promotes precision over recall. Thus, it should be acknowledged that when there is limited information to cluster the publications of a certain author, the algorithm may occasionally split the *ouvre* of an author into clusters with only one publication.

[6] The large percentage of authors with a single publication, the low between-field variation of this amount, as well as the low average number of articles per author are also observed in Table 1 in RCC.

Table 5
Average and coefficient of variation of the mean productivity and the first and second mean citations for very productive authors in each field, and average and coefficient of variation of these magnitudes over the 29 fields.

	Mean-size$_f$ (1)	*CV* (2)	m_{f1} (3)	*CV* (4)	m_{f2} (5)	*CV* (6)
AGRICULTURE & FOOD SCIENCE	19.9	0.85	15.6	1.23	41.6	1.93
ASTRONOMY & ASTROPHYSICS	153.0	0.55	29.4	0.63	126.1	0.94
BASIC LIFE SCIENCES	22.6	0.87	32.4	1.41	96.4	2.43
BASIC MEDICAL SCIENCES	14.4	0.93	16.7	1.14	43.5	2.21
BIOLOGICAL SCIENCES	19.3	0.78	25.3	1.97	79.3	2.95
BIOMEDICAL SCIENCES	22.8	0.87	24.7	1.24	69.1	2.00
CHEMISTRY & CHEMICAL ENG.	36.0	1.01	18.0	1.30	49.5	3.82
CIVIL ENG. & CONSTRUCTION	10.9	0.93	10.9	1.21	24.9	1.42
CLINICAL MEDICINE	39.4	0.92	25.0	1.17	77.3	2.23
COMPUTER SCIENCES	16.7	0.79	9.3	2.05	30.7	3.28
EARTH SCIENCES & TECHNOLOGY	28.4	0.73	19.8	0.94	55.7	1.69
ECONOMICS & BUSINESS	14.9	0.63	17.5	1.20	47.2	1.60
EDUCATIONAL SCIENCES	9.6	0.92	12.6	1.35	30.9	1.63
ELECTRICAL ENG. & TELECOMM.	20.5	0.91	9.4	2.23	30.3	6.61
ENERGY SC. & TECHNOLOGY	16.5	0.81	11.7	1.31	33.4	2.34
ENVIRONMENTAL SCS & TECH.	19.5	0.83	20.2	1.11	55.6	1.96
GENERAL & INDUSTRIAL ENG.	9.9	0.85	9.2	1.02	21.6	1.33
HEALTH SCIENCES	17.5	0.86	15.7	1.06	42.4	2.18
INFORMATION & COMM. SCS.	7.6	0.90	12.9	1.34	33.0	2.27
INSTS. & INSTRUMENTATION	18.9	0.66	9.2	1.47	32.7	2.39
LAW & CRIMINOLOGY	10.3	0.95	10.2	1.09	23.9	2.56
MANAGEMENT & PLANNING	9.0	0.65	20.5	1.16	52.1	1.59
MATHEMATICS	21.4	0.85	6.5	1.43	18.7	2.43
MECHANICAL ENG. & AEROSPACE	17.4	0.83	10.0	1.10	25.3	1.85
PHYSICS & MATERIALS SCIENCE	93.9	1.31	17.1	1.11	55.5	2.57
PSYCHOLOGY	20.6	0.83	20.7	1.07	58.5	1.91
SOCIAL & BEHAVIORAL SCIENCES	7.2	0.74	15.1	1.53	36.6	2.47
SOCIOLOGY & ANTHROPOLOGY	7.4	0.76	15.4	1.40	37.1	1.81
STATISTICAL SCIENCES	16.6	0.80	13.8	1.98	45.2	3.09
	Mean-size	*CV*	M_1	*CV*	M_2	*CV*
Average over the 29 fields	24.9	0.84	16.4	1.32	47.4	2.33
Coefficient of variation	1.18	0.16	0.40	0.27	0.51	0.44

Mean-size$_f$ = average of the sizes of distributions $c_{f1}(i)$, or individual productivities $n_f(i)$, over the $I_f{}^*$ very productive authors in field *f*.
m_{f1} = average of the individual first means $m_{f1}(i)$ of distributions $c_{f1}(i)$ over the $I_f{}^*$ very productive authors in field *f*.
m_{f2} = average of the individual second means $m_{f2}(i)$ of distributions $c_{f1}(i)$ over the $I_f{}^*$ very productive authors in field *f*.

3.2. Within- and between-field variation of individual citation distributions for very productive authors

Recall that $n_f(i)$ is the size of the individual citation distribution $c_f(i)$. For every very productive author *i* in field *f*, let $m_{f1}(i)$ and $m_{f2}(i)$ be the first and second means of $c_f(i)$. For every field, denote the average of these three quantities over the $I_f{}^*$ authors by *Mean-size$_f$*, m_{f1} and m_{f2}; that is, *Mean-size$_f$* = $\Sigma_i\, n_f(i)/I_f{}^*$, and $m_{fj} = \Sigma_i\, m_{fj}(i)/I_f{}^*$ for *j* = 1, 2, where the sum in these expressions goes over the $I_f{}^*$ very productive authors. In turn, the average of *Mean-size$_f$*, m_{f1} and m_{f2} over the 29 fields are denoted by *Mean-size*, M_1 and M_2; that is, *Mean-size* = Σ_f *Mean-size$_f$*/29, and $M_j = \Sigma_f\, m_{f1}/29$ for *j* = 1, 2. The results for all these magnitudes are in Table 5. Wide differences in citation impact among very productive authors will manifest themselves in large CVs of their mean citation rates. This is exactly what we observe for all fields in columns 4 and 6 in Table 5. On the other hand, large CVs over the 29 fields reflect large differences in citation practices across fields.

Of course, given the wide differences between authors' citation impact in each field, and in production and citation practices across fields, large within- and between-field differences in mean citation rates come as no surprise. The key question for our purposes concerns the skewness of individual citation distributions. For every very productive author *i* in field *f*, we denote the CSS results by ($p_{f1}(i)$, $p_{f2}(i)$, $p_{f3}(i)$) and ($s_{f1}(i)$, $s_{f2}(i)$, $s_{f3}(i)$), where $p_{fj}(i)$ is the proportion of articles in distribution $c_f(i)$ in category *j* = I, II, III, and $s_{fj}(i)$ is the share of total citations in distribution $c_f(i)$ accounted for by category *j* = I, II, III. For every field, we denote the average of these individual results over the $I_f{}^*$ authors by (p_{f1}, p_{f2}, p_{f3}) and (s_{f1}, s_{f2}, s_{f3}), that is, for every *j*,

$$p_{fj} = \Sigma_i\ \mathrm{p_{fj}}(i)/I_f*, \tag{1}$$

and

$$s_{fj} = \Sigma_i\ \mathrm{s_{fj}}(i)/I_f*, \tag{2}$$

where the sum in expressions (1) and (2) goes over the $I_f{}^*$ very productive authors. The corresponding CVs over the $I_f{}^*$ very productive authors are denoted by (cv_{f1}, cv_{f2}, cv_{f3}) and (cv_{f4}, cv_{f5}, cv_{f6}), respectively. The results for (p_{f1}, p_{f2}, p_{f3}), (cv_{f1}, cv_{f2},

Table 6
The skewness of individual citation distributions according to the CSS approach for very productive authors in the double extended count, and average and coefficient of variation of the CSS results over the 29 fields.

	p_1 (1)	p_2 (2)	P_3 (3)	cv_1 (4)	$cv_{2.}$ (5)	cv_3 (6)	s_1 (7)	s_2 (8)	s_3 (9)	cv_4 (12)	cv_5 (11)	cv_6 (12)
AGRICULTURE & FOOD SC.	65.8	20.8	13.3	0.14	0.36	0.44	24.2	32.2	43.7	0.34	0.36	0.27
ASTRONOMY & ASTROPHYSICS	76.9	17.3	5.8	0.11	0.26	0.73	25.8	28.6	45.7	0.13	0.20	0.13
BASIC LIFE SCIENCES	67.4	20.1	12.5	0.14	0.36	0.45	25.2	31.1	43.7	0.30	0.34	0.26
BASIC MEDICAL SCIENCES	64.5	21.2	14.3	0.17	0.39	0.47	25.9	32.7	41.4	0.38	0.40	0.34
BIOLOGICAL SCIENCES	66.8	20.3	12.9	0.15	0.37	0.46	24.0	32.0	44.0	0.35	0.38	0.29
BIOMEDICAL SCIENCES	66.7	20.5	12.7	0.14	0.35	0.44	25.4	31.3	43.3	0.30	0.33	0.25
CHEMISTRY & CHEMICAL ENG.	67.1	20.8	12.1	0.12	0.30	0.41	25.0	31.7	43.4	0.27	0.27	0.21
CIVIL ENG. & CONSTRUCTION	63.5	22.0	14.5	0.19	0.39	0.58	22.8	37.1	40.0	0.51	0.48	0.47
CLINICAL MEDICINE	68.8	20.1	11.1	0.12	0.29	0.42	24.9	31.1	44.0	0.24	0.27	0.20
COMPUTER SCIENCES	68.5	19.4	12.2	0.15	0.41	0.52	18.8	34.4	46.5	0.52	0.47	0.35
EARTH SCIENCES & TECH.	67.4	20.5	12.1	0.13	0.31	0.42	24.6	31.4	44.0	0.27	0.29	0.22
ECONOMICS & BUSINESS	66.7	20.1	13.2	0.15	0.38	0.46	22.4	32.6	45.0	0.39	0.39	0.30
EDUCATIONAL SCIENCES	63.8	21.7	14.5	0.19	0.39	0.58	23.2	36.9	39.8	0.50	0.48	0.49
ELECTRICAL ENG. & TEL.	67.4	20.1	12.5	0.15	0.38	0.47	20.7	33.3	46.0	0.44	0.40	0.29
ENERGY SC. & TECHNOLOGY	66.6	20.2	13.1	0.16	0.41	0.48	22.2	33.3	44.4	0.46	0.43	0.33
ENVIRONMENTAL SCS & TECHNOLOGY	66.3	20.6	13.1	0.15	0.36	0.45	24.6	31.8	43.6	0.33	0.36	0.27
GENERAL & INDUSTRIAL ENGINEERING	63.4	22.2	14.4	0.19	0.41	0.59	22.9	37.6	39.1	0.53	0.49	0.50
HEALTH SCIENCES	65.6	20.8	13.6	0.15	0.37	0.45	24.0	32.3	43.7	0.36	0.37	0.29
INFORMATION & COMM. SCS.	63.1	23.5	13.4	0.21	0.39	0.75	22.6	42.9	34.2	0.58	0.50	0.68
INSTS. & INSTRUMENTATION	71.0	17.8	11.2	0.15	0.45	0.52	19.2	33.1	47.7	0.49	0.44	0.32
LAW & CRIMINOLOGY	62.9	22.4	14.6	0.19	0.39	0.56	23.4	37.2	39.2	0.51	0.46	0.47
MANAGEMENT & PLANNING	64.0	21.5	14.5	0.19	0.39	0.57	23.2	36.6	40.1	0.47	0.46	0.48
MATHEMATICS	67.5	20.2	12.3	0.14	0.37	0.46	18.5	34.0	47.5	0.50	0.39	0.28
MECHANICAL ENG. & AEROSPACE	65.2	21.1	13.7	0.15	0.37	0.45	22.9	33.0	44.1	0.41	0.38	0.29
PHYSICS & MATERIALS SCIENCE	71.1	19.4	9.5	0.10	0.24	0.42	23.0	31.3	45.7	0.24	0.21	0.16
PSYCHOLOGY	66.9	20.2	12.9	0.14	0.36	0.45	23.7	31.7	44.6	0.33	0.35	0.26
SOCIAL & BEHAVIORAL SCIENCES	62.4	23.5	14.1	0.21	0.37	0.73	23.7	41.6	34.6	0.54	0.48	0.66
SOCIOLOGY & ANTHROPOLOGY	62.6	23.4	14.0	0.21	0.37	0.73	23.4	41.6	34.9	0.54	0.48	0.65
STATISTICAL SCIENCES	67.3	19.9	12.8	0.15	0.40	0.48	21.1	33.2	45.7	0.44	0.42	0.32
Average over the 29 fields	66.4	20.8	12.8	0.16	0.36	0.52	23.2	34.0	42.8	0.40	0.39	0.34
Coefficient of variation	0.05	0.07	0.14	0.19	0.12	0.20	0.08	0.10	0.09	0.28	0.21	0.42

