UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper
Economic Series 12 – 29
November 2013

Departamento de Economía
Universidad Carlos III de Madrid
C/ Madrid, 126, 28903 Getafe (Spain)

"DIFFERENCES IN CITATION IMPACT ACROSS COUNTRIES"

Pedro Albarrán[a], Antonio Perianes-Rodríguez[b], and Javier Ruiz-Castillo[c]

[a] Departamento de Fundamentos del Análisis Económico, Universidad de Alicante
[b] Departamento de Biblioteconomía y Documentación, Universidad Carlos III, SCImago Research Group
[c] Departamento de Economía, Universidad Carlos III

Abstract

Using a large dataset, indexed by Thomson Reuters, consisting of 4.4 million articles published in 1998-2003 with a five-year citation window for each year, this paper studies country citation distributions in a partition of the world into 36 countries and two geographical areas in the all-sciences case and eight broad scientific fields. The key findings are the following two. Firstly, the shape of country citation distributions is highly skewed and very similar to each other across all fields. Secondly, differences in country citation distributions appear to have a strong scale factor component. The implication is that, in spite of the skewness of citation distributions, international comparisons of citation impact in terms of country mean citations capture well such scale factors. The empirical scenario described in the paper helps understanding why, in each field and the all-sciences case, the country rankings according to (i) mean citations and (ii) the percentage of articles in each country belonging to the set formed by the 10% of the more highly cited papers are so similar to each other.

Keywords: citation impact, citation inequality, country rankings, country normalization

Acknowledgements

This is the second version of a paper with the same title that appeared in this series in December 2012. The authors acknowledge financial support by Santander Universities Global Division of *Banco Santander*. Albarrán acknowledges additional financial support from the Spanish MEC through grants ECO2009-11165 and ECO2011-29751, and Ruiz-Castillo through grant SEJ2007-67436. Conversations with Ludo Waltman are greatly appreciated. All remaining shortcomings are the authors' sole responsibility.

I. INTRODUCTION

Among other factors, differences in the average number of authors per paper, the average paper length, the average number of papers per author in a given period of time, the average number of references per paper, the percentage of internationally co-authored papers, or the speed at which the citation process evolves generate large differences between citation distributions across scientific disciplines. However, using large datasets indexed by Thomson Reuters, recent research has systematically established the following two fundamental regularities for the 22 broad fields distinguished by Thomson Reuters, and the 219 sub-fields identified with Web of Science subject-categories. Firstly, broadly speaking, field and sub-field citation distributions share the same (highly skewed) shape (Albarrán and Ruiz-Castillo, 2011, and Albarrán *et al.*, 2011) Secondly, differences in publication and citation practices generate scale and other differences between citation distributions. However, in so far as scale differences appear to be very important and are well captured by mean citations, the traditional procedure of taking mean citations as normalization factors works very well in practice (Crespo *et al.*, 2013a, b, Radicchi and Castellano, 2012a, b, and Li *et al.*, 2013).[1]

This paper searches for similar regularities among country citation distributions in eight broad fields –Physics, Chemistry, Clinical Medicine, Biology & Biochemistry, Materials Science, Geosciences, Engineering, and Economics & Business– and in the all-sciences case. We consider a partition of the world into 36 countries and two residual geographical areas using a Thomson Reuters dataset consisting of the 4.4 million articles published in 1998-2003, and the citations they receive during a five-year citation window for each year in that period. In principle, differences in resources, intellectual traditions, organization, the structure of incentives, science policy and many other factors lead us to expect large differences between country citation distributions in any science. We analyze the following three aspects of the question.

[1] This does not preclude the interest of other normalization procedures, such as the two-parameter reverse engineering procedure introduced by Radicchi and Castellano (2012a), or a variety of source, or citing side procedures (see *inter alia* Waltman and van Eck, 2013).

Firstly, we abstract from size and mean citation differences and focus on the shape of citation distributions. As in Albarrán and Ruiz-Castillo (2011), and Albarrán *et al.* (2011), we use for this purpose a scale- and size-independent statistical technique, the Characteristic Scores and Scales (CSS hereafter) first introduced in scientometrics by Schubert *et al.* (1987). Moreover, we use the robust index of skewness suggested by Groeneveld and Meeden (1984). We find that although country citation distributions are highly skewed, this skewness is somewhat different across countries in each field.

