Universidad
Carlos III de Madrid

e-Archivo

Institutional Repository

This is a postprint version of the following published document:

University Citation Distributions

Antonio Perianes-Rodriguez
Departamento de Biblioteconomía y Documentación, Universidad Carlos III, SCImago Research Group, Madrid 126, Getafe, Madrid 28903, Spain. E-mail: antonio.perianes@uc3m.es

Javier Ruiz-Castillo
Departamento de Economía, Universidad Carlos III, Madrid 126, Getafe, Madrid 28903, Spain. E-mail: jrc@eco.uc3m.es

We investigate the citation distributions of the 500 universities in the 2013 edition of the Leiden Ranking produced by The Centre for Science and Technological Studies. We use a Web of Science data set consisting of 3.6 million articles published in 2003 to 2008 and classified into 5,119 clusters. The main findings are the following. First, the universality claim, according to which all university-citation distributions, appropriately normalized, follow a single functional form, is not supported by the data. Second, the 500 university citation distributions are all highly skewed and very similar. Broadly speaking, university citation distributions appear to behave as if they differ by a relatively constant scale factor over a large, intermediate part of their support. Third, citation-impact differences between universities account for 3.85% of overall citation inequality. This percentage is greatly reduced when university citation distributions are normalized using their mean normalized citation scores (MNCSs) as normalization factors. Finally, regarding practical consequences, we only need a single explanatory model for the type of high skewness characterizing all university citation distributions, and the similarity of university citation distributions goes a long way in explaining the similarity of the university rankings obtained with the MNCS and the Top 10% indicator.

Introduction

Universities constitute a key vehicle in the production of knowledge in contemporary societies. However, the evaluation of the quality or the relevance of the research done by universities in a myriad of scientific fields is a very difficult problem. For the assessment of the performance of research units of all types during recent decades, academic bodies, public officials in charge of science policy, and specialists in the field of scientometrics have been paying increasing attention to one observable aspect of research in all fields: the citation impact of publications in the periodical literature.

In this article, we focus on this aspect of scientific research for the 500 universities included in the 2013 edition of the The Centre for Science and Technology Studies' (CWTS) Leiden Ranking (Waltman et al., 2012a)—the LR universities hereafter. We use a Web of Science (WoS) data set consisting of 3.6 million publications in the 2005 to 2008 period, the citations they receive during a 5-year citation window for each year in that period, and a classification system consisting of 5,119 clusters (Ruiz-Castillo & Waltman, 2015).

The construction of university citation distributions in the all-sciences case requires the prior solution of two methodological problems: (a) the assignment of responsibility for publications with two or more coauthors belonging to different institutions and (b) the aggregation of the citation impact achieved by research units working in different scientific clusters. We solve these problems using a fractional counting approach in the case of coauthorship and the standard field-normalization procedure where cluster mean citations are used as normalization factors.[1]

Once these two problems have been solved, specialists typically debate the properties of alternative citation impact indicators. In this article, we study a basic aspect of the research evaluation problem that comes *before* the comparison of the advantages and shortcomings of specific indicators: namely, the characteristics of the university-citation distributions themselves. Differences between these

---

[1]Nevertheless, we study the robustness of some of our key results using a multiplicative rather than a fractional counting approach in the coauthorship case, and considering the university unnormalized citation distributions where every article receives the raw number of citations that appear in the initial data set.

distributions arise from the interplay of a complex set of economic, sociological, and intellectual factors that influence the research performance of each university in every field in a way that is difficult to summarize. In this scenario, it is well known that some universities are more productive or successful than are others in terms of the mean number of citations that these publications receive. However, little is known concerning the shape of university citation distributions not taking size and mean citation differences into account. Thus, in this article we investigate the following four issues.

- First, we inquire whether university citation distributions are universally distributed. The universality condition, borrowed from statistical physics, means that, appropriately normalized, citation distributions follow a unique functional form within the bounds set by random variation. Radicchi, Fortunato, and Castellano (2008) suggested a statistical test of this condition in their study of 14 WoS journal subject categories. According to this test, the universality condition is not satisfied by our 500 university citation distributions.[2]
- Second, despite the fact that the universality condition is not satisfied, are university citation distributions as highly skewed and as similar among one another as previous results have indicated for field citation distributions? Using the same size- and scale-independent techniques that have been used in previous research, we confirm that this is the case in our data set.[3]
- Third, using the measurement framework introduced in Crespo, Li, and Ruiz-Castillo (2013), we investigate the importance of the effect in the overall citation inequality that can be attributed to differences in citation impact between LR universities. Furthermore, we inquire up to what point this effect can be accounted for by scale factors captured by the universities' mean normalized citation score (MNCS). The answer is that citation impact differences between universities account for 3.85% of overall citation inequality—a much smaller percentage than that found in the context of production- and citation-practice differences between scientific fields (Crespo, Herranz, Li, & Ruiz-Castillo, 2014; Crespo et al., 2013; Waltman & Van Eck, 2013; Perianes-Rodriguez & Ruiz-Castillo, 2014; Ruiz-Castillo, 2014; Ruiz-Castillo & Waltman, 2015). This percentage is greatly reduced when university citation distributions are normalized using their MNCS values as normalization factors.

- Finally, we discuss the implications of these results for understanding the high correlation between the university rankings according to two citation impact indicators: the MNCS, and the Top 10% indicator of scientific excellence (the T indicator), defined as the percentage of an institution's output included into the set formed by 10% of the world's most cited papers in the different scientific fields. The latter indicator has been recently adopted by institutions such as the CWTS in the Netherlands and SCImago in Spain.[4]

Thus far, we have mentioned previous research on individual productivity distributions and citation distributions at the field level in different classification systems. Now we cite the contributions closer to our own system in which research publications are aggregated into the type of organization unit to which the authors belong. First, using a large WoS data set consisting of 4.4 million articles published in 1998 to 2003 with a 5-year citation window for each year, Albarrán et al. (2015) found that at least in some broad fields and in the all-sciences case, the citation distributions of 36 countries and two residual geographical areas are not only highly skewed but also very similar across countries—a result parallel to our own for the 500 LR universities. Second, Perianes-Rodriguez and Ruiz-Castillo (2015) studied a set of 2,530 highly productive economists who in 2007 worked in a selection of the top-81 economics departments in the world. Contrary to previous results for field or country citation distributions, we found that productivity distributions were very different across the 81 economics departments. Finally, Chatterjee et al. (2014) studied 42 academic institutions across the world, their publications in 4 years (1980, 1990, 2000, 2010), and the citations they have received up to July 2014 according to the WoS.[5] Contrary to our first result, these authors claimed that their 42 citation distributions satisfy the universality condition. As we will see, when examined in detail, the results of the last two articles can be reconciled with ours.

The rest of the article is organized into six sections. In the first section, we present the data and the methods applied in the construction of university citation distributions: the standard field-normalization procedure and the address-line fractional counting method. The next three sections include the empirical results: We examine the failure of the universality condition in our data set, and then discuss the skewness and similarity between university citation distributions, as well as the robustness of the results when we consider unnormalized university citation distributions and the address-line multiplicative counting method. Next, we study the effect in the overall citation inequality of the citation impact differences between universities as well as the

---

[2]This is consistent with previous results for large classification systems in WoS data sets consisting of complete field citation distributions that include publications with zero citations (Albarrán & Ruiz-Castillo, 2011; Albarrán et al., 2011a; Perianes-Rodriguez & Ruiz-Castillo, 2014; Waltman et al., 2012b;).

[3]This result has been established for field citation distributions at different aggregation levels, publication years, and citation window lengths, and independently of whether the problem of the multiple assignment of publications to subfields in WoS data sets is solved by following a multiplicative or a fractional approach (Albarrán & Ruiz-Castillo, 2011; Albarrán et al., 2011a; Glänzel, 2007; Herranz & Ruiz-Castillo, 2012; Li et al., 2013; Perianes-Rodriguez & Ruiz-Castillo, 2014; Radicchi et al., 2008; Radicchi & Castellano, 2012; Ruiz-Castillo & Waltman, 2015; Waltman et al., 2012b). Similar conclusions concerning the skewness and similarity of individual productivity distributions are found when authors are classified into 30 broad scientific fields (Ruiz-Castillo & Costas, 2014).

[4]The CWTS introduced the T indicator as of the Leiden Ranking 2011/ 2012 (http://www.leidenranking.com/methodology.aspx), based on a WoS database, while SCImago did so as of the SCImago Institutions Rankings (SIR) 2011 World Report (http://www.scimagoir.com/pdf/sir_2011_world _report.pdf), based on the Scopus database (Elsevier B.V.).