Within-field results: average and coefficient of variation of the CSS results for all individuals in each field.
Between-field results: average and coefficient of variation of the CSS results over the 29 fields.

cv_{f3}), (s_{f1}, s_{f2}, s_{f3}) and (cv_{f4}, cv_{f5}, cv_{f6}) in each field and are in columns 1–12 in Table 6. In turn, the average of p_{fj} and s_{fj} over the 29 fields for j = I, II, III are denoted by P_j and S_j, respectively, whereas the average of (cv_{f1}, cv_{f2}, cv_{f3}) and (cv_{f4}, cv_{f5}, cv_{f6}) over the 29 fields are denoted by (CV_{f1}, CV_{f2}, CV_{f3}) and (CV_{f4}, CV_{f5}, CV_{f6}). The results on (P_1, P_2, P_3), (CV_{f1}, CV_{f2}, CV_{f3}), (S_1, S_2, S_3) and (CV_{f4}, CV_{f5}, CV_{f6}), as well as their corresponding CVs over the 29 fields, are in the last two rows in Table 6. Finally, the information concerning (p_{f1}, p_{f2}, p_{f3}) for f = 1,. . ., 29 is illustrated in Fig. 1 where fields are ordered by p_{f1}.

There are three main results. Firstly, as expected, the CVs in columns 4 to 6 and 10 to 12 in Table 6 indicate that the skewness of individual citation distributions exhibit a very large within-field variability. Secondly, recall that uniform or normal distributions would yield percentages of articles in categories I, II, and II equal to 50%, 25%, and 25% in the first case, and 50%, 28.8% and 21.2% in the second one. However, on average over all fields, mean citation rates are approximately 16 points above the median, and less than 13% of articles in category III account for almost 43% of all citations. In brief, on average individual citation distributions within each field are considerably skewed. Thirdly, judging from the size of CVs over the 29 fields in columns 1 to 3 and 7 to 9, the degree of skewness across fields is very similar indeed. Fig. 1 clearly illustrates this important result.

4. Within- and between-field results concerning field citation distributions

4.1. The extended versus the double extended count for all authors

In this Section we investigate the connection between the skewness at the individual and field levels. But to do this, we must determine which type of field citation distribution we wish to select: field citation distributions in the extended count or in the double extended count. In order to facilitate the reading of the text, a detailed discussion of this issue is relegated to **Part 2. The extended vs. the double extended count** in the Supplementary Material. Fortunately, the difference between the CSS results for field citation distributions in both counts is so small that, for all practical purposes, we may continue the analysis focusing on either case. In what follows, we will restrict ourselves to the double extended count.

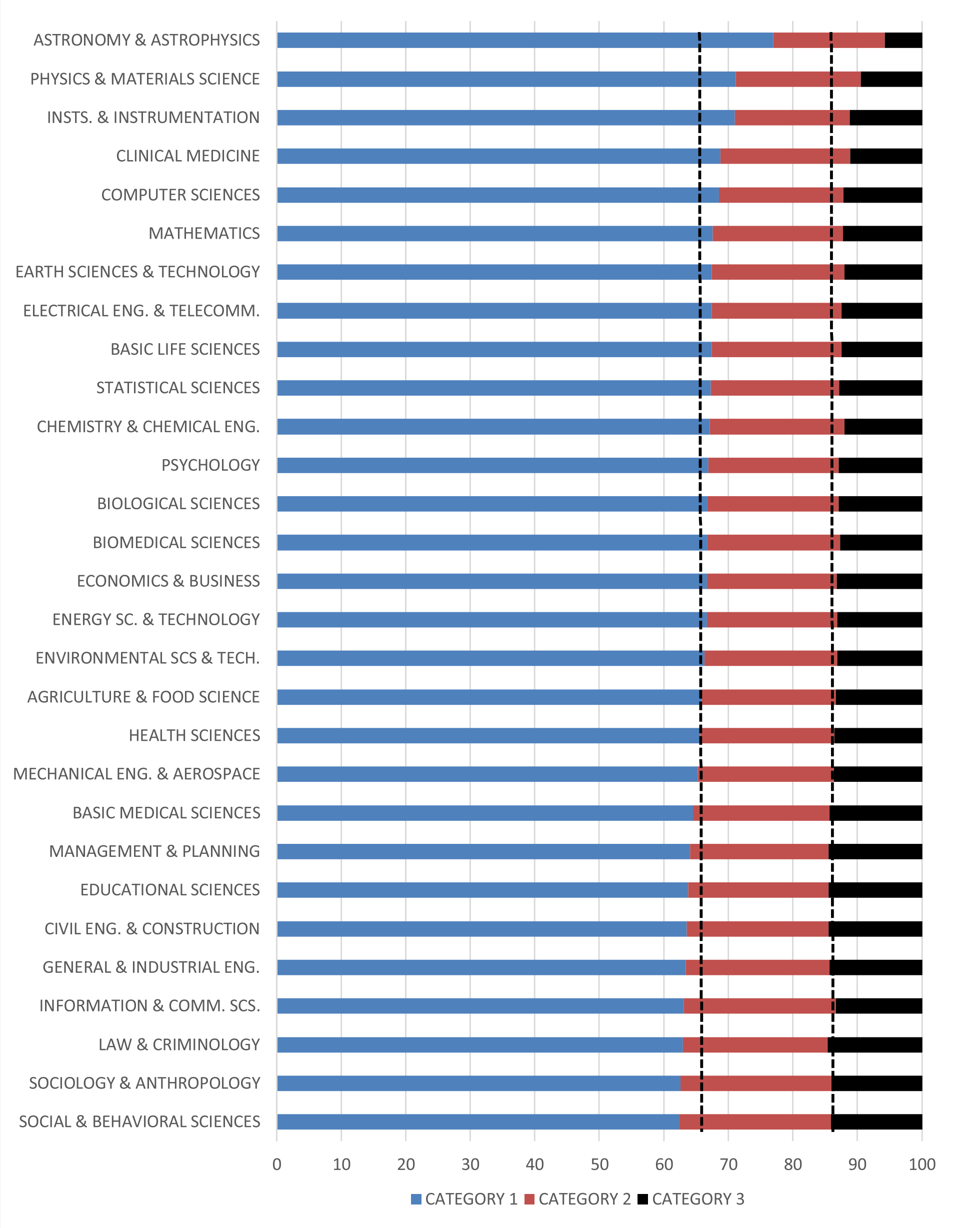

Fig. 1. The skewness of average individual citation distributions in each field according to the CSS approach. Very productive authors in the double extended count.

4.2. The gap between the skewness of the field citation distribution and the average skewness of the individual citation distributions for very productive authors

In Section 3 we considered very productive authors with citation distributions of a certain minimal size. In principle, it is natural to focus on field citation distributions consisting of articles published only by authors of this type. However, it is also important to consider the unrestricted field citation distributions of articles published by all authors. Although, as we will see, the difference is relatively small, we first study the field citation distributions consisting of articles published by very productive authors in the double extended count.

The information for the first and the second means for these distributions, denoted by $\mu^{D*}{}_{fj}$ for $j = 1, 2$, is in Table 7. It is interesting to compare the means of field citation distributions in the double extended count in Table 7 with the average of the mean citations of very productive authors (columns 3 and 5 in Table 5). For any f, the individual citation distributions

Table 7
First and second means of field citation distributions for very productive authors in the double extended count, and average and coefficient of variation of these magnitudes over the 29 fields.