Secondly, it should be emphasized that what matters is how important these differences in skewness within each field really are. Using a measuring framework first introduced in Crespo *et al.* (2013a), we quantify in each field the effect on citation inequality of differences of all sorts between country citation distributions. Using an additively decomposable citation inequality index, this effect is seen to be well captured by a between-group term in a certain partition by fields and quantiles (see Crespo *et al.*, 2013a, b, Waltman and Van Eck, 2013, and Li *et al.*, 2013). In this paper we apply this method to estimate a *IDCC* term that measures the effect on citation *I*nequality of *D*ifferences in *C*itation impact across *C*ountries.

Thirdly, we study by how much is the *IDCC* term reduced when we normalize the raw citation counts using country mean citations (MC hereafter) as normalization factors. As an alternative, we also study the consequences of normalization using the exchange rates concept introduced in Crespo *et al.* (2013a) as normalization factors.

Articles are assigned to countries according to the institutional affiliation of their authors on the basis of what had been indicated in the by-line of the publications. We must confront the technical difficulty posed by international cooperation, namely, the existence of articles written by authors belonging to two or more countries. The problem, of course, is that international articles as opposed to, say domestic articles, tend to be highly cited. Although this old question admits different solutions (see *inter alia* Anderson *et al.*, 1988, for a discussion), in this paper we focus on a *multiplicative* strategy according to which in every internationally co-authored article a whole count is credited to each contributing country.[2] However, in two important instances –Physics, and the all-sciences case– we find that our results are robust to a *fractional* strategy where each international article is fractioned into as many pieces as countries appear among its authors.

The rest of the paper is organized in four Sections. Section II introduces some notation, describes the data, and presents the results concerning the shape of citation distributions. Section III has three aims. Firstly, it summarizes the method for quantifying the effect on overall citation inequality of differences in citation impact across countries under a multiplicative strategy, and presents the results about its importance as well as the consequences of country MC normalization. Secondly, it studies how this effect varies along the support of citation distributions, and explores the distinction between domestic and international articles. Thirdly, it reports on two extensions concerning the use of exchange rates as normalization factors, and the consequences of adopting a fractional approach to the treatment of international articles. Finally, Section IV summarizes and discusses the results, and offers some suggestions for possible extensions.

II. THE SIMILARITY OF THE SHAPE OF COUNTRY CITATION DISTRIBUTIONS

II. 1. Notation in the Multiplicative Approach

Consider a certain scientific field, say Physics, consisting of N distinct articles, indexed by $l = 1,\ldots, N$. Let $\boldsymbol{Q} = (c_1,\ldots, c_l, \ldots, c_N)$ be the initial citation distribution, where c_l is the number of citations received by article l. The total number of citations is denoted by $\gamma = \Sigma_l c_l$. Assume that there are P countries, indexed by $p = 1,\ldots, P$. For any l, let X_l be the non-empty set of countries to which the author(s) of article l belongs to, and let x_l be the cardinal of this set, i. e. $x_l = |X_l|$. Since, at most, an article can be written by authors in P countries, we have that $x_l \in [1, P]$.

[2] Among the many contributions that follow a multiplicative strategy, see May (1997), and King (2004), as well as the references in Section II in Albarrán *et al.* (2010). On the other hand, see Aksnes *et al.* (2012) for a recent defense of the fractional strategy.

Let N_p be the total number of distinct articles in p, indexed by $i = 1,\ldots, N_p$. In the multiplicative approach to international co-authorship, country p's ordered citation distribution can be described by $\boldsymbol{C_p} = (c_{p1}, \ldots, c_{pi}, \ldots, c_{pNp})$, where $c_{pi} = c_l$ for some article l in the initial distribution Q, and $c_{p1} \leq c_{p2} \leq \ldots \leq c_{pNp}$. What we call the *geographical extended count* is simply the union of these distributions, $\boldsymbol{C} = \cup_p \boldsymbol{C_p}$, whose total number of articles is $M = \Sigma_p N_p = \Sigma_l x_l$. Only domestic articles, or articles exclusively authored by one or more scientists affiliated to research centers in a single country are counted once, in which case $x_l = 1$. Otherwise, $x_l \in [2, P]$. As long as $x_l > 1$ for some l, we have that $M > N$.