[5]They also study 30 popular academic journals across physics, chemistry, biology, and medicine. However, the characteristics of journal citation distributions are beyond the scope of this article.

reduction of this effect when the university citation distributions are normalized using their MNCS values as normalization factors. Finally, we discuss the practical implications of our results and then offer some conclusions and suggestions for further research.

Construction of the University Citation Distributions

The Data

Our data set results from the application of the publication-level algorithmic method introduced by Waltman and Van Eck (2012) to a WoS data set consisting of 9,446,622 publications from 2003 to 2012. This is done along a sequence of 12 independent classification systems, in each of which the same set of publications is assigned to an increasing number of clusters. Ruiz-Castillo and Waltman (2015) discussed the advantages of this method at high disaggregation levels over the standard WoS classification system consisting of 236 journal subject categories. Note that the 2015 edition of the CWTS LR already uses a publication-level, algorithmically constructed classification system consisting of 3,822 clusters (see the Methodology section in http://www.leidenranking.com/methodology/fields). In this article, we use the classification system recommended in Ruiz-Castillo and Waltman (2015), consisting of 5,119 clusters.

Only publications of the document types article and review, referred to in the sequel as "articles" or "publications," are considered. Publications in local journals, as well as popular magazines and trade journals, have been excluded. We work with journals in the sciences, the social sciences, and the arts and humanities, although many arts and humanities journals are excluded because they are of a local nature. In this article, we focus on the set of 3,614,447 articles published in 2005 to 2008, and the citations they receive during a 5-year citation window for each year in that period (for further details, see Ruiz-Castillo & Waltman, 2015).

The All-Sciences Aggregation Problem

Using a measurement framework introduced in Crespo et al. (2013), recent research has established that different normalization procedures perform quite well in eliminating most of the effect in overall citation inequality that can be attributed to differences in production and citation practices between fields (Crespo et al., 2014; Crespo et al., 2013; Li, Castellano, Radicchi, & Ruiz-Castillo, 2013; Ruiz-Castillo, 2014; Waltman & Van Eck, 2013). We believe that the reason for the good performance of target (or cited-side) normalization procedures is that field citation distributions, although not universal, are extremely similar. Li et al. (2013) indicated that the best alternative among a wide set of target normalization procedures is the two-parameter system developed in Radicchi and Castellano (2012). However, recent results have indicated that the standard field-normalization procedure where field mean citations are used as normalization factors performs well in the sense already indicated (Crespo et al., 2014; Crespo et al., 2013; Li et al., 2013; Leydesdorff, Radicchi, Bornmann, Castellano, & de Nooye, 2012; Radicchi et al., 2008; Ruiz-Castillo, 2014). Consequently, in this article, we adopt this procedure as the solution to the all-sciences aggregation problem, so that the raw citations to the 3.6 million articles in the original data set are normalized using the 5,119 cluster mean citations as normalization factors.

Assignment of Responsibility in the Case of Coauthorship

We know the total number of address lines appearing in each publication, but only for the 500 LR universities do we have information concerning the correspondence between address lines and research organizations. As in Waltman et al. (2012a), the 2,420,054 distinct articles, or 67% of the total, with at least one address line belonging to an LR university are assigned to universities using the following fractional counting method. An article is fully assigned to an LR university only if all addresses mentioned in the publication belong to the university in question. If a publication is coauthored by two or more LR universities, then it is assigned fractionally to each of them in proportion to the number of address lines in each case. For example, if the address list of an article contains five addresses and two of them belong to a particular university, then 0.4 of the article is assigned to this university, and only 0.2 of the article is assigned to each of the other three universities. Finally, consider a publication coauthored by an LR university and an unknown number of other institutions outside the LR. Assume, for example, that the publication has four address lines, two of which correspond to the LR university. In this case, only 0.5 of the article will be assigned to the LR university.

This completes the construction of the 500 university field-normalized citation distributions according to the address-line fractional counting method. For simplicity, in the sequel they are referred to as *university citation distributions* and are denoted as c_i, $i = 1, \ldots, 500$, where articles are ranked in increasing order of the fractional number of field-normalized citations. Note that for each university, the mean citation of c_i is precisely the MNCS. Finally, note also that the fractional counting method implies that the total (fractional) number of articles assigned to LR universities is necessarily smaller than the 2.4 million articles with at least one address line belonging to an LR university. This total is 1,886,106.1, or 52.2% of the 3.6 million articles in the initial data set. The distribution of the 1.9 million articles among the 500 universities is in column 1 in Table A1 in the Supplementary Material (SMS hereafter; available at wileyonlinelibrary.com), where universities are ordered by their MNCS values.

3

Following a suggestion from one referee, we introduce a numerical example to illustrate the construction of university citation distributions from the raw data. This example will be used to illustrate other aspects of our argument. Consider a situation with two universities, *A* and *B*, and two clusters, *1* and *2*. For simplicity, we assume that every article has been written by only one university (i.e., there is no coauthorship). Each university and each cluster has six articles. The *original university citation distributions*, u_A and u_B, consisting of the raw citations received by each article are:

	Cluster 1	Cluster 2
u_A	(2, 26, 54)	(0, 1, 22)
u_B	(0, 4, 10)	(1, 7, 11)

The corresponding *cluster raw citation distributions*, r_1 and r_2, as well as their mean citations μ_1 and μ_2, are:

Cluster raw citation distributions	Cluster mean citations
$r_1 = (0, 2, 4, 10, 26, 54)$	$\mu_1 = 96/6 = 16$
$r_2 = (0, 1, 1, 7, 11, 22)$	$\mu_2 = 42/6 = 7$

The overall raw citation distribution, $R = r_1 \cup r_2$, is:

$R = (0, 0, 1, 1, 2, 4, 7, 10, 11, 22, 26, 54)$

On the other hand, the *field-normalized cluster citation distributions*, r_1^* and r_2^*, using μ_1 and μ_2 as normalization factors, are:

$r_1^* = (0, 1/8, 2/8, 5/8, 13/8, 27/8)$
$r_2^* = (0, 1/7, 1/7, 1, 11/7, 22/7)$

Therefore, the field-normalized university citation distributions, or simply the *university citation distributions*, c_A and c_B, are:

$c_A = (1/8, 13/8, 27/8) \cup (0, 1/7, 22/7) = (0, 1/8, 1/7, 13/8, 22/7, 27/8)$
$c_B = (0, 2/8, 5/8) \cup (1/7, 1, 11/7) = (0, 1/7, 2/8, 5/8, 1, 11/7)$

Note that the overall field-normalized citation distribution, or simply the *overall citation distribution*, $C = c_A \cup c_B = r_1^* \cup r_2^*$, is:

$C = (0, 0, 1/8, 1/7, 1/7, 2/8, 5/8, 1, 11/7, 13/8, 22/7, 27/8)$

Finally, the means of the university citation distributions, which coincide with their MNCS values, are $MNCS_A = 1.4$, and $MNCS_B = 0.6$. Of course, ½ $MNCS_A$ + ½ $MNCS_B = 1$.

Universality of University Citation Distributions

Methods

Let *C* be the *overall citation distribution*, $C = \cup_i \{c_i\}$, where c_i is the (field-normalized) citation distribution of university *i*. Let X_z be the set of publications in the top *z*% of distribution *C*, and let x_{zi} be the publications in X_z that belong to the *i*th university, so that $X_z = \cup_i \{x_{zi}\}$. Naturally, universities with a high citation impact would have large x_{zi} sets. If we let n_{zi} and N_i be the number of articles in x_{zi} and c_i, then we expect that the proportion that n_{zi} represents in N_i will be larger than *z*%. On the contrary, universities with a low citation impact would have small x_{zi} sets, so that the proportion that n_{zi} represents in N_i will be smaller than *z*%.