FIELDS	$\mu^{D*}{}_{f1}$ (1)	CV (2)	$\mu^{D*}{}_{f2}$ (3)	CV (4)
AGRICULTURE & FOOD SCIENCE	16.0	2.46	46.3	1.47
ASTRONOMY & ASTROPHYSICS	29.2	5.37	116.1	3.01
BASIC LIFE SCIENCES	35.3	3.43	122.5	1.94
BASIC MEDICAL SCIENCES	17.0	2.58	47.8	1.61
BIOLOGICAL SCIENCES	25.3	4.48	96.6	2.47
BIOMEDICAL SCIENCES	26.1	2.79	81.8	1.63
CHEMISTRY & CHEMICAL ENG.	19.6	3.36	58.5	2.08
CIVIL ENG. & CONSTRUCTION	10.9	1.98	30.4	1.09
CLINICAL MEDICINE	27.5	3.40	89.5	2.00
COMPUTER SCIENCES	9.5	4.50	36.1	2.42
EARTH SCIENCES & TECHNOLOGY	21.6	2.51	63.6	1.48
ECONOMICS & BUSINESS	18.5	2.55	61.9	1.34
EDUCATIONAL SCIENCES	13.5	2.28	40.6	1.27
ELECTRICAL ENG. & TELECOMM.	9.8	5.55	32.2	3.32
ENERGY SC. & TECHNOLOGY	11.4	3.21	37.2	1.83
ENVIRONMENTAL SCS & TECH.	21.8	2.60	64.1	1.55
GENERAL & INDUSTRIAL ENG.	10.0	1.94	26.5	1.11
HEALTH SCIENCES	16.8	2.70	47.1	1.67
INFORMATION & COMM. SCS.	14.4	2.79	44.8	1.60
INSTS. & INSTRUMENTATION	8.3	3.59	31.9	1.85
LAW & CRIMINOLOGY	10.8	2.35	28.0	1.47
MANAGEMENT & PLANNING	23.4	2.33	73.4	1.26
MATHEMATICS	7.0	3.11	22.0	1.76
MECHANICAL ENG. & AEROSPACE	10.5	2.43	30.0	1.44
PHYSICS & MATERIALS SCIENCE	19.6	4.71	66.1	2.81
PSYCHOLOGY	22.4	2.74	67.1	1.62
SOCIAL & BEHAVIORAL SCIENCES	15.9	2.65	44.5	1.63
SOCIOLOGY & ANTHROPOLOGY	16.6	2.42	49.6	1.39
STATISTICAL SCIENCES	13.8	5.02	50.4	2.82
Average over the 29 fields	17.3	3.2	55.4	1.8
Coefficient of variation	0.41	0.32	0.47	0.32

of very productive authors form a partition of the corresponding field citation distribution. Consequently, for the first mean we have:

$$\mu^{D*}{}_{f1} = \Sigma_i \, \mathrm{w}_{*f}(\mathrm{i})\mathrm{m}_{f1}(\mathrm{i}),$$

where $w^*{}_f(i) = n_f(i)/N^{D*}{}_f$ is the proportion of the publications of author i, $n_f(i)$, with respect to the total number of publications in the double extended count, i.e. $N^{D*}{}_f = \Sigma_i \, n_f(i)$, where all summations are over the $I^*{}_f$ very productive authors in the field. Instead, the average of the mean citations of very productive authors is

$$m_{f1} = [\Sigma_i \, \mathrm{m}_{f1}(i)]/I*_f.$$

Therefore, as long as $m_{f1}(i)$ tends to increase with $n_f(i)$, we expect $\mu^{D*}{}_{f1} > m_{f1}$. This is what we find for every f when we compare column 1 in Table 7 with column 3 in Table 5. Hence, on average over the 29 fields, we have $(\Sigma_f \, \mu^{D*}{}_{f1})/29 = 17.3 > (\Sigma_f \, m_{f1})/29 = 16.4$. However, the differences are relatively small, indicating that $m_{f1}(i)$ does not increase much with $n_f(i)$, i.e. that the scaling relationship between mean citations and the number of scientific publications among very productive authors is rather weak. Similar results hold for the second means.

We denote by $(P^*{}_{f1}, P^*{}_{f2}, P^*{}_{f3})$, $(S^*{}_{f1}, S^*{}_{f2}, S^*{}_{f3})$, $f = 1, \ldots, 29$, the CSS results for field citation distributions of articles published by very productive authors in the double extended count. In turn, we denote by $(P^*{}_1, P^*{}_2, P^*{}_3)$ and $(S^*{}_1, S^*{}_2, S^*{}_3)$ the average of these quantities over the 29 fields. The CSS results are presented in Table 8.

We first note that, except for the proportion of articles in category III, the small CVs for the other five parameters over the 29 fields indicate that the skewness of field citation distributions is very similar indeed. This is clearly illustrated in Fig. 2 representing the proportion of articles in the three categories ordered by $P^*{}_{f1}$.

Finally, we arrive to the most important comparison in this Section between the average skewness of individual citation distributions for very productive authors in a given field, i.e. (p_{f1}, p_{f2}, p_{f3}) and (s_{f1}, s_{f2}, s_{f3}) in Table 6 and Fig. 1, and the skewness of the field citation distribution consisting of the articles these authors produce, i.e. $(P^*{}_{f1}, P^*{}_{f2}, P^*{}_{f3})$ and $(S^*{}_{f1}, S^*{}_{f2}, S^*{}_{f3})$ in Table 8 and Fig. 2. The key observation is that the average skewness of the individual citation distributions in each field is considerably smaller than the skewness of field citation distributions. In Agriculture and Food Science, for example, in the first case the mean is 15.8 points greater than the median and 13.3% of highly cited articles account for 43.7% of all citations, whereas in the second case the mean is 23.8 points greater than the median and only 6.9% of highly cited articles account for 43.2% of all citations.

Table 8
The skewness of field citation distributions according to the CSS approach for very productive authors in the double extended count, and average and coefficient of variation of the CSS results over the 29 fields.

	% of articles in category:			% of total citations in category:		
	I (1)	**II** (2)	**III** (3)	**I** (4)	**II** (5)	**III** (6)
ASTRONOMY & ASTROPHYSICS	81.4	15.7	2.9	25.8	27.7	46.4
BASIC LIFE SCIENCES	78.1	16.8	5.1	23.8	29.8	46.4
BASIC MEDICAL SCIENCES	73.5	19.5	7.0	25.7	32.0	42.3
BIOLOGICAL SCIENCES	79.8	16.1	4.1	23.0	29.2	47.8
BIOMEDICAL SCIENCES	75.9	18.1	6.0	24.7	30.9	44.4
CHEMISTRY & CHEMICAL ENG.	74.5	19.0	6.5	23.9	31.7	44.5
CIVIL ENG. & CONSTRUCTION	71.8	19.9	8.3	21.0	32.4	46.6
CLINICAL MEDICINE	76.6	18.0	5.4	23.8	31.0	45.1
COMPUTER SCIENCES	78.6	16.6	4.7	18.9	31.3	49.8
EARTH SCIENCES & TECHNOLOGY	74.2	19.1	6.6	23.9	31.8	44.3
ECONOMICS & BUSINESS	76.2	17.4	6.4	20.7	31.1	48.3
EDUCATIONAL SCIENCES	74.3	18.8	6.9	22.5	32.0	45.5
ELECTRICAL ENG. & TELECOMM.	75.8	18.5	5.7	20.2	32.3	47.5
ENERGY SC. & TECHNOLOGY	75.6	18.0	6.3	20.9	32.2	46.9
ENVIRONMENTAL SCS & TECH.	74.0	19.4	6.6	23.6	32.0	44.4
GENERAL & INDUSTRIAL ENG.	70.4	21.0	8.6	21.4	33.1	45.5
HEALTH SCIENCES	72.4	20.3	7.2	23.0	33.0	44.1
INFORMATION & COMM. SCS.	74.9	18.5	6.7	21.9	31.8	46.2
INSTS. & INSTRUMENTATION	78.8	15.5	5.7	18.7	30.5	50.9
LAW & CRIMINOLOGY	69.9	21.1	9.0	22.0	32.8	45.1
MANAGEMENT & PLANNING	75.2	18.0	6.9	22.1	31.3	46.7
MATHEMATICS	73.9	19.0	7.1	17.8	31.7	50.5
MECHANICAL ENG. & AEROSPACE	72.7	19.5	7.7	22.5	32.0	45.4
PHYSICS & MATERIALS SCIENCE	77.2	17.9	4.8	23.2	31.4	45.4
PSYCHOLOGY	74.4	18.9	6.7	23.2	32.2	44.6
SOCIAL & BEHAVIORAL SCIENCES	72.5	20.0	7.5	22.8	32.8	44.4
SOCIOLOGY & ANTHROPOLOGY	74.1	19.1	6.8	22.7	31.8	45.6
STATISTICAL SCIENCES	78.2	17.3	4.4	20.3	31.2	48.5
Average over the 29 fields	75.1	18.5	6.4	22.3	31.6	46.1
Coefficient of variation	0.04	0.08	0.21	0.09	0.04	0.05

As we will presently see, a possible explanation is the following. Together with the skewness of individual citation distributions, the skewness of a field citation distribution may essentially arise from two additional factors: differences between individual productivity, measured by the number of articles per author, and differences between individual mean citation rates.

The following examples in the Appendix A illustrate the situation. In the first place, differences in individual productivity in a given field may have no impact on the skewness of the field citation distribution. For example, if all individual citation distributions in a field have the same first and second mean and the same skewness, the average skewness of individual citation distributions will coincide with the skewness of the corresponding field citation distribution regardless of any difference in the size of individual citation distributions. However, when individual citation distributions have different skewness, within-field differences in individual productivity may affect the skewness of the field citation distribution. Example 1 in the Appendix A illustrates this case for two individuals in a single field with the same first mean. In the second place, when individuals are equally productive we may still have a skewness gap. Example 2 in the Appendix A considers two individuals in a single field with citation distributions not only of equal size but also equal skewness. Naturally, in this case the average skewness coincides with the skewness of the individual citation distributions. However, the difference in individual mean citation rates causes the average skewness to be smaller than the skewness of the field citation distribution. More generally, in practice it is likely that individuals will have different number of publications, different means, and different skewness –a case illustrated in Example 3 in the Appendix A.

Given the between-field results at the individual and field levels, the gap between the skewness of any field citation distribution and the average skewness of individual citation distributions in that field is of the same order of magnitude. Therefore, for simplifying purposes we will restrict ourselves to the skewness results for the average over the 29 fields, namely, (P_1, P_2, P_3) and (S_1, S_2, S_3) versus (P^*_1, P^*_2, P^*_3) and (S^*_1, S^*_2, S^*_3), which are reproduced in rows I and II in Table 9.