II. 2. The Data

Since we wish to address a homogeneous population, in this paper only research articles or, simply, articles are studied. As indicated in the Introduction, we begin with a large sample, consisting of more than 4.4 million articles published in 1998-2003, as well as the citations these articles receive using a common, five-year citation window for each year in that period. Table A in the Appendix I presents the number of articles by field in the original and the geographically extended count, as well as the MC in each of the 22 fields. The number of distinct articles in the original dataset is $N = 4{,}472{,}332$, while the number of articles in the geographical extended count is $M = 5{,}450{,}309$, a total which is 21.9% larger than N. In turn, the number of distinct articles in Physics, for example, is $N_P = 456{,}144$, while the number of articles in the corresponding geographically extended count is 626,304, a total which is 37.3% larger than N_P. It should be noted that the two distributions of the number of articles by field are very similar. In the extended count there are 1.5% more articles in the Physical Sciences, and slightly less in the Life and the Social Sciences. On the other hand, as expected, MCs are always greater in the geographical extended count reflecting the fact that internationally co-authored articles tend to be more cited than the domestic ones.

We consider 36 countries and two residual geographical areas, described in Table B in Appendix I, that have published at least 10,000 articles in all sciences in 1998-2003. To save space, this Table includes the number of articles by country, both in the original dataset and the extended count, only in Physics and the all-fields case. In the latter, the U.S. publishes about 27% of the total under the multiplicative approach, while the EU, namely, the 15 countries forming the European Union before the 2004 accession, is responsible for approximately one third. The remaining 23 countries and the two geographical areas publish almost 39% of the total. In the original dataset, these three areas publish 29%, 34%, and 37% of the total. In Physics, the U.S., and the EU publish 18%, and 37% of the total in the extended count, and 19%, and 33% of the total in the original dataset. For later reference, Table C in Appendix I includes the number of articles per country only in the extended count in the remaining seven fields.

Given the wide differences in publication and citation practices, in scientometrics is customary to proceed to some normalization before aggregating all fields into what we call the all-sciences case. Recent results indicate that, among target or cited-side normalization procedures, the standard practice of using field MCs as normalization factors generates good results (Radicchi and Castellano, 2012a, b, Leydesdorff *et al.*, 2012, Crespo *et al.*, 2013, a, b, and Li *et al.*, 2013). Furthermore, Li and Ruiz-Castillo (2013) establish that the ability of the normalization at the field level to reduce the problem of differences in citation practices works well even at the sub-field level. Therefore, in the sequel all references to the all-sciences case take place after the standard field normalization.

II. 3. Characteristics of Country Citation Distributions

It is important to know whether or not country citation distributions present the fundamental features that have been appreciated for entire broad fields, sub-fields, and the all-sciences case under both the fractional and the multiplicative approaches to the problem of the assignment of articles to Web of Science subject-categories (Albarrán and Ruiz-Castillo, 2010, Albarrán *et al.*, 2011, and Herranz and Ruiz-Castillo, 2011). For this purpose, we use the CSS technique, a scale- and size invariant statistical method that allows us to focus on the shape of citation distributions.

The following *characteristic scores* are determined: μ_1 = mean citation for the entire distribution; μ_2 = mean citation for articles with citations above μ_1, and μ_3 = mean citation for articles with citations above μ_2. Consider the partition of the distribution into five broad classes: (i) articles with no citations; (ii) articles with few citations below μ_1; (iii) fairly cited articles, with citations above μ_1 and below μ_2; (iv) articles with a remarkable number of citations above μ_2 and below μ_3, and (v) articles with an outstanding number of citations above μ_3. Figures 1 to 9 illustrate the partition of citation distributions into the five classes for all countries and geographical areas in the eight fields and the all-sciences case (numerical results are available upon request). In each case, countries are ordered in terms of their MC.