Next, let c_i^* be the *normalized citation distribution* of university *i* using the university MNCS as the normalization factor, and let $C^* = \cup_i \{c_i^*\}$ be the *overall normalized citation distribution*. Let Y_z be the set of publications in the top *z*% of distribution C^*, and let y_{zi} be the publications in Y_z that belong to the *i*th university, so that $Y_z = \cup_i \{y_{zi}\}$. For each *i*, let that n_{zi}^* be the number of publications in y_{zi}. If the university citation distributions follow a unique functional form under the universality condition, so that—in the terminology of Radicchi et al. (2008)—the ranking in Y_z is fair, or unbiased, then the percentage that n_{zi}^* represents in N_i should be near *z*%, with small fluctuations. Assuming that publications of the various universities are scattered uniformly along the rank axis, for any value *z*%, one would expect the average relative frequency of the number of articles in any university to be *z*% with an *SD*

$$\sigma_z = \left\{ \left[z(100-z) \sum_i (1/N_i) \right] \big/ 500 \right\}^{1/2},$$

which is Equation 2 in Radicchi et al. (2008).

Results

For each *z* value, columns 2 and 3 in Table 1 present the theoretical *SD* and coefficient of variation, σ_z and σ_z/z. Columns 4 to 6 contain the values for the average *z*, the *SD* σ_z, and the coefficient of variation σ_z/z obtained empirically over the 500 values $(100\, n_{zi}^*)/N_i$ in distribution C^*. For comparison purposes, Columns 7 to 9 report the average *z*, the *SD* σ_z, and the coefficient of variation σ_z/z obtained empirically over the 500 values $(100\, n_{zi})/N_i$ in distribution *C* before the normalization of university distributions by MNCS values.

The following three points should be emphasized. First, although σ_z varies nonlinearly with *z*, the theoretical coefficient of variation in column 3 increases from 0.01 to 0.20 when we proceed from *z* = 90% toward *z* = 1%. Second, in the union of unnormalized university citation distributions, the range of the coefficients of variation in column 9 is [0.05, 0.63], indicating that the distributions are very different. Third, the normalization using university MNCS values clearly decreases the coefficients of variation at all *z* values (column 6). Nevertheless, the differences with the

TABLE 1. Percentage of articles in each university that appear in the top z% of the global rank, together with the standard deviation, σ_z, and the coefficient of variation, σ_z/z.

			Empirical values in:					
Theoretical values			Normalized distribution C^*			Un-normalized distr. C		
$z\%$	σ_z	σ_z/z	$z\%$	σ_z	σ_z/z	$z\%$	σ_z	σ_z/z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	0.20	0.20	0.96	0.29	0.30	0.82	0.52	0.63
5	0.43	0.09	4.95	0.90	0.18	4.33	1.93	0.45
10	0.59	0.06	10.00	1.46	0.15	8.91	3.22	0.36
20	0.79	0.04	20.03	2.41	0.12	18.30	5.10	0.28
30	0.91	0.03	30.04	3.11	0.10	27.90	6.44	0.23
40	0.97	0.02	40.00	3.49	0.09	37.67	7.25	0.19
50	0.99	0.02	49.88	3.76	0.08	47.59	7.63	0.16
75	0.86	0.01	74.73	4.08	0.05	73.08	6.57	0.09
90	0.59	0.01	88.94	4.08	0.05	88.93	4.07	0.05

theoretical values in column 3, above all for lower values of z, indicate that for our set of university citation distributions, the universality condition is not satisfied.

Following up on Waltman, Van Eck, and Van Raan (2012b), the situation is illustrated for $z = 10\%$, in which case the theoretical value of σ_z is 0.59 (Table 1). The histogram of the percentages $(100\ n^*_{10i})/N_i$ is represented in Figure 1A. Only 132 universities, representing 25.5% of all articles in X_{10}, are within the (9, 11) interval. Naturally, the situation improves when we consider the histogram of the percentages $(100\ n_{10i})/N_i$ in Figure 1B. Now, 295 universities, representing 68.4% of all articles in Y_{10}, are within the (9, 11) interval. However, the large number of universities outside the theoretical interval illustrates the lack of universality in our data set.

This conclusion contrasts with the universality claim in Chatterjee et al. (2014). Note that their article has a number of technical problems. The criterion for selecting their 42 academic institutions is not given, and there is no information on how the following three problems have been solved: the assignment of publications in WoS data sets to multiple journal subject categories, the assignment of responsibility for coauthored publications, and the all-sciences aggregation problem. Nevertheless, we will discuss their results.

Chatterjee et al. (2014) explained that for each of four publication years, the university normalized citation distributions fit well to a lognormal for most of the range, although the poorly cited publications seem to follow another distribution while the upper tail is better described by a power law. This is quite different from the claim that there is a single functional form for the entire domain of definition of the 42 institutions in their sample. Instead, our statistical approach tests whether the universality claim is supported by the data over the *entire domain* of the 500 LR universities. In this sense, our results do not contradict each other. We both agree that the universality claim over the entire domain is not the case in our respective samples.

On the other hand, apart from the technical shortcomings already noted, the main problem with the still-unpublished version of Chatterjee et al. (2014) is that in our opinion, their statistical method is open to question. Specifically, the authors do not explain the following three aspects: (a) how the partition of the domain into three segments is estimated for each university, and whether this partition is common to all of them; (b) which tests have been used to determine the functional form chosen in each segment versus possible alternatives; and (c) how the confidence interval for the power law parameter has been estimated, and which is the confidence interval for the lognormal parameters. The only clear evidence for the distributions' collapse into a universal curve is the graphical illustration provided for a sample—whose selection is unexplained—of 24 of the original 42 academic institutions (see Chatterjee et al., 2014, Figure 1).

Skewness and Similarity of University Citation Distributions

Methods

Universities are known to be very different in size, measured by the number of articles as well as in mean number of citations per article which is simply the MNCS (columns 1 and 2 in Table A1 in the SMS). Therefore, we should focus on the shape of university citation distributions abstracting from size and scale differences between fields. The skewness of citation distributions is assessed in the following two complementary approaches.

First, we study the broad features of the skewness phenomenon by simply partitioning citation distributions into three classes of articles: low, fair, and very high number of citations. For this purpose, we follow the characteristic scores and scales (CSS) approach, first introduced in scientometrics by Schubert, Glänzel, and Braun (1987). In our application of the CSS technique, the following two *characteristic scores* are determined for every university: μ_1 = mean citation, and μ_2 = mean citation for articles with citations greater than μ_1. We consider the partition of the distribution into three broad categories: (a) articles with a low number of citations, smaller than or equal to μ_1; (b) fairly cited articles, with a number of citations greater than μ_1 and smaller than or equal to μ_2; and (c) articles with a remarkable or outstanding number of citations greater than μ_2. For each citation distribution, we measure the percentages of publications in the three categories as well as the percentages of the total citations accounted for by the three categories.

Second, we summarize the skewness of citation distributions with a single scalar. The problem, of course, is that extreme observations of publications with a very large number of citations are known to be prevalent in citation distributions. Fortunately, robust measures of skewness based on quartiles have been developed in the statistics literature. Among the size- and scale-independent measures that also are robust to extreme observations, in this article,
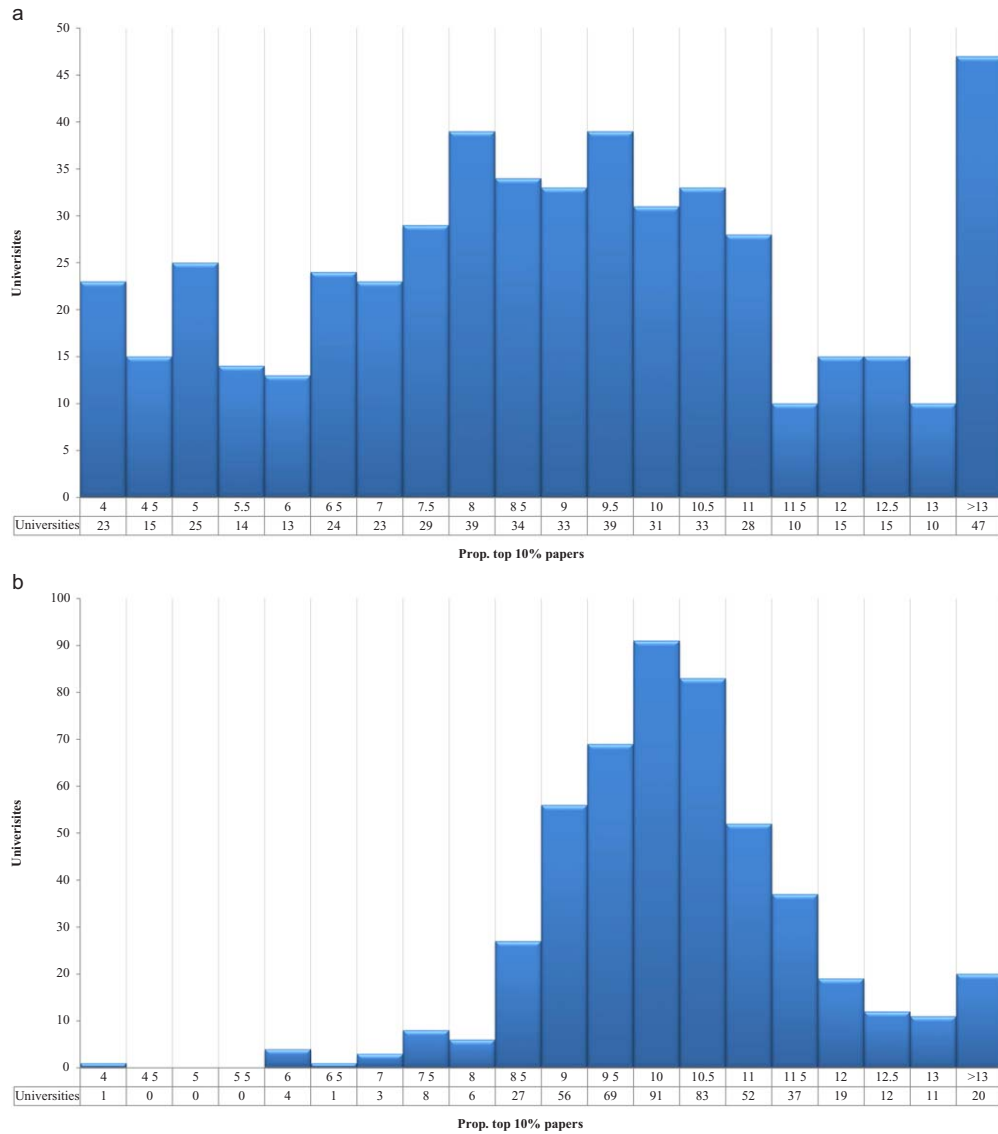
a