Our aim is the following. We want to establish that the skewness gap between rows I and II in Table 9 can be mostly explained by differences in individual productivity and individual mean citations. We begin with the first factor. If the only source of the skewness gap were differences in individual productivity, then a solution would be to consider the weighted average skewness of individual citation distributions, with weights equal to the proportion that the number of articles of each author represents with respect to the total number of articles in the field. In this case, as illustrated in Example 1, the

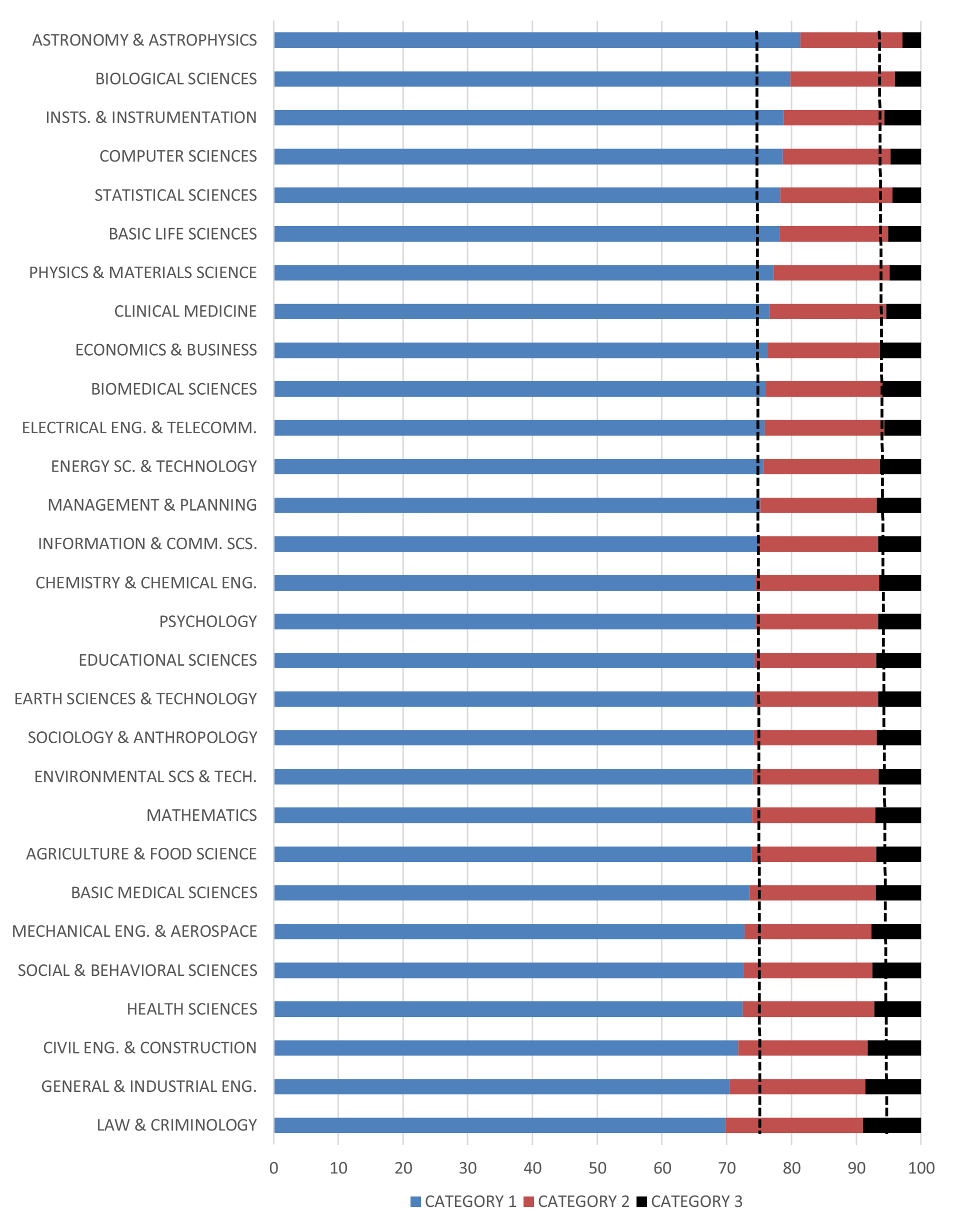

Fig. 2. The skewness of field citation distributions according to the CSS approach. Articles published by very productive authors in the double extended count.

skewness gap would disappear. Therefore, given the CSS individual results $(p_{f1}(i), p_{f2}(i), p_{f3}(i))$, $(s_{f1}(i), s_{f2}(i), s_{f3}(i))$, $i = 1, \ldots, I_f{}^*$ in every field, instead of the simple averages in Eqs. (1) and (2) in Section III.2, we would estimate, for every $j = 1, 2, 3$,

$$p'_{fj} = \Sigma_i \, w*_f(i) \, p_{fj}(i), \tag{3}$$

and

$$s'_{fj} = \Sigma_i \, w*_f(i) \, s_{fj}(i), \tag{4}$$

where, as before, $w^*_f(i) = n_f(i)/N^{D*}{}_f$, $N^{D*}{}_f = \Sigma_i \, n_f(i)$, and the sum in expressions (3) and (4) goes over the $I_f{}^*$ very productive authors. The average of $(p'_{f1}, p'_{f2}, p'_{f3})$ and $(s'_{f1}, s'_{f2}, s'_{f3})$ over all fields are denoted by (P'_1, P'_2, P'_3) and (S'_1, S'_2, S'_3). The results for $(p'_{f1}, p'_{f2}, p'_{f3})$ and $(s'_{f1}, s'_{f2}, s'_{f3})$ in all fields are in Table A4 in the Appendix A, while the results for (P'_1, P'_2, P'_3) and (S'_1, S'_2, S'_3) are reproduced in row III in Table 9.

Recall that the within-field variability of individual productivity is rather high (column 2 in Table 5). Therefore, as long as the skewness of individual citation distribution for author i increases with $n_f(i)$, we expect that the results for (p'_{fj}, s'_{fj}),

Table 9
Average and coefficient of variation of the CSS results over the 29 fields. Selected individual and field citation distributions.

	% of articles in category:			% of total citations in category:		
	I	II	III	I	II	III
A. Very productive authors in the double extended count						
I. Average individual citation distributions of very productive authors in each field						
Average over 29 fields	66.4	20.8	12.8	23.2	34.0	42.8
Coefficient of variation	0.05	0.07	0.14	0.08	0.10	0.09
II. Field citation distributions						
Average over 29 fields	75.1	18.5	6.4	22.3	31.6	46.1
Coefficient of variation	0.04	0.08	0.21	0.09	0.04	0.05
III. Weighted average individual citation distributions of very productive authors in each field						
Average over 29 fields	67.5	20.5	12.0	23.3	33.0	43.7
Coefficient of variation	0.05	0.06	0.16	0.08	0.07	0.06
IV. Field citation distributions controlling for individual productivity within each field						
Average over 29 fields	75.4	18.3	6.3	22.3	31.5	46.2
Coefficient of variation	0.03	0.08	0.20	0.08	0.04	0.04
V. Field citation distributions controlling for individual productivity and individual mean citations within each field						
Average over 29 fields	66.4	21.9	11.7	23.2	31.5	46.2
Coefficient of variation	0.05	0.05	0.17	0.08	0.02	0.04
B. All authors in the double extended count						
VI. Field citation distributions						
Average over 29 fields	76.8	17.7	5.5	21.6	31.3	47.1
Coefficient of variation	0.03	0.08	0.22	0.08	0.05	0.05

j = 1, 2, 3, reflect a greater skewness than the results for (p_{fj}, s_{fj}), j = 1, 2, 3. However, by comparing the field results in Table 6 and Table A4 in the Appendix A, we observe that the skewness of the weighted average is only slightly greater than the skewness of the simple average. This indicates that the skewness of citation distributions among very productive authors does not vary much with individual productivity. Therefore, we conclude that differences in individual productivity play a minor role in explaining the skewness gap in each field. Given the similarity across fields (see the low CVs in row III in Table 9), this is also what we observe by comparing rows I and III in Table 9.

There is another way of studying the role of differences in individual productivity. We can estimate the skewness of field citation distributions controlling for the within-field differences in individual productivity by equalizing the number of articles per author. Since the CSS technique is size-independent, the skewness of individual citation distributions is preserved. As illustrated in Example 1, if the only source of the skewness gap were differences in individual productivity, then after this normalization the skewness gap would disappear. In our case, we proceed by weighting every article of an individual i in field f by the quantity $[n_*/n_f(i)]$, where n_* is an arbitrary amount. In this way, individual productivity in each field becomes equal to n_*. The CSS results for the average over the 29 fields appear in row IV in Table 9 (detailed field results are available on request). Given the small role that differences in individual productivity have in explaining the skewness gap in each field, we expect minor differences in the skewness in each field. This is exactly what we find when we compare rows II and IV in Table 9.

Next, we must study the role of within-field differences in mean citations in generating the skewness gap. For that purpose, we can estimate the skewness of field citation distributions controlling for these differences by equalizing the first mean of all authors in a given field. Since the CSS technique is scale-independent, the skewness of individual citation distributions is preserved. Instead, as a consequence of the equalization of individual mean citations, the skewness of the new field distribution should be reduced. The size of the reduction in skewness will inform us of the role of differences in mean citations in explaining the skewness gap. As Example 2 illustrates, when the main difference between authors is the difference in the first mean citation, this procedure completely eliminates the skewness gap. However, when authors also differ in their second mean citation by a sufficient amount –as it is the case in Example 3– the gap does not completely vanish.

For our dataset, the task is to explain the skewness gap at the field level between Tables 8 and A4 in the Appendix A, and between rows I and IV in the aggregate, *once* we have controlled the skewness of field citation distributions by differences in individual productivity. Thus, after weighting every article of an individual i in field f by the quantity $[n_*/n_f(i)]$, we now multiply the citation count $c_{fk}(i)$ for all k by the quantity $[\mu_*/\mu_{f1}(i)]$, where μ_* is an arbitrary amount. In this way, individual mean citations in each field become equal to μ_*.[7] Note that the total number of articles in each field will be the product of $I_f{}^*$and n_*, the field mean citation will be equal to μ_*, and the percentage of articles in category I, as well as the percentage of total citations accounted by articles in this category, will coincide with the average of the corresponding individual percentages. The results for all fields are in Table B1 in the Appendix A, while the results for the average over the 29 fields

[7] This normalization can only be applied for authors with a positive mean citation. However, very productive authors receiving no citations only represent 0.022% of the total (details by field are available on request).

are reproduced in row V in Table 9. The resulting skewness after the double normalization is called the *basic skewness* of field citation distributions.

Note that the within-field variability of the first mean among very productive authors is very large (column 4 in Table 5). Consequently, by comparing Tables A4 and B1 in the Appendix A we observe that, as a consequence of the second normalization the skewness of field citation distributions is greatly reduced. Given the similarity across fields (see the low CVs in row V in Table 9), this is also the case when comparing rows II and IV in Table 9. The minimal resulting gap between the basic skewness in any field (row V) and the weighted or unweighted average skewness at the individual level (rows II and I) is due to differences in the second mean of the individual normalized citation distributions that cannot be eliminated without changing the original individual skewness.

The conclusion is that, in every field, the skewness of the field citation distribution can be essentially accounted for by the average skewness of individual citation distributions and the within-field differences in the size and the mean of the latter. Differences in individual mean citations are much more important that differences in individual productivity in explaining the initial skewness gap. Furthermore, the relative order of magnitude of these two sources of skewness is the same for all fields.

4.3. The gap between the skewness of the field citation distribution and the average skewness of the individual citation distributions for merely productive authors

Very productive authors have been defined using a relative benchmark that takes into account field differences in production practices. An alternative is to define *merely productive* authors as those who publish at least five articles in the 2000–2015 period. The percentages of merely productive authors in each field are in column 4 in Table 4. Except for three fields where very productive authors publish only four or more articles –Information & Communication Sciences, Social & Behavioral Sciences, and Sociology & Anthropology–, the percentage of merely productive authors is greater than the percentage of very productive authors. Consequently, in 26 fields merely productive authors are responsible for greater percentages of articles than very productive researchers. In Agriculture and Food Science, for example, merely productive authors represent 9.1% of all authors and are responsible for 48.5% of all articles in the field. On average over all fields, merely productive authors represent 9.4% of all authors and are responsible for 47.9% of all articles (for details, see column 6 in Table 4).