Two points should be emphasized. Firstly, note that there are large differences in the percentage of uncited articles across countries. However, there is a strong negative correlation between the first two classes of articles, so that, on average, all countries have a very similar percentage of articles below μ_1. To illustrate this point, we have drawn a first vertical line in Figures 1 to 9 at the average over all countries and geographical areas of the percentage of articles in the sum of classes 1 and 2. Visually, it is clear that for most countries this percentage is not very far from the average. Secondly, something similar takes place at the upper tail of citation distributions: in spite of noticeable differences in, say, the percentage of articles in class 5, the percentage of articles in the sum of classes 4 and 5 are quite similar. To illustrate this point, we have drawn a second vertical line in Figures 1 to 9 at the average over all countries of the percentage of articles in the sum of classes 4 and 5. Again, it is clear that for most countries this percentage is not very far from the average.

These results suggest focusing the attention on a partition of citation distributions into three broad categories: the sum of classes 1 plus 2; class 3, and the sum of classes 4 and 5, which will be denoted as categories I, II, and III, respectively. Figures 1 to 3 have illustrated that country citation distributions share some stylized basic features in each of the selected cases. More precisely, Table 1 includes the average and standard deviation over all countries and geographical areas in each case for the percentage of articles in the three classes, as well as the corresponding statistics for the percentages of the total number of citations accounted for by each class.