Prop. top 10% papers	4	4 5	5	5.5	6	6 5	7	7.5	8	8 5	9	9.5	10	10.5	11	11 5	12	12.5	13	>13
Universities	23	15	25	14	13	24	23	29	39	34	33	39	31	33	28	10	15	15	10	47

b

Prop. top 10% papers	4	4 5	5	5 5	6	6 5	7	7 5	8	8 5	9	9 5	10	10.5	11	11 5	12	12.5	13	>13
Universities	1	0	0	0	4	1	3	8	6	27	56	69	91	83	52	37	19	12	11	20

FIG. 1. A. Distribution over the 500 universities of the top 10% articles in the unnormalized citation distribution C. Histogram of the percentage that these articles represent with respect to the total number of articles in each university.
B. Distribution over the 500 universities of top 10% articles in the overall normalized citation distribution C^*. Histogram of the percentage that these articles represent with respect to the total number of articles in each university.
[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

we use the one suggested by Groeneveld and Meeden (1984).[6] Given a process $\{y_t\}$, $t = 1, \ldots, T$, where the y_ts are independent and identically distributed with a cumulative distribution function F, the Groeneveld–Meeden (GM) robust measure is defined as

$$GM = (\mu - \Theta_2)/\mathrm{E}|y_t - \Theta_2|, \qquad (1)$$

where $\Theta_2 = \mathrm{F}^{-1}(0.5)$ is the second quartile of y_t, or the median of the distribution, and the expectation in the

[6]For references concerning robust measures of skewness in the context of the financial literature on stock market returns, and for the properties of the Groeneveld and Meeden's measure, see Albarrán et al. (2015).

denominator in Equation (1) is estimated by the sample mean of the deviations from the median in absolute value. The GM index is bounded in the interval $[-1, 1]$, and whenever the mean is greater than the median—as is always the case in our data set—the GM index takes positive values.

For the interpretation of results, recall that the absence of skewness in a uniform or a normal distribution corresponds to a value of the GM index equal to zero, and to a partition of the population into three classes in the CSS approach equal to 50, 25, and 25%. Finally, the between-group variability of the results of the CSS approach and the GM index is measured by the coefficient of variation of the results over the 500 universities.

The information concerning the second mean, μ_2, as well as the percentages of articles and citations in the three categories are presented in columns 3 to 9 in Table A1 in the SMS. Finally, the estimates of the GM index of skewness are in column 10 of this table.

The average, *SD*, and coefficient of variation for the 500 university values are presented in Panel A in Table 2. This panel also includes the results of the CSS approach for two important citation distributions: (a) the union of the 500 field-normalized university citation distributions, $C = \cup_i \{c_i\}$ (the LR union), consisting of 1.9 million articles according to the fractional approach; and (b) the overall citation distribution consisting of the 3.6 million distinct articles in the original data set before the assignment of articles to universities.

The results are remarkable. In principle, differences in resources, intellectual traditions, organization, the structure of incentives, and other factors lead us to expect large differences between the 500 LR university citation distributions in different parts of the world. However, judging from the size of the *SD*s and the coefficients of variation for the 500 universities, we find that university citation distributions are extremely similar (row I in Table 2). At the same time, the distributions are highly skewed: On average, the MNCS values of the 500 universities is 12.9 percentage points above the median while the 12.5% of outstanding articles account for 44.4% of all normalized citations. Figure 2 clearly illustrates the situation. It comes as no surprise that the union of the 500 field-normalized university citation distributions exhibits practically the same skewness as the average of the 500 universities (row II.1 in Table 2). Furthermore, the overall field-normalized citation distribution exhibits very similar characteristics (row II.2 in Table 2).

On the other hand, the results concerning the GM index confirm that university citation distributions are both highly skewed and extremely similar (row I in Panel A in Table 3). The GM index is somewhat smaller for the LR union, and the overall citation distribution (rows II.1 and II.2 in Panel A in Table 3).

Robustness Analysis

For the sake of robustness, we conducted two more sets of computations. First, in the case of coauthorship, we have assigned publications to universities following a multiplicative approach. Thus, any coauthored article is multiplied as many times as the number of address lines that appear in the byline of the publication in question. For example, assume that the address list of an article contains six addresses, two of which belong to a particular LR university (university A), two other to a non-LR university (university B), and the remaining two addresses belong to two other non-LR universities (universities C and D). Then, two articles are assigned to university A, another two articles to university B, and only one article is assigned to each of the universities C

and D. In this way, we construct what we call an *extended count* of 8,329,951 articles, or 230.4% of the 3.6 million articles published from 2005 to 2008.[7]

Next, field normalization proceeds using the cluster mean citations in the extended count as normalization factors. In turn, the restriction of the extended count to the set of publications with at least one address line in one of the LR universities now has 4,351,584 million articles, or 180% more than the original 2.4 million. To save space, the information concerning the characteristics included in Table A1 in the SMS for each LR university is available upon request. The key results for the set of 500 LR universities the 4.3 million articles in the LR union, and the 8.3 million articles in the overall citation distribution are presented in Panel B in Table 2. The results for the GM index are in Panel B in Table 3.

Second, we studied the raw citation distributions without the benefit of any field-normalization procedure. Consider the raw citation distribution consisting of the 2.4 million articles in which there is at least one address line corresponding to one LR university. In this case, we assign coauthored publications to universities according to the fractional counting method, so that the LR unnormalized union has, again, approximately 1.9 million articles. As before, the information concerning the characteristics included in Table A1 in the SMS for each LR university is available upon request. The key results for the set of 500 LR universities as well as for the corresponding LR union and the overall citation distribution are presented in Panel C in Table 2. The results for the GM index are in Panel C in Table 3.[8]

The results in the two exercises are very similar to those obtained for field-normalized university citation distributions in the fractional case. Note that the GM values in Panel C in Table 3 indicate that university citation distributions in the absence of field normalization are somewhat more skewed than when we consider the standard solution to the all-sciences aggregation problem. Thus, we conclude that the characteristics of university citation distributions are robust to the way the assignment of publications to universities in the presence of coauthorship and the all-sciences aggregation problem are solved.

By way of comparison, we include in Panel D in Table 2 the results from the CSS approach to subfield citation

---

[7]Ideally, we would have preferred to assign publications to universities without taking into account the number of address lines corresponding to them. Thus, in the example, we would have multiplied the article only four times, assigning them to each of the four universities A to D. Unfortunately, as we noted earlier, we only have information about the number of address lines of specific institutions for the 500 LR universities. Consequently, we could only use the total number of address lines in the construction of the extended count.