In order to facilitate the reading of the text, the CSS analysis of individual citation distributions for merely productive authors is relegated to **Part 3. Merely productive authors** in the Supplementary Material. Interestingly, the results are very similar to those obtained for very productive authors. Three points should be noted. Firstly, individual citation distributions are slightly less skewed for merely productive authors than for very productive ones. Secondly, field citation distributions consisting of the articles published by merely productive authors turned out to be essentially as skewed as field citation distributions for very productive authors in Table 8. Thirdly, as a consequence of these two facts, the gap between the skewness of field citation distributions and the average skewness of the individual citation distributions for merely productive authors in each field is slightly greater than for very productive ones. However, controlling for differences in individual productivity and individual mean citations within each field, the skewness of field citation distributions for merely productive authors is essentially the same as for very productive authors.

In brief, the only difference between the two cases is that, since there are more authors involved, the within-field variation of individual productivity and individual mean citations is greater for merely productive authors than for very productive ones. Consequently, as we have seen, the gap between the skewness of field citation distributions and the average skewness of the individual citation distributions for merely productive authors in each field is slightly greater than for very productive ones. However, controlling for such differences, we arrive to a very similar basic skewness in each field. The consequence of this result is very helpful: for all practical purposes, our analysis can be equally conducted in terms of the two notions of productive authors. Generally, we will restrict our attention to very productive authors.

4.4. The gap between the skewness of field citation distributions for all authors and the average skewness of the individual citation distributions for very productive authors in each field

As indicated before, it is also important to consider the unrestricted field citation distributions of articles published by all authors. Recall that the CSS results on the skewness of field citation distributions in this case are in Table SM7 in Part 2 in the Supplementary Material. CSS results for the average over the 29 fields are reproduced in row VI in Table 9.

Our task is to explain the skewness gap between rows I and VI in Table 9, which is slightly greater than the gap between rows I and II for very productive authors. In line with our previous argument, the explanation is that the within-field variation of individual productivity and individual mean citations is greater for all authors than for very productive authors. For individual productivity, this is exactly what we find when we compare column 2 in Table 3 and column 2 in Table 5. For individual mean citations, this is also what we find when we compare column 2 in Table SM1 in Part 1 in the Supplementary Material and column 4 in Table 5.

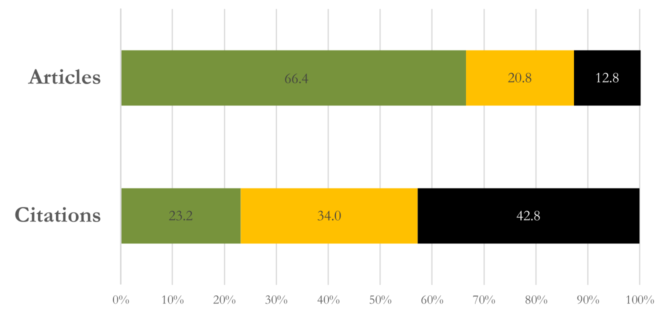

Fig. 3. Percentage of articles in categories I, II, III, and percentage of citations accounted for by each category. Individual citation distributions of very productive authors in each field. Average of CSS results over the 29 fields.

5. Discussion

It will be useful to organize the discussion of the results around the following three issues: patterns of individual citation distributions, patterns of field citation distributions, and the relationship between the two.

5.1. Patterns of individual citation distributions

Within each field, individual scientists are extremely heterogeneous. In our dataset, we observe a well-known large within-field variation in the following three dimensions: individual productivity, measured by the number of articles; individual citation impact, measured by mean citation rates, and the pattern of co-authorship, measured by the number of authors per article.[8] This large variability in individual productivity and citation impact is also present for very productive and merely productive authors).

In addition, we have investigated the skewness characteristics of individual citation distributions for very productive and merely productive authors that represent, on average, 5.2% and 9.4% of the total number of authors. Focusing on the former, two results stand out. Firstly, not surprisingly, we also find a large within-field variability for the CSS results of individual citation distributions. Secondly, the average of the individual CSS results in each field exhibit a clear skewness pattern. Furthermore, judging from the small size of CVs over the 29 fields, in spite of wide differences in production and citation practices across fields this skewness pattern is very similar for all of them. Note that this is at variance with the results in Costas et al. (2009), where field characteristics influence the research performance of individual authors in the sense that the size-dependent cumulative advantage for receiving citations tends to be larger in low citation-density fields. For later reference, the average results over the 29 fields for very productive authors are illustrated in Fig. 3.

It is important to emphasize that the CSS results concerning the within- and between-field variation just summarized, are only slightly less pronounced for merely productive authors (Table SM9 in Part 3 in the Supplementary Material). Essentially, this indicates that, within each field, the average CSS results for individual citation distributions conditional on the number of articles per person do not change much as we increase the authors' individual productivity.

The similarity of the average characteristics of individual citation distributions across fields has important conceptual and practical consequences: to explain the skewness of individual citation distributions we do not need a different model for each field. On the contrary, since a certain degree of average skewness seems to be generic, all we need is a single model for individual researchers in any scientific field. A good example can be found in Sinatra, Wang, Dweville, Song, and Barabási (2016), where an author's high-impact work, resulting from a combination of her ability to take advantage of the available knowledge and a random element, is randomly distributed within her career.

5.2. Patterns of field citation distributions

Previous results for large WoS datasets for classification systems at different aggregation or granularity levels with a fixed five-year citation window indicate that field citation distributions are highly skewed, and that between-field variability is very reduced. For example, as documented in Li et al. (2013), CSS results evolve smoothly during the 1980–2004 period. As the citation window increases from seven years for documents published in 2004 up to 31 years for documents published in 1980, sub-field citation distributions become somewhat more skewed (the increase in the degree of skewness with the length of the citation window is amply documented in Katz, 2016). The evidence for more than two decades is summarized

[8] The within-field variability in individual productivity, individual citation impact, and the pattern of co-authorship are also characteristics of RCC's dataset.

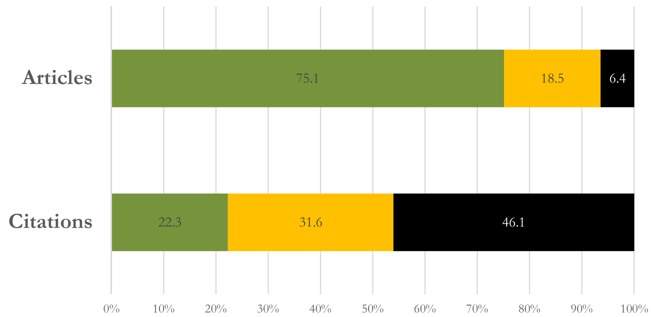

Fig. 4. Percentage of articles in categories I, II, III, and percentage of citations accounted for by each category. Field citation distributions in the double extended count for very productive authors. Average of CSS results over the 29 fields.

in Li et al. (2013) by the following percentages of publications and total citations in categories I, II, and III: (70.9, 20.4, 8.7) and (22.7, 32.7, 44.6).[9]

In the extended count in our dataset, these percentages are (74.9, 18.6, 6.5) and (21.5, 32.0, 46.5), whereas in the double extended count the averages over the 29 fields are (76.8, 17.7, 5.5) and (21.6, 31.3, 47.1). As we have seen in Part 2 in the Supplementary Material, the similarity between the extended and the double extended counts indicate that, essentially, the skewness characteristics of citation distributions of articles conditional on the number of authors do not change much as we vary the number of authors per publication. In any case, the skewness of field citation distributions in our dataset with a variable 16-year citation window is somewhat more pronounced, but still of a comparable order of magnitude, than the skewness documented in the previous contributions referred to.[10]

Publication and citation practices are very different across scientific disciplines at all aggregation levels. As a result, certain key statistics –such as the number of authors per paper, the first and second means of the number of publications per author and the mean citation rates, as well as the mean number of references or a variety of indicators of citation impact amply documented in the literature– exhibit a large range of variation across scientific fields. However, the reduced between-field variability of the CSS results presented in this paper and previous contributions indicate that the degree of skewness of field citation distributions is very similar indeed.

Three comments are in order. Firstly, as emphasized in Albarrán et al. (2011) and Waltman, Van Eck, and Van Raan (2012), this similarity should not be confused with the universality claimed by Radicchi et al. (2008). Secondly, nevertheless, the similarity between field citation distributions opens the possibility of meaningful comparisons of citation counts across fields (Crespo, Li, & Ruiz-Castillo, 2013; Crespo, Herranz, Li, & Ruiz-Castillo, 2014; Glänzel, 2011; Li et al., 2013; Radicchi et al., 2008; Radicchi & Castellano, 2012; Ruiz-Castillo, 2014). Thirdly, the similarity of the degree of skewness at the field level is at variance with the results in van Raan (2006a, 2006b, 2008) concerning the scaling relationship between the number of citations and the number of scientific publications: in these contributions the size-dependent cumulative advantage for receiving citations tends to be larger in low citation-density fields (although this difference in the advantage between low and high field-citation-density for research groups is larger than the difference for individuals found in Costas et al., 2009).

5.3. Relationships between individual and field citation distributions

It is useful to begin investigating the macro-micro relationship for very productive authors. The CSS results for the field citation distributions in the double extended count are in Table 8. The average results over the 29 fields, (75.1, 18.5, 6.4) and (22.3, 31.6, 46.1), are illustrated in Fig. 4.

The comparison between Figs. 1 and 2 illustrates the extent of the skewness gap between the average of individual citation distributions and the citation distribution consisting of the articles published by very productive authors in each field, whereas the comparison of Figs. 3 and 4 illustrates the skewness gap for the average over the 29 fields. However, a key result of this paper is that the skewness of field citation distributions can be explained in terms of the average skewness of individual citation distributions combined with the skewness in individual productivity and citation impact. Although the skewness gap is somewhat greater for merely productive authors, the explanation of the skewness of field citation distributions in terms of the average skewness of individual citation distributions combined with the differences in individual productivity and citation impact is maintained for merely productive authors.

[9] The situation closely resembles the one described in Albarrán et al. (2011) for 3.7 million articles with a common, five-year citation window published in 1998-2003 in a wide array of 219 WoS sub-fields. Similar results are also obtained for selected publication-level, algorithmically constructed classification systems consisting of 3.7 million articles classified into 2,272 and 4,161 significant clusters with at least 100 publications in Ruiz-Castillo and Waltman (2015). For a rationalization of these regularities in terms of a lognormal distribution, see Vîiu (2018).