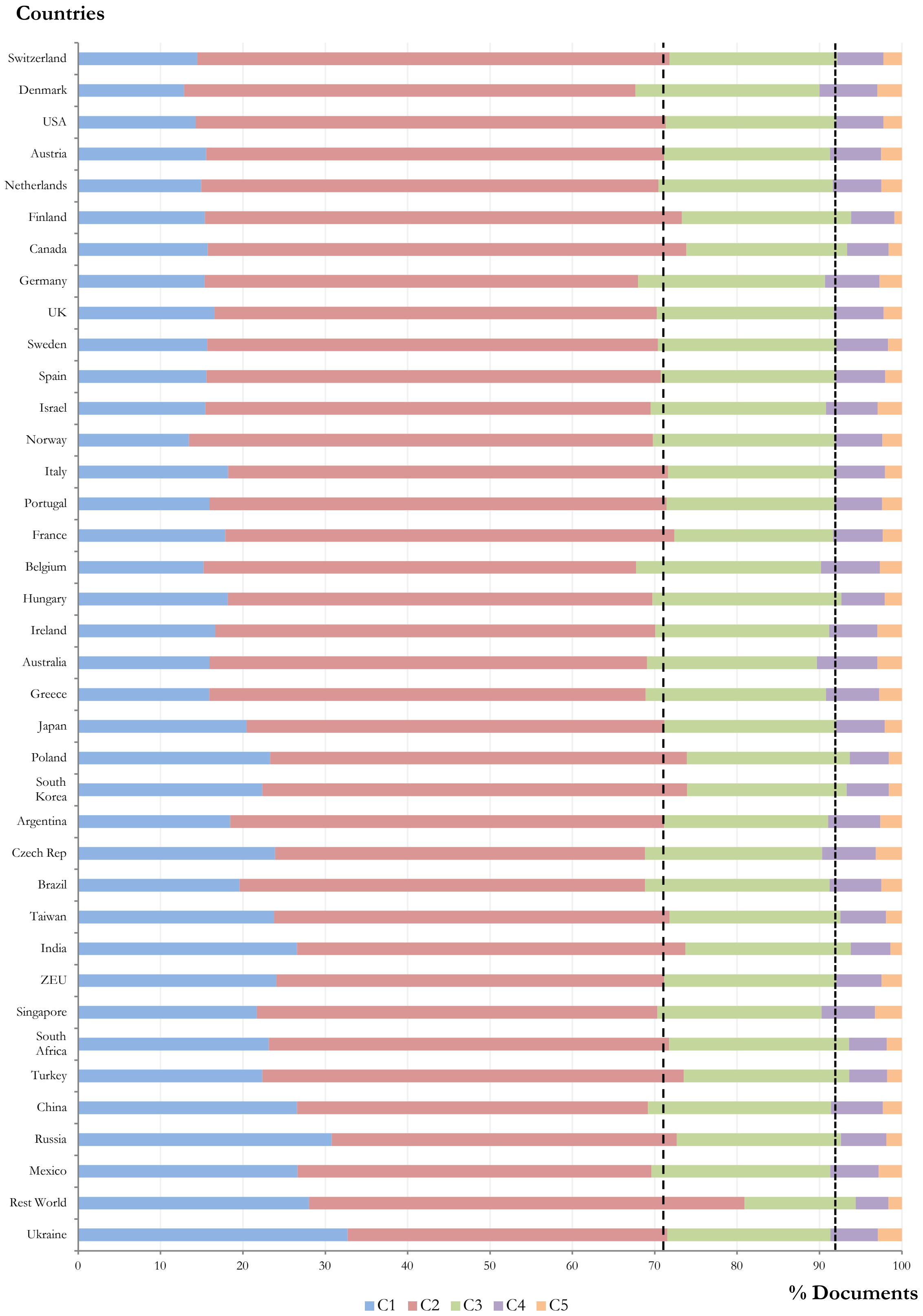

Figure 1. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Physics

Countries

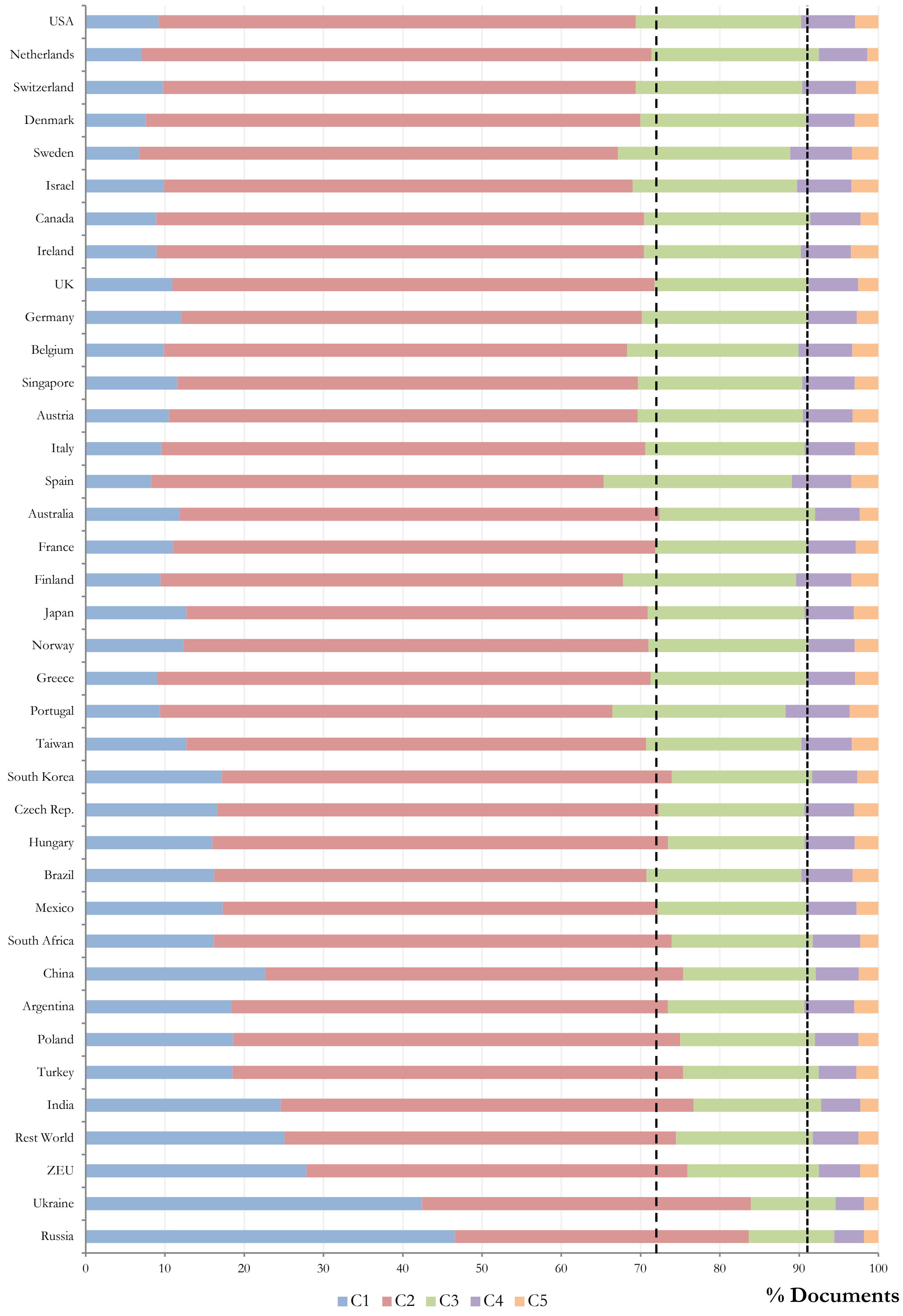