[8]Note that the overall citation distribution in this case coincides with the distribution corresponding to the granularity level 0 in Table 2 in Ruiz-Castillo and Waltman (2015).

TABLE 2. The skewness of citation distributions according to the CSS approach. Percentages of articles and citations by category. Average, standard deviation, and coefficient of variation over the 500 LR universities, results for the union of the LR universities, and the overall citation distribution.

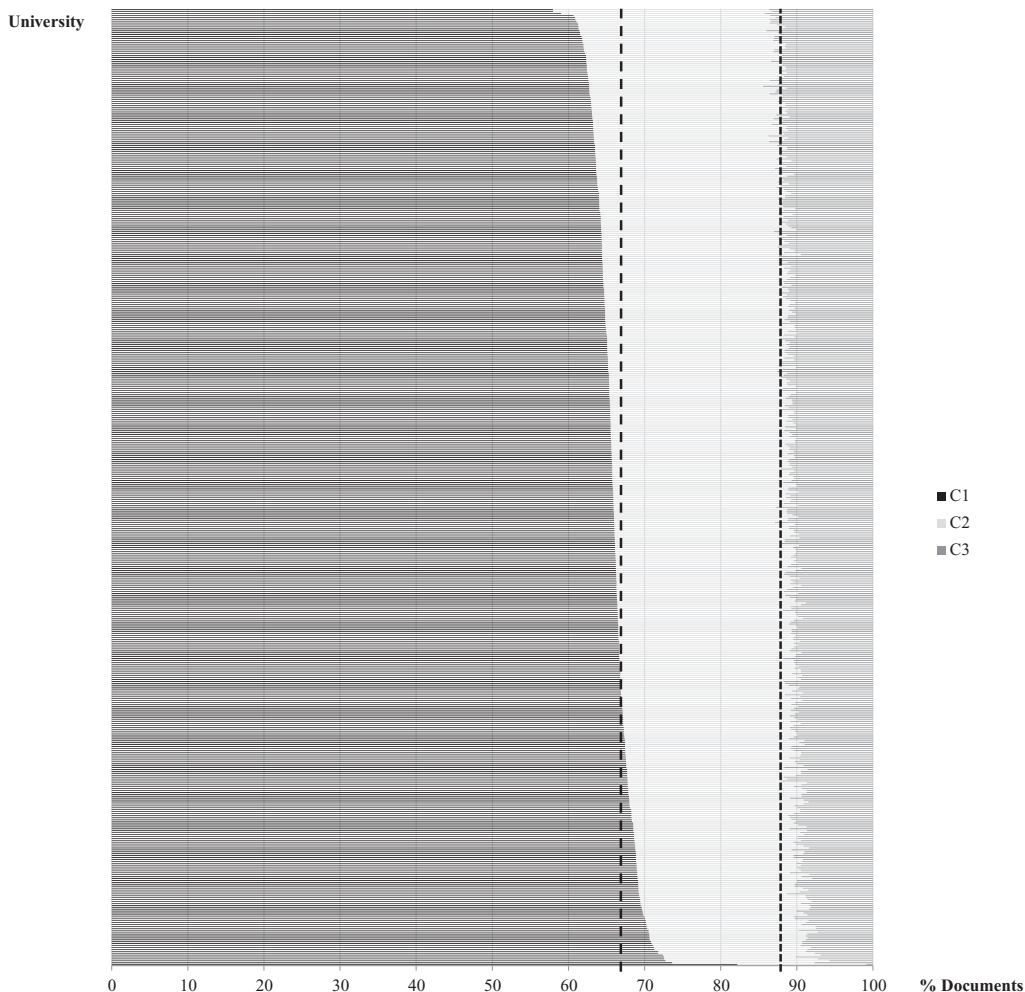|	Percentage of articles in category			Percentage of articles in category		
	1	2	3	1	2	3
A. Field-normalized citation distributions. Assignment of articles to universities according to the fractional method						
I. Average (*SD*)	62.9 (1.9)	24.6 (1.2)	12.5 (1.2)	22.9 (1.7)	32.7 (0.8)	44.4 (1.5)
Coefficient of variation	0.03	0.05	0.10	0.08	0.02	0.03
II.1. LR union (1.9 million articles)	62.8	24.8	12.4	23.4	32.6	44.0
II. 2. Overall citation distribution (3.6 million articles)	69.2	21.7	9.1	23.7	33.7	42.6
B. Field-normalized citation distributions. Assignment of articles to universities according to the multiplicative method						
III. Average (*SD*)	66.3 (2.3)	22.7 (1.0)	10.9 (1.6)	24.4 (1.8)	32.7 (1.0)	42.9 (1.5)
Coefficient of variation	0.03	0.04	0.14	0.08	0.03	0.03
IV.1. LR union (4.3 million articles)	65.9	23.0	11.1	25.1	32.3	42.6
IV. 2. Extended count (8.3 million articles)	62.3	24.4	13.3	22.1	32.2	45.7
C. Raw citation distributions. Assignment of articles to universities according to the fractional method						
V. Average (*SD*)	65.9 (2.4)	23.3 (1.5)	10.8 (1.3)	22.7 (1.8)	32.3 (1.1)	45.0 (1.8)
Coefficient of variation	0.04	0.07	0.12	0.08	0.03	0.04
VI.1. LR union (1.8 million articles)	70.9	20.7	8.4	26.3	31.6	42.1
VI.2. Overall citation distribution (3.6 million articles)	72.0	20.2	7.8	22.6	32.2	45.2
D. Previous results for citation distributions in a comparable case. Articles published in 1998–2002 in 219 subfields with a fixed, 5-year citation window. Panel A in Table 1, in Albarrán and Ruiz-Castillo (2011):						
VII. Average (*SD*)	68.6 (3.7)	-	10.0 (1.7)	21.1 (5.0)	-	44.9 (4.6)
Coefficient of variation	0.05		0.17	0.24		0.10

FIG. 2. Partition of the citation distributions for the 500 Leiden Ranking universities into three categories according to the CSS technique.

distributions in a classification system consisting of 219 WoS journal subject categories, or subfields, in Albarrán, Crespo, Ortuño, and Ruiz-Castillo (2011a). The results for university citation distributions and for subfield citation distributions are of the same order of magnitude. The same can be said for authors' productivity distributions in a WoS data set with a classification system consisting of 30 broad scientific fields for two measures of individual productivity (Ruiz-Castillo & Costas, 2014).

Finally, we should mention the results of two contributions closer to our own in which research publications are aggregated into the type of organization unit to which the authors belong. First, Albarrán et al. (2015) studied the partition of world citation distributions into 36 countries and two residual geographical areas using a data set, comparable to ours, consisting of 4.4 million articles published in 1998 to 2003, with a 5-year citation window for each year. As indicated earlier, Albarrán et al. (2015) found that at least in some broad fields and in the all-sciences case, the country citation distributions are not only highly skewed but also

TABLE 3. The skewness of citation distributions according to the GM index. Average, standard deviation, and coefficient of variation over the 500 LR universities, results for the union of the LR universities, and the overall citation distribution.

A. Field-normalized citation distributions: Fractional method	
I. Average (*SD*)	0.63 (0.03)
Coefficient of variation	0.04
II.1. LR union (1.9 million articles)	0.56
II.2. Overall citation distribution (3.6 million articles)	0.58

B. Field-normalized citation distributions: Multiplicative method	
III. Average (*SD*)	0.62 (0.03)
Coefficient of variation	0.05
IV.1. LR union (4.3 million articles)	0.53
IV. 2. Extended count (8.3 million articles)	0.72

C. Raw citation distributions: Fractional method	
V. Average (*SD*)	0.68 (0.5)
Coefficient of variation	0.07
VI.1. LR union (1.9 million articles)	0.75
VI.2. Overall citation distribution(3.6 million articles)	0.79

9

very similar across countries—a result parallel to our own for the 500 LR universities.

Second, Perianes-Rodriguez and Ruiz-Castillo (2015) studied a set of 2,530 highly productive economists who worked in 2007 in a selection of the top-81 economics departments in the world. Contrary to previous results for field or country citation distributions, we found that productivity distributions are very different across the 81 economics departments. However, certain characteristics of the data set may help to explain the different results. First, the data in Perianes-Rodriguez and Ruiz-Castillo (2015) do not consist of department citation distributions of articles published in a certain time period with a citation window of common length, but of the individual productivity of faculty members in each department, where individual productivity is measured as a quality index that weights differently the articles published up to 2007 by each researcher in four journal-equivalent classes. Second, information about researchers' publications and academic age was taken from the departments' and the individuals' web pages in 2007. Nevertheless, we cannot rule out that the similarity of citation distributions is a phenomenon present at certain levels of aggregation. To settle this issue, we need more work at the department level with citation distributions consisting of articles published in a certain period of time with a common citation window.