[10] As a matter of fact, our results are closer to those reported in Glänzel (2007) for 450,000 papers published in 1980 with a 20-year citation window and classified into 12 major fields and 60 subfields according to the publication-level Leuven/Budapest classification system (Glänzel & Schubert, 2003). The proportion of the 450,000 publications in categories I, II, and III are (74.7, 18.5, 6.7).

We have established that differences in individual productivity are a very minor source of skewness at the field level, so that most of the skewness gap is accounted for by differences in individual mean citations. What we cannot do in this framework is to measure the relative contribution to the total of the two main sources of skewness. That is, we cannot answer which part of the skewness of a field citation distribution can be attributed to the average skewness of individual citation distributions, and which part can be attributed to differences in individual mean citations. The reason, of course, is that the CSS technique is not decomposable by population subgroup. As a matter of fact, all real valued measures of skewness involve highly non-linear transformations of the data. Consequently, we do not know of any skewness index which is decomposable by population subgroup.

An alternative, of course, is to study the relationship between the citation inequality at the field and the individual level. There are size- and scale-independent citation inequality indices which are decomposable by population subgroup in the sense that for any partition of the population, for example the partition of a field into productive authors, the citation inequality at the field level can be expressed as the sum of a within-group and a between-group term. The within-group term is the weighted average of the citation inequality of individual authors, with weights equal to the proportion that the number of articles of any individual represent with respect to the total number of articles in the field. The between-group term is equal to the citation inequality of a field distribution in which the number of citations of any article is replaced by the mean citation of the author to which the article belongs.[11] In that case, it is possible to measure the relative contribution to the total of the within- and the between-group terms.

Coming back to the CSS approach, Thijs et al. (2017) indicate that the average CSS results for individual citation distributions in each field, (p_{f1}, p_{f2}, p_{f3}) and (s_{f1}, s_{f2}, s_{f3}), constitute a natural benchmark for the assessment of the CSS results $(p_{f1}(i), p_{f2}(i), p_{f3}(i))$ and $(s_{f1}(i), s_{f2}(i), s_{f3}(i))$ for any individual author i in that field. But the proximity that we have established between this average and the basic skewness of each field citation distribution, controlling for differences in individual productivity and individual mean citations, reinforces this choice of a benchmark.

6. Conclusions and further research

Using a large dataset consisting of 35.1 million authors and 105.3 million articles published in the period 2000–2016, we have investigated the possibility of explaining the skewness of field citation distributions in 29 broad fields in terms of the characteristics of the individual citation distributions of the authors in each field. Our main findings can be summarized in the following four points.

1 The vast majority of authors in all fields publish only less than five articles in the period 2000-2016. To study individual citation distributions of a certain size, recognizing the fact that individual productivity differs greatly between fields, it suffices to focus on very productive authors that, on average, represent only 5.2% of the population but are responsible for 38.0% of all articles. The following results for this set of very productive authors are robust to an alternative notion of merely productive authors with a minimum of five articles *per capita*.
2 Very productive authors in a given field are very different from each other in many respects. Nevertheless, on average within each field, individual citation distributions exhibit a characteristic skewness pattern. Furthermore, the degree of average skewness is very similar across all fields.
3 As in previous contributions, we find that field citation distributions are highly skewed, and the degree of skewness is very similar across all fields. However, the typical pattern at the field level can be explained in terms of three ingredients: the average skewness of the field very productive authors' citation distributions, and the heterogeneity of individuals with respect to the number of publications per author and their mean citation rates.
4 The role of these ingredients is similar in all fields. Thus, the basic skewness of field citation distributions controlling for individual differences in productivity and mean citation rates is of the same order of magnitude in all fields. We emphasize that such basic skewness is already present at the individual level. Therefore, to understand the skewness of science at the field level we must simply explain the skewness of the individual citation distributions of very productive authors.

This view of science does not crucially depend on the way we have built our dataset. It is true that the disambiguation algorithm we have used admits further improvements. It is also true that it will be very instructive to experiment with other datasets and classification systems. However, previous results indicate that our conclusions are robust to the way we have solved the assignment of articles to two or more WoS subject categories, as well as to the assignment of responsibility in the case of multiple authorship.

As far as further research, we will mention two directions. Firstly, it might be useful to treat the skewness of citation distributions by means of a real valued skewness index that is robust to extreme observations. Although it is not decomposable by population subgroup, the Groeneveld and Meeden (1984) index, which we have used in other contributions (RCC, Albarrán, Perianes-Rodriguez, & Ruiz-Castillo, 2015; Perianes-Rodríguez and Ruiz-Castillo, 2015b), might constitute an appropriate choice. Secondly, as long as one is interested in investigating whether research questions receive similar or

[11] The Generalized Entropy (GE hereafter) family of inequality indices are the only measures of relative inequality that satisfy the usual properties required from any inequality index and, in addition, are decomposable by population subgroup (Bourguignon, 1979 Shorrocks, 1980, 1984).

different answers across fields –as is the case in the present paper– it is advisable to treat distinct authors publishing in two or more fields as different authors. However, there are other contexts where this methodology is not adequate. For example, in many instances one may be interested in the ranking of universities or other research units which include distinct authors working in different fields. It would be interesting to extend our approach to this case.

Author contributions

Javier Ruiz-Castillo: Concieved and designed the analysis; Performed the analysis; Wrote the paper.
Rodrigo Costas: Collected the data; Contributed the data or analysis tool; Performed the analysis.

Acknowledgements

J. Ruiz-Castillo acknowledges financial support from the Spanish MEC through grants ECO2014-55953-P and MDM2014-0431, as well as grant MadEco-CM (S2015/HUM-3444) from the Comunidad Autónoma de Madrid. Research assistantship by Patricia Llopis, as well as conversations with Ricardo Mora, and especially Vincent Traag, are gratefully acknowledged. Three referee reports have contributed very much to the improvement of the original version of the paper. All remaining shortcomings are the authors' sole responsibility.

Appendix A.

In the following three examples, it suffices to study the first part of the CSS results for any citation distribution, namely, the percentages of articles (p_1, p_2, p_3) in categories I, II, and III.

EXAMPLE 1

1. *Individual citation distributions*

Consider two authors with the following citation distributions:

Author 1: $c(1) = (0, 0, 5, 6, 6, 8, 10)$

Author 2: $c(2) = (0, 2, 2, 2, 3, 3, 4, 6, 6, 6, 8, 8, 10, 10)$

The number of articles, and the two means for both authors and the field citation distribution are under *Raw data* in Table A1.

Table A1
Number of articles, first and second mean for authors and the field citation distribution. Raw and productivity normalized data.

	Author 1	Author 2	Field	
Raw data				
$\boldsymbol{n(i)}$	7	14	$\boldsymbol{N}$	21
$\boldsymbol{m_1(i)}$	5	5	$\boldsymbol{M_1}$	5
$\boldsymbol{m_2(i)}$	7.5	7.714	$\boldsymbol{M_2}$	7.636
Productivity normalized data				
$\boldsymbol{n'(i)}$	14	14	$\boldsymbol{N}$	28
$\boldsymbol{m'_1(i)}$	5	5	$\boldsymbol{M'_1}$	5
$\boldsymbol{m'_2(i)}$	7.5	7.714	$\boldsymbol{M'_2}$	7.6

The CSS results on the skewness of the authors' distributions are:

$$p(1) = (p_1(1), p_2(1), p_3(1)) = (3/7, 2/7, 2/7) = (42.8, 28.6, 28.6),$$

$$p(2) = (p_1(2), p_2(2), p_3(2)) = (7/14, 3/14, 4/14) = (50.0, 21.4, 28.6).$$

Therefore, the average skewness is:

$$p = (p_1, p_2, p_3) = (46.4, 25.0, 28.6),$$

where $p_j = (\Sigma_i\, p_j(i))/2$, $j = 1, 2, 3$.

2. *Field citation distribution*

As indicated under *Raw data* in Table A2, the number of articles, as well as the first and the second means of the field citation distribution are:

$$N = 7 + 14 = 21, M_1 = 5, M_2 = 7.636.$$

Therefore, the skewness of the field citation distribution is:

$$P = (P_1, P_2, P_3) = (10/21, 5/21, 6/21) = (47.6, 23.8, 28.6). P = (P_1, P_2, P_3) = (74.6, 17.5, 7.9).$$

The difference in skewness between p and P is due to differences in their individual productivity.

Table A2
Number of articles, first and second mean for authors and the field citation distribution. Raw and citation impact normalized data.

	Author 1	Author 2	Field	
Raw data				
$n(i)$	7	7	N	14
$m_1(i)$	5	11	M_1	8
$m_2(i)$	7.5	16.75	M_2	12.125
Citation impact normalized data				
$m'_1(i)$	8	8	M'_1	8
$m'_2(i)$	12	12.18	M'_2	12.09

3. *Correcting for differences in individual productivity*

There are two ways of correcting for differences in individual productivity. In the first place, we can consider the weighted average skewness where, for every j,

$$p'_j = \Sigma_i w(i) p_j(i),$$

where $w(i) = n(i)/N$. In this example, $w(1) = 7/21 = 1/3$, and $w(2) = 14/21 = 2/3$. Therefore, the weighted average skewness will be:

$$p' = (p'_1, p'_2, p'_3) = (1/3)(42.8, 28.6, 28.6) + (2/3)(50.0, 21.4, 28.6) = (47.6, 23.8, 28.6),$$

which coincides with the skewness of the field citation distribution. Therefore, in this case the skewness gap is entirely explained by the difference in individual productivity.

In the second place, we will correct for productivity differences replicating the first individual citation distribution so that it has 14 articles, that is:

$$c''(1) = (0, 0, 0, 0, 5, 5, 6, 6, 6, 6, 8, 8, 10, 10).$$

The mean citations for both authors and the field citation distribution are presented under *Productivity normalized data* in Table A2.

Since the CSS technique is replication invariant, the values

$$p''(1) = (p''_1(1), p''_2(1), p''_3(1)) = (42.8, 28.6, 28.6),$$

and

$$p'' = (p''_1, p''_2, p''_3) = (46.4, 25.0, 28.6),$$

remain invariant. However, at the field level we have:

$$P'' = (P''_1, P''_2, P''_3) = (13/28, 7/28, 8/28) = (46.4, 25.0, 28.6).$$

Therefore, the skewness of the field citation distribution controlling for differences in individual productivity, i. e. the *basic field skewness* coincides with the simple average skewness. As before, we conclude that the skewness gap is entirely explained by the difference in individual productivity.