Figure 2. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Chemistry

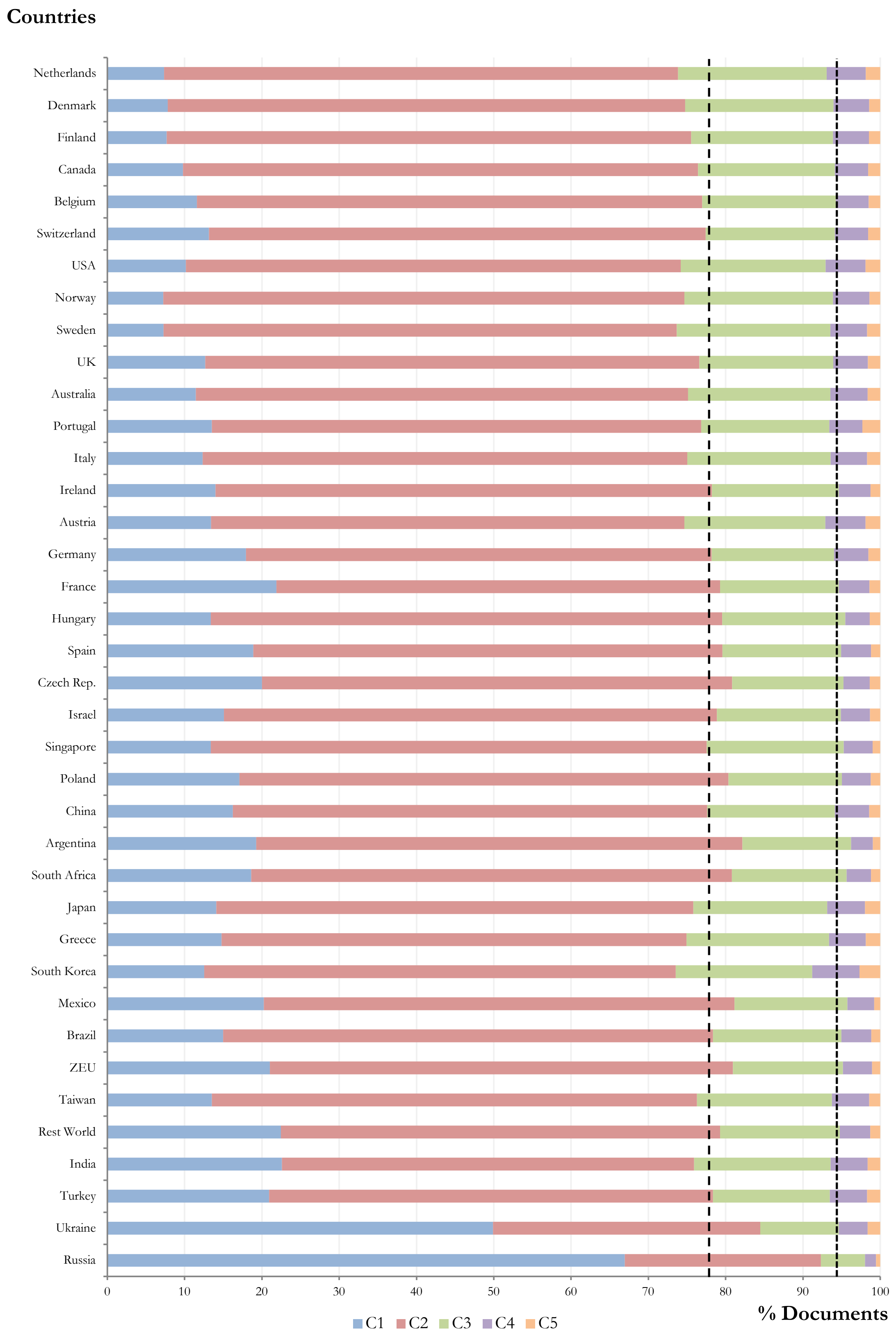

Figure 3. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Clinical