Importance of Citation Impact Differences Between Universities

Methods

We are interested in measuring how important the citation impact differences are between universities. Formally, this problem is analogous to the measurement of the importance of differences in production and citation practices between scientific fields. For the latter, Crespo et al. (2013) suggested measuring the impact of such differences on the overall citation inequality for the entire set of field citation distributions by applying an additively decomposable citation inequality index to a double partition into scientific fields and quantiles. Similarly, in our case, we measure how much of the overall citation inequality exhibited by the union of the 500 LR university citation distributions can be attributed to the citation impact differences between universities (This also is the approach adopted in Albarrán et al., 2015, to assess the effect of citation impact between countries.)

For that purpose, we begin with the partition of, for example, each university citation distribution into Π quantiles, indexed by $\pi = 1, \ldots, \Pi$. In practice, in this article, we use the partition into percentiles; that is, we choose $\Pi = 100$. Assume that in any university u, we disregard the citation inequality within every percentile by assigning to every article in that percentile the mean citation of the percentile itself, μ_u^π. The interpretation of the fact that, for example,

$\mu_u^\pi = 2\,\mu_v^\pi$ is that, on average, the citation impact of university u is twice as large as is the citation impact of university v, despite the fact that both quantities represent a common underlying phenomenon: the same *degree of citation impact* in both universities. In other words, for any π, the distance between μ_u^π and μ_v^π is entirely attributable to the difference in the citation impact that prevails in the two universities for publications with the same degree of excellence in each of them. Thus, the citation inequality between universities at each percentile, denoted by $I(\pi)$, is entirely attributable to the citation impact differences between the 500 LR universities holding constant the degree of excellence in all universities at quantile π. Hence, any weighted average of these quantities provides a good measure of the total impact on overall citation inequality that can be attributed to such differences.

Recall that the overall citation distribution C is the union of the universities citation distributions, $C = \cup_i \{c_i\}$. Using an additively decomposable citation inequality index, it can be shown that the overall citation inequality, $I(C)$, can be decomposed into three terms, one of which, denoted by *I*nequality due to *D*ifferences in *C*itation impact between *U*niversities (*IDCU*), is a weighted average of the $I(\pi)$ values. Therefore, we use the ratio

$$100(IDCU)/I(C) \qquad (2)$$

to assess the relative effect on overall citation inequality, attributed to citation impact differences between universities (for details, see Crespo et al., 2013).

As an illustration, consider the example introduced earlier, where the university citation distributions, c_A and c_B, as well as the overall citation distribution $C = c_A \cup c_B$ were the following:

$$c_A = (0, 1/8, 1/7, 13/8, 22/7, 27/8),$$

$$c_B = (0, 1/7, 2/8, 5/8, 1, 11/7),$$

$$C = (0, 0, 1/8, 1/7, 1/7, 2/8, 5/8, 1, 11/7, 13/8, 22/7, 27/8).$$

Using our citation inequality index, we have $I(C) = 0.6456$. Next, consider the partition of the university citation distributions into three quantiles:

π	1	2	3
c_A	(0, 1/8)	(1/7, 13/8)	(22/7, 27/8)
c_B	(0, 1/7)	(2/8, 5/8)	(1, 11/7)

The array of mean citations μ_A^π and μ_B^π in every quantile π, as well as the citation inequality of the column vector (μ_A^π, μ_B^π) at every quantile, $I(\pi)$, are the following:

π	1	2	3
μ_A^π	0.06	0.88	3.26
μ_B^π	0.07	0.44	1.29
$I(\pi)$ terms	0.0022	0.05820	0.0974

The *IDCU* term, which is equal to a weighted average of the $I(\pi)$ terms, is 0.0867. Therefore, the value of Equation (2) is:

$$100(IDCU)/I(C) = 13.4\%.$$

In other words, in the example, the citation impact differences between universities A and B is responsible for 13.4% of the overall citation inequality exhibited by the union of the 500 LR university citation distributions.

On the other hand, we are interested in estimating how important scale differences between university citation distributions are in accounting for the effect measured by Equation (2). Following earlier practice (Albarrán et al., 2015; Crespo et al., 2014; Crespo et al., 2013), we choose the university mean citations—the MNCS values—as normalization factors. As before, let C^* be the union of the LR universities normalized citation distributions, $C^* = \cup_i \{c^*_i\}$, and denote by *IDCU** the term that measures the effect on overall citation inequality, $I(C^*)$, attributed to the differences in citation impact between the universities after the normalization of their citation distributions using the MNCS values as normalization factors. To assess the importance of such scale factors, we use the relative change in the *IDCU* term; that is, the ratio

$$[IDCU - IDCU^*]/IDCU. \qquad (3)$$

As an illustration, consider again the example introduced earlier. The *normalized university citation distributions c_A^** and c_B^* using the MNCS values $MNCS_A = 1.4$ and $MNCS_B = 0.6$ as normalization factors are the following:

$c_A^* = (0, 0.09, 0.10, 1.16, 2.24, 2.41)$,
$c_B^* = (0, 0.24, 0.42, 1.04, 1.66, 2.62)$.

Therefore, the overall normalized citation distribution, $C^* = c_A^* \cup c_B^*$, is the following:

$C^* = (0, 0, 0.09, 0.10, 0.24, 0.42, 1.04, 1.16, 1.66, 2.24, 2.41, 2.62)$.

Using our citation inequality index, we have $I(C^*) = 0.5321$. Next, consider the partition of the university citation distributions into three quantiles:

π	1	2	3
$c_A{}^*$	(0, 0.09)	(0.10, 1.16)	(2.24, 2.41)
$c_B{}^*$	(0, 0.24)	(0.42, 1.04)	(1.66, 2.62)

The array of mean citations $\mu^*_A{}^\pi$ and $\mu^*_B{}^\pi$ in every quantile π, as well as the citation inequality of the column vector $(\mu^*_A{}^\pi, \mu^*_B{}^\pi)$ at every quantile, $I(\pi)$, are the following:

π	1	2	3
$\mu^*_A{}^\pi$	0.045	0.631	2.32
$\mu^*_B{}^\pi$	0.119	0.728	2.14
$I(\pi)$ terms	0.1072	0.0027	0.0007

Since the *IDCU** term turns out to be 0.0041, the value of Equation (3) is:

$$100(IDCU^*)/I(C^*) = 0.77\%.$$

In other words, the citation impact differences between universities that were responsible for 13.4% of the overall citation inequality before normalization only represent 0.8% of the overall citation inequality after applying the normalization procedure where the MNCS values are used as normalization factors. Thus, the reduction of the *IDCU* term after normalization is:

$$100(IDCU - IDCU^*)/IDCU = 95.3\%.$$

Results

The estimates of Equations (2) and (3) are presented in panel A in Table 4. It is interesting to compare these figures

TABLE 4A. The effect on overall citation inequality, $I(C)$, of the differences in citation impact between universities before and after MNCS normalization, and the impact of normalization on this effect.

	Normalization impact $=100$ [$IDPD - IDCP^*/IDCP$]	
Before MNCS normalization, 100 [$IDPU/I(C)$]	3.85%	–
After MNCS normalization, 100 [$IDPU^*/I(C)$]	0.72%	81.3%

TABLE 4B. [a]The effect on overall citation inequality, $I(C)$, of the differences in citation impact between countries before and after MNCS normalization, and the impact of normalization on this effect.

	Normalization impact $=100$ [$IDPD - IDCP^*/IDCP$]	
Before MNCS normalization, 100 [$IDPU/I(C)$]	5.6%	–
After MNCS normalization, 100 [$IDPU^*/I(C)$]	0.9%	83.8%

TABLE 4C. [b]The effect on overall citation inequality, $I(C)$, of the differences in citation impact between subfields before and after mean normalization, and the impact of normalization on this effect.