EXAMPLE 2

1. *List Individual citation distributions*

Consider two authors with the following citation distributions:

Author 1: $c(1) = (0, 0, 5, 6, 6, 8, 10) p(1) = (p_1(1), p_2(1), p_3(1)) = (3/7, 2/7, 2/7) = (42.8, 28.6, 28.6)$.

Author 2: $c(2) = (0, 3, 7, 12, 13, 18, 24)$

The number of articles, and the two means for both authors and the field citation distribution are under *Raw data* in Table A2.

For both authors, we have

$$p(i) = (p_1(i), p_2(i), p_3(i)) = (3/7, 2/7, 2/7) = (42.8, 28.6, 28.6), i = 1, 2.$$

Therefore, the average skewness is also

$$p = (p_1, p_2, p_3) = (42.8, 28.6, 28.6),$$

where $p_j = (\Sigma_i p_j(i))/2, j = 1, 2, 3$.

2. *Field citation distribution*

As indicated under *Raw data* in Table A2, the number of articles, as well as the first and the second means of the field citation distribution are:

$N = 7 + 7 = 14$, $M_1 = (5 + 11)/2 = 8$, $M_2 = (7.5 + 16.75)/2 = 12.125$.

Therefore, the skewness of the field citation distribution is:

$P = (P_1, P_2, P_3) = (9/14, 2/14, 3/14) = (64.3, 14.3, 21.4)$. $P = (P_1, P_2, P_3) = (74.6, 17.5, 7.9)$.

3. *Correcting for differences in individual mean citations*

Why is there such a difference in skewness between p and P? Because individuals differ in their citation impact. Therefore, to establish the basic skewness of the field citation distribution we must control for individual differences in the first mean citation.

Consider multiplying every citation count $c_k(i)$ for article $k = 1, \ldots, 7$ published by author $i = 1, 2$ by the quantity $\mu*/m_1(i)$, where $\mu* = 8$. The normalized citation distributions become:

Author 1: $c'(1) = (0, 0, 8, 9.6, 9.6, 12.8, 16)$$p(1) = (p_1(1), p_2(1), p_3(1)) = (3/7, 2/7, 2/7) = (42.8, 28.6, 28.6)$.

Author 2:$c'(2) = (0, 2.18, 5.09, 8.73, 9.45, 13.09, 17.45)$

The mean citations for both authors and the field citation distribution are presented under *Citation impact normalized data* in Table A2. Since the CSS technique is scale independent, the skewness of the individual citation distributions does not vary. However, when all individual citation distributions have the same size, the same first mean citation, and the same skewness in spite of a very small difference in the second mean citation, it can be verified that the basic skewness of the field citation distribution, $P' = (P'_1, P'_2, P'_3)$, becomes equal to the average skewness of the individual distributions:

$$P' = (P'_1, P'_2, P'_3) = (6/14, 4/14, 4/14) = (42.8, 28.6, 28.6).$$

Therefore, in this case we conclude that the skewness gap is entirely explained by the difference between the individuals' first mean citations.

EXAMPLE 3

1. *Individual citation distributions*

Consider three authors with the following citation distributions:

Author 1:$c(1) = (0, 0, 5, 6, 6, 8, 10)$,

Author 2:$c(2) = (0, 0, 0, 0, 3, 3, 4, 4, 6, 6, 6, 10, 14, 21)$,

Author 3:$c(3) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5, 5, 5, 5, 5, 5, 5, 5, 7, 7, 7, 7, 15, 15, 15, 15, 15, 15, 15, 15, 36, 36, 38, 38)$.

The number of articles, and the two means for the three authors and the field citation distribution are under *Raw data* in Table A3.

Table A3
Number of articles, first and second mean for authors and the field citation distribution. Raw data, citation impact normalized data, and both citation impact and productivity normalized data.

	Author 1	Author 2	Author 3	Field	
Raw data					
$n(i)$	7	14	42	N	63
$m_1(i)$	5	5.5	8	M_1	7.11
$m_2(i)$	7.5	10.5	22.33	M_2	19.47
First mean citation normalized data					
$n'(i)$	7	14	42	N	63
$m'_1(i)$	8	8	8	M'_1	8
$m'_2(i)$	12	15.27	22.33	M'_2	18.52
First mean citation and productivity normalized data					
$n''(i)$	14	14	14	N''	42
$m''_1(i)$	8	8	8	M''_1	8
$m''_2(i)$	12	15.27	22.33	M''_2	15.39

The CSS results on the skewness of the authors' distributions are:

$$p(1) = (p_1(1), p_2(1), p_3(1)) = (3/7, 2/7, 2/7) = (42.8, 28.6, 28.6),$$

$$p(2) = (p_1(2), p_2(2), p_3(2)) = (8/14, 4/14, 2/14) = (57.1, 28.6, 14.3),$$

$$p(3) = (p_1(3), p_2(3), p_3(3)) = (30/42, 8/42, 4/42) = (71.4, 19.1, 9.5).$$

Thus, productivity, mean citation, and skewness increase as we go from author 1 to author 3. On the other hand, the average skewness is:

$$p = (p_1, p_2, p_3) = (57.1, 25.4, 17.5),$$

where $p_j = (\Sigma_i\, p_j(i))/3, j = 1, 2, 3$.

2. *Field citation distribution*

As indicated under *Raw data* in Table A3, the number of articles, as well as the first and the second means of the field citation distribution are:

$$N = 7 + 14 + 42 = 63, M_1 = 7.11, M_2 = 18.05.$$

Therefore, the skewness of the field citation distribution is:

$$P = (P_1, P_2, P_3) = (46/63,\ 12/63,\ 5/63) = (73.02,\ 19.05,\ 7.93). P = (P_1, P_2, P_3) = (74.6,\ 17.5,\ 7.9).$$

Why is there such a difference in skewness between p and P? Because individuals differ in their productivity (number of articles per author) and their citation impact (first and second mean citations per author). Therefore, to establish the basic skewness of the field citation distribution preserving the skewness of the individual citation distributions, we must control for individual differences in productivity and in the first mean citation.

3. *Correcting for differences in individual mean citations*

We will correct for mean citation differences giving all authors the same first mean citation, say 8 citations per article. Thus, citations received by authors 1 and 2 must be multiplied by the following factors: 8/5 = 1.6 and 8/5.5 = 1.45, respectively. Thus,

$$c'(1) = (0,\ 0,\ 8,\ 9.6,\ 9.6,\ 12.8,\ 16)$$

$$c'(2) = (0,\ 0,\ 0,\ 0,\ 4.36,\ 4.36,\ 5.8,\ 5.8,\ 8.7,\ 8.7,\ 8.7,\ 14.5,\ 20.4,\ 30.54).$$

The number of articles, and the two means for the three authors and the field citation distribution are under *First mean citation normalized data* in Table A3.

Since the CSS technique is scale-invariant, the values

$$p'(i) = (p'_1(i), p'_2(i), p'_3(i)) = (p_1(i), p_2(i), p_3(i)), i = 1, 2,$$

do not change. Therefore, the average skewness remains constant too. However, as indicated under *First mean citation normalized data* in Table A3, in the new field citation distribution we have:

$$N' = \Sigma_i \mathrm{n}'(i) = 63, \mu_1 = 8, \mu_2 = 19.602,$$

so that

$$P' = (P'_1, P'_2, P'_3) = (41/63,\ 16/63,\ 6/63) = (65.08,\ 25.4,\ 9.52).$$

As expected, the skewness of this new distribution is smaller than before, and the difference between P and P' is due to individual differences in the first mean citation.

4. *Correcting for differences in individual productivity*

We will correct for productivity differences modifying individual citation distributions so that all have 63/3 = 21 articles. Thus, citations received by the three authors are now weighted by a factor $21/n(i)$, i = 1, 2, 3, so that they all have 21 articles:

$$c''(1) = (0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 8,\ 8,\ 8,\ 9.6,\ 9.6,\ 9.6,\ 9.6,\ 9.6,\ 9.6,\ 12.8,\ 12.8,\ 12.8,\ 16,\ 16,\ 16),$$

$$c''(2) = (0 \text{ a total of 6 times},\ 4.36 \text{ a total of 3 times},\ 5.8 \text{ a total of 3 times},\ 8.7 \text{ a total of 4.5 times, and } 14.5,\ 20.3,\ 30.4 \text{ a total of 1.5 times each}),$$

$c''(3)$ = (0, 0, 0, 0, 0, 0, 0, 0, 0, 5, 5, 5, 5, 7, 7, 15, 15, 15, 15, 36, 38).

The number of articles, and the two means for the three authors and the field citation distribution are under *First mean citation and productivity normalized data* in Table A3.

Since the CSS technique is replication invariant, the values

$$p''(i) = (p''_1(i), p''_2(i), p''_3(i)) = (p_1(i), p_2(i), p_3(i)), i = 1, 2, 3,$$

remain invariant for all authors. Therefore, the average skewness is the same as before. However, at the field level we now have:

$$P'' = (36/63,\ 22/63,\ 5/63) = (57.1,\ 30.2,\ 12.7). P'' = (P''_1, P''_2, P''_3) = (57.1,\ 34.9,\ 8.0).$$

The difference between P' and P'' is due to individual differences in productivity. On the other hand, the skewness of the field citation distribution controlling for differences in the first mean citation and productivity, i. e. the *basic field skewness* in P'' = (57.1, 30.2, 12.7) is better approximated by the average skewness p = (57.1, 25.4, 17.5). Nevertheless, the difference between P'' and p is due to the remaining differences in the authors' second means that cannot be eliminated without changing the original individual skewness.

Table A4
The weighted average skewness of individual citation distributions for very productive authors in the double extended count, and average and coefficient of variation of the CSS results over the 29 fields.