Medicine

Countries

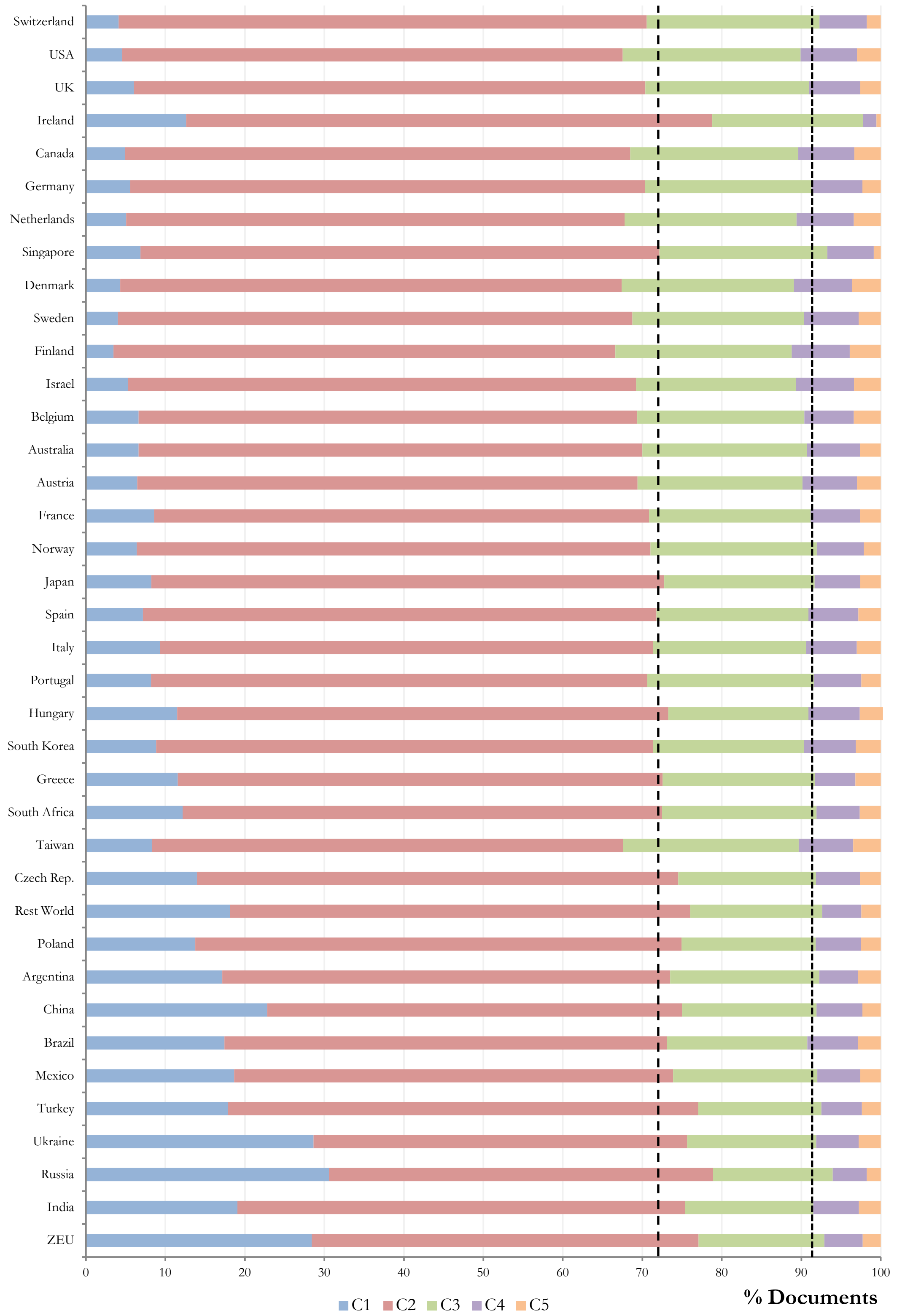

Figure 4. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Biology & Biochemistry