	Normalization impact $=100$ [$IDPD - IDCP^*/IDCP$]	
Before MNCS normalization, 100 [$IDPU/I(C)$]	17.9%	–
After MNCS normalization, 100 [$IDPU^*/I(C)$]	3.45%	87.1%

Note. [a]Table 3 in Crespo, Li, and Ruiz-Castillo (2013). [b]Table 3 in Albarrán, Perianes-Rodriguez, and Ruiz-Castillo. (2015).
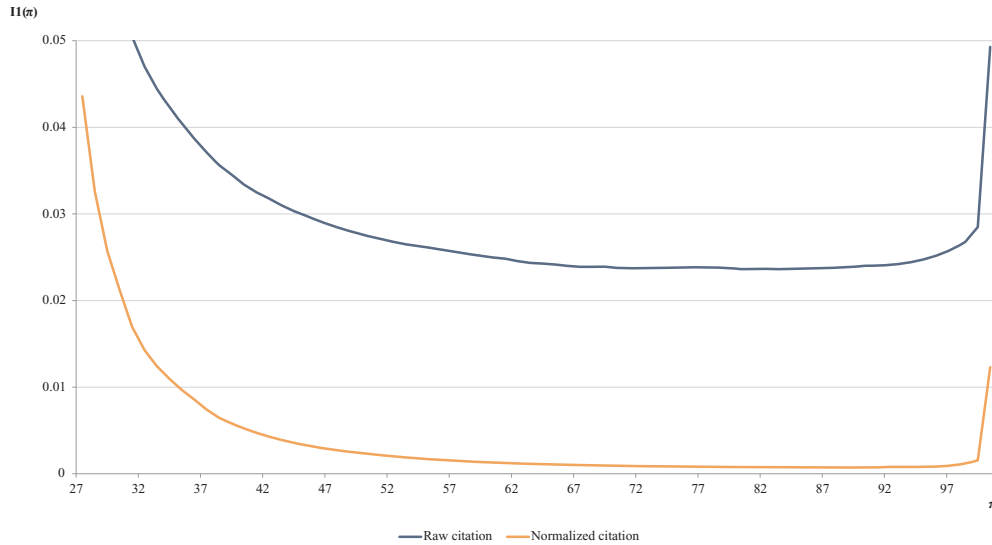
FIG. 3. Citation inequality due to differences in citation impact between universities, I(π), as a function of π. Results for the [27, 100] quantile interval. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

with what was obtained in two instances in previous literature. The first case concerns the partition into 36 countries and two residual geographical areas in the all-sciences case (Albarrán et al., 2015) while the second case refers to 219 WoS subfields (Crespo, et al., 2014).

Two comments are in order. First, the effect on overall citation inequality due to citation impact differences between the 500 LR universities (3.85%) is comparable to the effect due to citation impact differences between countries (5.4%). However, both of them are considerably smaller than is the corresponding effect on overall citation inequality attributable to differences in production and citation practices across the 219 subfields (~18%). Second, the reduction of the total effect generated by MNCS normalization in our data set (81.3% of the total effect) is of a comparable order of magnitude to the same phenomenon in the context of country (85.2%) or subfield citation distributions (83.2%).

Note that these results summarize in a pair of scalars a complex phenomenon that takes place along the entire support of our university citation distributions. As a matter of fact, the term *IDCU* is simply a weighted average of the I(π) terms, π = 1, . . ., 100, which capture the effect on overall inequality of the citation impact differences between the 500 LR universities holding constant the degree of excellence in all universities at percentile π. Therefore, it is instructive to study how I(π) changes with π both before and after the MNCS normalization. The results appear in Figure 3 [since I(π) is very high for π < 27, for clarity, these percentiles are omitted from Figure 3].

Figure 3 warrants two comments. First, the strong impact of MNCS normalization is readily apparent. Second, it is useful to informally partition the support into the following three intervals: [0, 57], [58, 96], and [98, 100]. In the first and the third intervals, I(π) values are very high. This means

that since in these two intervals university citation distributions differ by more than a scale factor, the universality condition can hardly be satisfied in them. However, I(π) is approximately constant for a wide range of intermediate values in the second interval. Thus, this is the range of values where the search for a single functional form—as in Chatterjee et al. (2014)—may give good results in our data set.

Discussion

Our results have two types of practical implications. In the first type, assume for example, that the top, intermediate, and weakest universities have different types of citation distributions or that the European, the North American, and the rest of the world universities have different types of citation distributions. In these two cases, we would need to build different models to explain the citation impact variability within the universities of the three types. On the contrary, since we have found that—although not universal—university citation distributions are rather similar, we need a single model to explain the high within-universities variability.

In the second type, recall the move in the CWTS and SCImago rankings from an average-based citation impact indicator—such as the MNCS—toward a rank percentile approach that throws all the weight on the top x% of most cited papers, such as the T indicator. This shift in emphasis is surely due to the idea that for highly skewed citation distributions, average-based indicators might not represent well the excellence in citation impact.[9] However, we should ask "Under what conditions will this move yield a ranking of

---

[9]Among other authors, one of us is on the record as advocating this idea (Albarrán, Ortuño, & Ruiz-Castillo, 2011b).
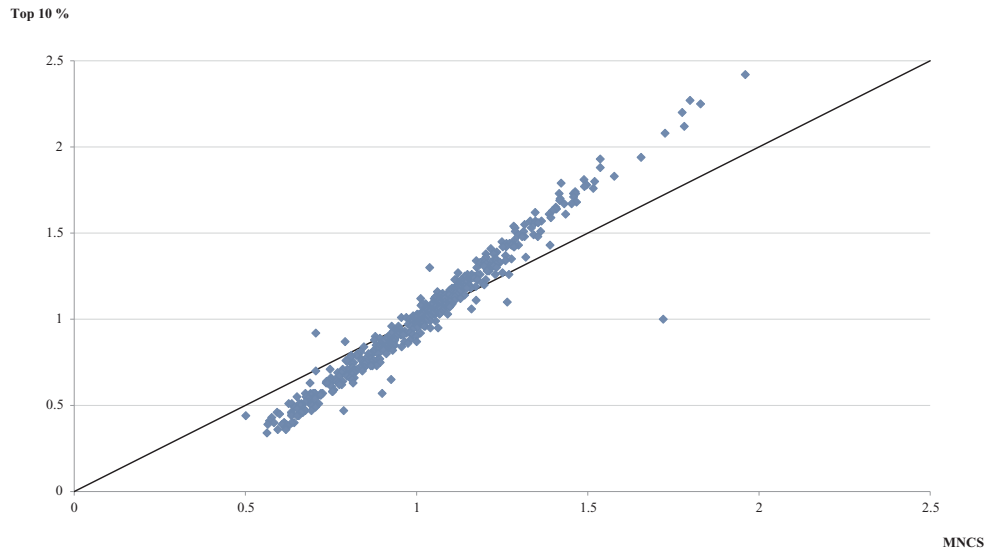
Top 10 %

MNCS

FIG. 4. Scatterplot of the relation between the MNCS indicator and the PPtop 10% indicator for the 500 Leiden Ranking universities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE 5. University ranking differences.

	From the WoS classification system to granularity level 8[a]			From T_i to T^*_i[b]			From $MNCS_i$ to $PP_{top10\%}$		
	First 100 universities (1)	Remaining 400 universities (2)	Total = (1) + (2) (3)	First 100 universities (4)	Remaining 400 universities (5)	Total = (1) + (2) (6)	First 100 universities (7)	Remaining 400 universities (8)	Total = (1) + (2) (9)
>50 positions	0	81	81	0	12	12	1	18	19
26–50	7	107	114	1	54	55	4	55	59
16–25	13	74	87	2	68	70	9	78	87
6–15	36	81	117	36	135	171	32	168	200
≤5 positions	44	57	101	61	131	192	54	81	135
Total	100	400	500	100	400	500	100	400	500

Note. [a]Table 6A in Ruiz-Castillo and Waltman (2015). [b]Table 3A in Perianes-Rodriguez and Ruiz-Castillo (2014).

research units different from a ranking according to average-based indicators?" Of course, the skewness of citation distributions is not a sufficient condition. If any pair of citation distributions under comparison were to differ only by a scale factor over their entire domain, or in other words, if the universality condition were to be satisfied, then the ratio of their MNCS values would coincide with the ratio of their T values for all choices of x. Therefore, for the two rankings to differ, we need the research units' citation distributions to be sufficiently different.

However, the two rankings are rather similar. The Pearson correlation coefficient between university values is 0.981 while the Spearman correlation coefficient between ranks is 0.986. Note that high correlations between university values and ranks do not preclude important differences for individual universities. As a matter of fact, the positive slope in Figure 4 indicates that the T index entails larger cardinal differences between universities: to low (high) MNCS values, there correspond lower (higher) T values.