	% of articles in category:			% of citations in category:		
	I (1)	II (2)	III (3)	I (4)	II (5)	III (6)
AGRICULTURE & FOOD SCIENCE	66.6	20.9	12.5	24.2	32.0	43.7
ASTRONOMY & ASTROPHYSICS	78.1	16.9	5.0	25.8	28.4	45.8
BASIC LIFE SCIENCES	68.5	20.0	11.5	25.2	31.0	43.8
BASIC MEDICAL SCIENCES	65.4	21.2	13.4	25.9	32.2	41.9
BIOLOGICAL SCIENCES	67.6	20.3	12.1	24.1	31.8	44.1
BIOMEDICAL SCIENCES	67.7	20.5	11.8	25.5	31.3	43.2
CHEMISTRY & CHEMICAL ENG.	68.0	20.8	11.2	25.1	31.7	43.1
CIVIL ENG. & CONSTRUCTION	64.8	21.5	13.7	22.8	34.8	42.4
CLINICAL MEDICINE	70.0	19.9	10.2	24.9	31.1	44.0
COMPUTER SCIENCES	69.4	19.3	11.4	19.1	33.5	47.5
EARTH SCIENCES & TECHNOLOGY	68.2	20.4	11.3	24.8	31.4	43.8
ECONOMICS & BUSINESS	67.5	20.0	12.5	22.5	32.2	45.4
EDUCATIONAL SCIENCES	65.0	21.2	13.8	23.2	34.9	41.9
ELECTRICAL ENG. & TELECOMM.	68.5	20.0	11.5	21.0	32.7	46.3
ENERGY SC. & TECHNOLOGY	67.5	20.2	12.3	22.3	32.7	45.0
ENVIRONMENTAL SCS & TECH.	67.2	20.5	12.2	24.7	31.7	43.6
GENERAL & INDUSTRIAL ENG.	64.4	21.8	13.8	23.3	35.4	41.3
HEALTH SCIENCES	66.7	20.7	12.6	24.1	32.0	43.8
INFORMATION & COMM. SCS.	64.5	22.2	13.2	22.8	38.5	38.7
INSTS. & INSTRUMENTATION	72.2	17.6	10.2	18.6	32.0	49.4
LAW & CRIMINOLOGY	63.9	22.1	14.0	23.6	35.3	41.2
MANAGEMENT & PLANNING	65.1	21.1	13.9	23.4	34.8	41.9
MATHEMATICS	68.2	20.1	11.7	19.0	33.4	47.6
MECHANICAL ENG. & AEROSPACE	66.2	21.0	12.8	23.1	32.6	44.3
PHYSICS & MATERIALS SCIENCE	73.6	18.9	7.5	24.0	30.9	45.0
PSYCHOLOGY	67.9	20.2	11.9	23.8	31.6	44.6
SOCIAL & BEHAVIORAL SCIENCES	63.6	22.4	14.0	23.8	38.0	38.2
SOCIOLOGY & ANTHROPOLOGY	63.9	22.3	13.8	23.6	37.9	38.5
STATISTICAL SCIENCES	68.1	19.8	12.1	21.3	32.6	46.1
Average over the 29 fields	67.5	20.5	12.0	23.3	33.0	43.7
Coefficient of variation	0.05	0.06	0.16	0.08	0.07	0.06

Table B1
The skewness of field citation distributions for very productive authors in the double extended count after normalizing for differences in individual productivity and individual mean citations, and average and coefficient of variation of the CSS results over the 29 fields.

	% of articles in category:			% of citations in category:		
	I (1)	II (2)	III (3)	I (4)	II (5)	III (6)
AGRICULTURE & FOOD SCIENCE	65.8	22.3	11.9	24.1	33.3	42.5
ASTRONOMY & ASTROPHYSICS	76.9	18.6	4.5	25.8	31.4	42.8
BASIC LIFE SCIENCES	67.4	21.8	10.8	25.2	32.9	41.9
BASIC MEDICAL SCIENCES	64.5	23.0	12.5	25.9	33.3	40.8
BIOLOGICAL SCIENCES	66.7	22.1	11.2	24.0	33.4	42.6
BIOMEDICAL SCIENCES	66.7	22.1	11.2	25.4	33.0	41.6
CHEMISTRY & CHEMICAL ENG.	67.1	22.0	10.9	25.0	33.1	41.9
CIVIL ENG. & CONSTRUCTION	63.4	23.1	13.5	22.8	34.1	43.1
CLINICAL MEDICINE	68.8	21.4	9.9	24.9	32.8	42.3
COMPUTER SCIENCES	68.4	20.6	11.0	18.9	33.6	47.6
EARTH SCIENCES & TECHNOLOGY	67.4	21.8	10.8	24.6	33.0	42.4
ECONOMICS & BUSINESS	66.7	21.4	11.9	22.4	33.1	44.6
EDUCATIONAL SCIENCES	63.7	22.5	13.7	23.2	33.3	43.5
ELECTRICAL ENG. & TELECOMM.	67.4	21.4	11.2	20.7	33.7	45.6
ENERGY SC. & TECHNOLOGY	66.6	22.0	11.4	22.2	33.7	44.0
ENVIRONMENTAL SCS & TECH.	66.3	22.1	11.6	24.6	33.1	42.3
GENERAL & INDUSTRIAL ENG.	63.2	23.0	13.8	23.0	33.7	43.3
HEALTH SCIENCES	65.6	22.3	12.1	24.0	33.3	42.6
INFORMATION & COMM. SCS.	63.0	22.7	14.3	22.6	33.4	43.9
INSTS. & INSTRUMENTATION	71.0	19.2	9.8	19.2	32.0	48.8
LAW & CRIMINOLOGY	62.8	23.3	13.9	23.4	33.9	42.7
MANAGEMENT & PLANNING	64.0	22.5	13.5	23.2	33.3	43.5
MATHEMATICS	67.5	21.3	11.2	18.5	34.3	47.2
MECHANICAL ENG. & AEROSPACE	65.2	22.4	12.4	22.9	33.6	43.6

Table B1 (*Continued*)

	% of articles in category:			% of citations in category:		
	I (1)	II (2)	III (3)	I (4)	II (5)	III (6)
PHYSICS & MATERIALS SCIENCE	71.1	20.3	8.6	23.0	32.7	44.3
PSYCHOLOGY	66.9	21.7	11.4	23.7	33.1	43.3
SOCIAL & BEHAVIORAL SCIENCES	62.3	23.1	14.6	23.7	33.4	42.8
SOCIOLOGY & ANTHROPOLOGY	62.5	23.0	14.4	23.4	33.6	42.9
STATISTICAL SCIENCES	67.3	21.4	11.4	21.1	33.5	45.4
Average over the 29 fields	66.4	21.9	11.7	23.2	33.3	43.6
Coefficient of variation	0.05	0.05	0.17	0.08	0.02	0.04

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.joi.2018.07.002.

References

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology*, *62*, 40–49.

Albarrán, P., Crespo, J., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, *88*, 385–397.

Albarrán, P., Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2015). Differences in citation impact across countries. *Journal of the American Society for Information Science and Technology*, *66*, 512–525.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, *47*, 901–920.

Caron, E., & van Eck, N. J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. In E. Noyons (Ed.), *19th international conference on science and technology indicators. "Context counts: Pathways to master big data and little data"* (pp. 79–86).

Costas, R., Bordons, M., van Leeuwen, T. N., & van Raan, A. F. J. (2009). Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of individual researchers. *Journal of the American Society for Information Science and Technology*, *60*, 740–753.

Crespo, J. A., Li, Yunrong, & Ruiz-Castillo, J. (2013). The measurement of the effect on citation inequality of differences in citation practices across scientific fields. *PLoS One*, *8*, e58727.

Crespo, J. A., Herranz, N., Li, Yunrong, & Ruiz-Castillo, J. (2014). The effect on citation inequality of differences in citation practices at the web of science subject category level. *Journal of the American Society for Information Science and Technology*, *65*, 1244–1256.

Glänzel, W. (2007). Characteristic scores and scales –a bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, *1*, 92–102.

Glänzel, W. (2011). The application of characteristic scores and scales to the evaluation and ranking of scientific journals. *Journal of Information Science*, *37*, 40–48.

Glänzel, W., & Schubert, A. (1988). Characteristic scores and scales in assessing citation impact. *Journal of Information Science*, *14*, 123–127.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*, 357–367.

Groeneveld, R. A., & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, *33*, 391–399.

Katz, J. S. (2016). What is a complex innovation system? *PLoS One*, *11*, e0156150

Li, Y., Castellano, C., Radicchi, F., & Ruiz-Castillo, J. (2013). Quantitative evaluation of alternative field normalization procedures. *Journal of Informetrics*, *7*, 746–755.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, *16*, 317–323.

Perianes-Rodríguez, A., & Ruiz-Castillo, J. (2015a). Multiplicative versus fractional counting methods for co-authored publications. The case of the 500 universities in the Leiden ranking. *Journal of Informetrics*, *9*, 974–989.

Perianes-Rodríguez, A., & Ruiz-Castillo, J. (2015b). Within and across department variability in individual productivity. The case of economics. *Scientometrics*, *102*, 1497–1520.

Price, D. J. de S. (1965). Networks of scientific papers. *Science*, *149*, 510–515.

Radicchi, F., & Castellano, C. (2012). A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions. *PLoS One*, *7*, e33833.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*, 17268–17272.

Ruiz-Castillo, J. (2014). The comparison of classification-system-based normalization procedures with source normalization alternatives in Waltman and Van Eck. *Journal of Informetrics*, *8*, 25–28.

Ruiz-Castillo, J., & Costas, R. (2014a). The Skewness of scientific productivity. *Journal of Informetrics*, *8*, 917–934.

Ruiz-Castillo, J., & Costas, R. (2014b). *The skewness of scientific productivity. Working paper 14-02*. Universidad Carlos III. http://hdl.handle.net/10016/18286

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*, 102–117.

Schubert, A., Glänzel, W., & Braun, T. (1987). A new methodology for ranking scientific institutions. *Scientometrics*, *12*, 267–292.

Seglen, P. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*, 628–638.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, *48*, 613–625.

Shorrocks, A. F. (1984). Inequality decomposition by population subgroups. *Econometrica*, *52*, 1369–1388.

Sinatra, R., Wang, D., Dweville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Sciences*, *354*, aaf5239.

Thijs, B., Debackere, K., & Glänzel, W. (2017). Improved author profiling through the use of citation classes. *Scientometrics*, *111*, 829–839.

Tijssen, R., Hollanders, H., & van Steen, J. (2010). *Wetenschaps en Technologie Indicatoren 2010*. Nederland's Observatorium: Wetenschap en Technologie (NOWT).

van Raan, A. F. J. (2008). Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of research groups. *Journal of the American Society for Information Science and Technology*, *59*, 565–576.

van Raan, A. F. J. (2006a). Statistical properties of bibliometric indicators: Research group indicator distributions and correlations. *Journal of the American Society for Information Science and Technology*, *57*, 1919–1935.

van Raan, A. F. J. (2006b). Performance-related differences of bibliometric statistical properties of researchers: Cumulative advantages and hierarchically layered networks. *Journal of the American Society for Information Science and Technology*, *57*, 408–430.

Vîiu, G. A. (2018). The lognormal distribution explains the remarkable pattern documented by characteristic scores and scales in scientometrics. *Journal of Informetrics*, *12*, 401–415.

Waltman, L., & Van Eck, N. J. (2015). Field normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, *9*, 872–894.

Waltman, L., & Van Eck, N. J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*, 699–716.

Waltman, L., & Van Eck, N. J. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics*, *7*, 833–849.

Waltman, L., Van Eck, N. J., & Van Raan, A. F. J. (2012). Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology*, *63*, 72–77.