Table 5 informs about the re-rankings that take place in the move from the MNCS to the T while Table 6 compares the differences between the university values themselves. In both cases, there are two instances with which to compare our results: (a) the relatively large differences between the university rankings according to the T indicator in going from the WoS classification system with 236 subfields to the classification system with 5,119 clusters we use in this article (Ruiz-Castillo & Waltman, 2015); and (b) the small differences between two ways of solving the all-sciences aggregation problem using the Top 10% indicator, with and without prior field normalization using our own data set (Perianes-Rodriguez & Ruiz-Castillo, 2014). According to the notation used in this article, these two indicators are denoted by T and T^*, respectively, in Tables 5 and 6.

On one hand, as anticipated in view of Figure 4, cardinal differences in our case are relatively large: The percentage of differences greater than 0.10 is 42%, a figure greater than 2.6% in Perianes-Rodriguez and Ruiz-Castillo (2014), and 19% in

TABLE 6.   University differences in values.

	From the WoS classification system to granularity level 8[a]			From T_i to T^*_i[b]			From $MNCS_i$ to $PP_{top10\%}$		
	First 100 universities (1)	Remaining 400 universities (2)	Total = (1) + (2) (3)	First 100 universities (4)	Remaining 400 universities (5)	Total = (1) + (2) (6)	First 100 universities (7)	Remaining 400 universities (8)	Total = (1) + (2) (9)
>0.20	1	16	17	0	0	0	44	26	70
>0.10 and 0.2	12	66	78	4	9	13	40	100	140
>0.05 and 0.10	27	124	151	25	70	95	12	106	118
≤0.05	60	94	254	71	321	392	4	168	172
Total	100	400	500	100	400	500	100	400	500

Note. [a]Table 6B in Ruiz-Castillo and Waltman (2014).[b]Table 3B in Perianes-Rodriguez and Ruiz-Castillo (2014).

Ruiz-Castillo and Waltman (2015)—compare columns 9, 6, and 3 in Table 6. On the other hand, the percentage of universities experiencing relatively large re-rankings greater than 25 positions is 15.6%—a figure similar to 13.4%, and well below 39% in the other two contributions—compare columns 3, 6, and 9 in Table 5. Therefore, ordinal differences between the university rankings according to the MNCS and the T indicators are of a small order of magnitude. We actually find a strong, more or less linear relationship between the T and the MNCS in two other instances: for the 500 universities in the 2011/2012 edition of the LR (see Figure 2 in Waltman et al., 2012a) and for the partition of the world into 39 countries and eight geographical areas studied in Albarrán and Ruiz-Castillo (2012).

How can we explain these results? We already have seen that university citation distributions behave as if they differ by a relatively constant scale factor over the [58, 96] percentile interval in their support. In this empirical scenario, it is not surprising that the MNCS values, which are reached at approximately the 63th percentile of citation distributions, and the T indicator that focuses on the last 10 percentiles, provide very similar rankings. A convenient, practical consequence is that the citation impact university ranking provided by the MNCS indicator is an adequate one for most universities. The T indicator would only add greater cardinal differences between the best and the worse universities, with relatively few re-rankings.

Conclusions and Further Research

Conclusions

This article has investigated the citation distributions of the 500 universities in the 2013 edition of the CWTS LR. We have used a WoS data set consisting of 3.6 million articles published in 2003 to 2008 with a 5-year citation window and classified into 5,119 clusters. The all-sciences aggregation problem is solved by using the standard field-normalization procedure, where clusters mean citations are used as normalization factors. The assignment of responsibility of publications to universities in the presence of coauthorship is solved by applying a fractional approach. The main findings can be summarized in the following four points.

- The universality claim, according to which all university citation distributions, appropriately normalized, follow a single functional form, is not supported by the data.
- Nevertheless, the 500 university citation distributions are all highly skewed and very similar. This result is essentially maintained regardless of the way we solve the all-sciences aggregation problem and the assignment of publications to universities in the presence of coauthorship.
- Citation impact differences between universities account for 3.85% of overall citation inequality. However, these differences are greatly reduced when university citation distributions are normalized using their MNCS values as normalization factors.
- The aforementioned results have important practical consequences. First, we only need a single explanatory model for the type of high skewness characterizing all university citation distributions. Second, the similarity of university citation distributions goes a long way in explaining the similarity of the university rankings obtained with the MNCS and the T indicator.

Naturally, the robustness of these results must be investigated with other data sets characterized by other publication years, other citation windows, and other classification systems, as well as other data sources different from the WoS.

Further Research

We provide four possibilities for further research.

- The effect on overall citation inequality attributable to the differences in citation impact between universities shows a characteristic pattern: University citation distributions appear to behave as if they differ by a relatively constant scale factor over a large, intermediate part of their support. Consequently, it might be interesting to compute the exchange rates introduced in Crespo et al. (2014) and Crespo et al., (2013) to exploit this feature and to use them as normalization factors. More generally, one could experiment with other normalization approaches that have been found most useful in other contexts, notably the two parameter scheme introduced by Radicchi and Castellano (2012).
- Chatterjee et al.'s (2014) idea of fitting specific functional forms to university citation distributions in different intervals of their support is worth pursuing. The threshold determining

the upper tail where a power law might be the best alternative could be estimated following the methods advocated in Clauset, Shalizi, and Newman (2009). Similar grid techniques could be applied to determine the lower bound of the interval where a lognormal might be the best alternative. In any case, standard methods should be used to test which specific functional form is best in each interval as well as to estimate the parameters' confidence intervals (Brzezinski, 2015; Clauset et al., 2009, Thelwall & Wilson, 2014).

- As was seen earlier, differences in citation impact between universities after MNCS normalization tend to increase when we reach the last few percentiles, including the most highly cited articles. The question left for further research is how to complement average-based or percentile rank indicators with other measurement instruments that highlight the behavior of citation distributions over the last few percentiles. Given the important role of extreme observations in citation distributions, robustness of alternative high-impact indicators to these extreme situations will be an important element in the discussion.

- Consider an array of citation distributions with a smaller number of scientific fields than the number of clusters in this article in the columns, and the 500 LR universities in the rows. We already know a good deal concerning field citation distributions and university citation distributions in the all-sciences case. A possible next step is to study the characteristics of university citation distributions column by column; that is, restricted to each field. The results will determine to what extent the similarities between citation distributions is a question depending on the aggregation level at which the study is conducted.

Acknowledgments

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., & Ruiz-Castillo, J. (2012). The Measurement of Scientific Excellence Around the World. Working Paper, Economic Series 12-08, Universidad Carlos III Retrieved from http://e-archivo.uc3m.es/handle/10016/13896

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

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011b). The measurement of low- and high-impact in citation distributions: Technical results. Journal of Informetrics, 5, 48–63.

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.

Brzezinski, M. (2015). Power laws in citation distributions: Evidence from Scopus. Scientometrics, 103, 213–228.

Chatterjee, A., Ghosh, A., & Chakrabarty, B.K. (2014). Universality of citation distributions for academic institutions and journals, posted September 29, ArXiv:1409.8029 [physics.soc-ph].

Clauset, A., Shalizi, C.R., & Newman, M.E.J. (2009). Power-law distributions in empirical data. SIAM Review, 51, 661–703.

Crespo, J.A., Li, Y., & 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, Y., & 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.

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

Herranz, N., & Ruiz-Castillo, J. (2012). Multiplicative and fractional strategies when journals are assigned to several sub-fields. Journal of the American Society for Information Science and Technology, 63, 2195–2205.

Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., & de Nooye, W. (2012). Field-normalized impact factors: A comparison of rescaling versus fractionally counted ifs. Journal of the American Society for Information Science and Technology, 27, 292–306.

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

Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2014). An alternative to field-normalization in the aggregation of heterogeneous scientific fields. Working Paper 14-25, Departamento de Economía, Universidad Carlos III Retrieved from http://e-archivo.uc3m.es/handle/10016/19812

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

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 (2013). Journal of Informetrics, 8, 25–28.

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

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). Subject field characteristic citation scores and scales for assessing research performance. Scientometrics, 12, 267–291.

Thelwall, M., & Wilson, P. (2014). Distributions for cited articles from individual subjects and years. Journal of Informetrics, 8, 824–839.

Waltman, L., & Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. Journal of the American Society for Information Science and Technology, 63, 2378–2392.

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

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., Van Eck, N.J., ... Wouters, P. (2012a). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. Journal of the American Society for Information Science and Technology, 63, 2419–2432.

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