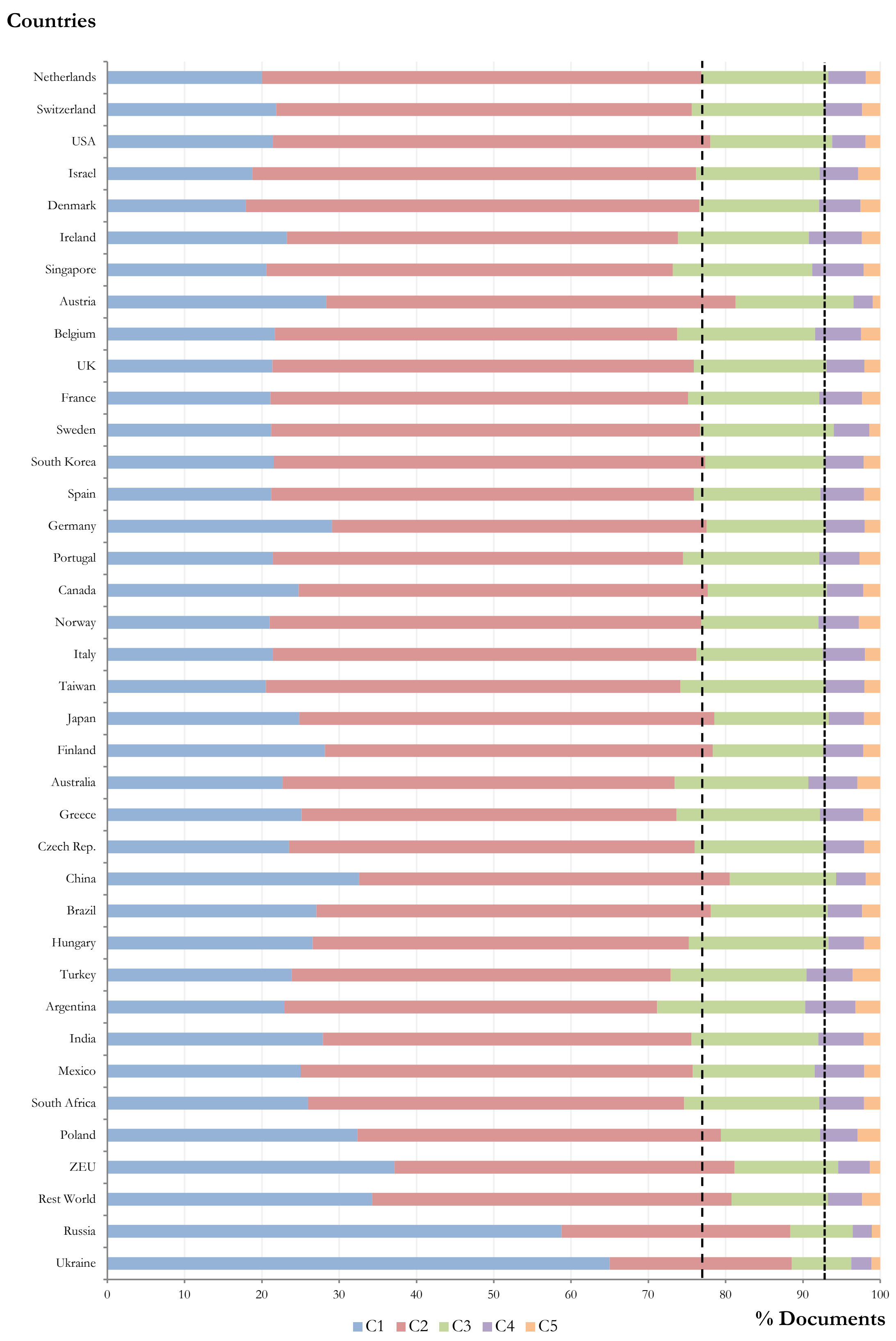

Figure 5. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Materials Science

Countries

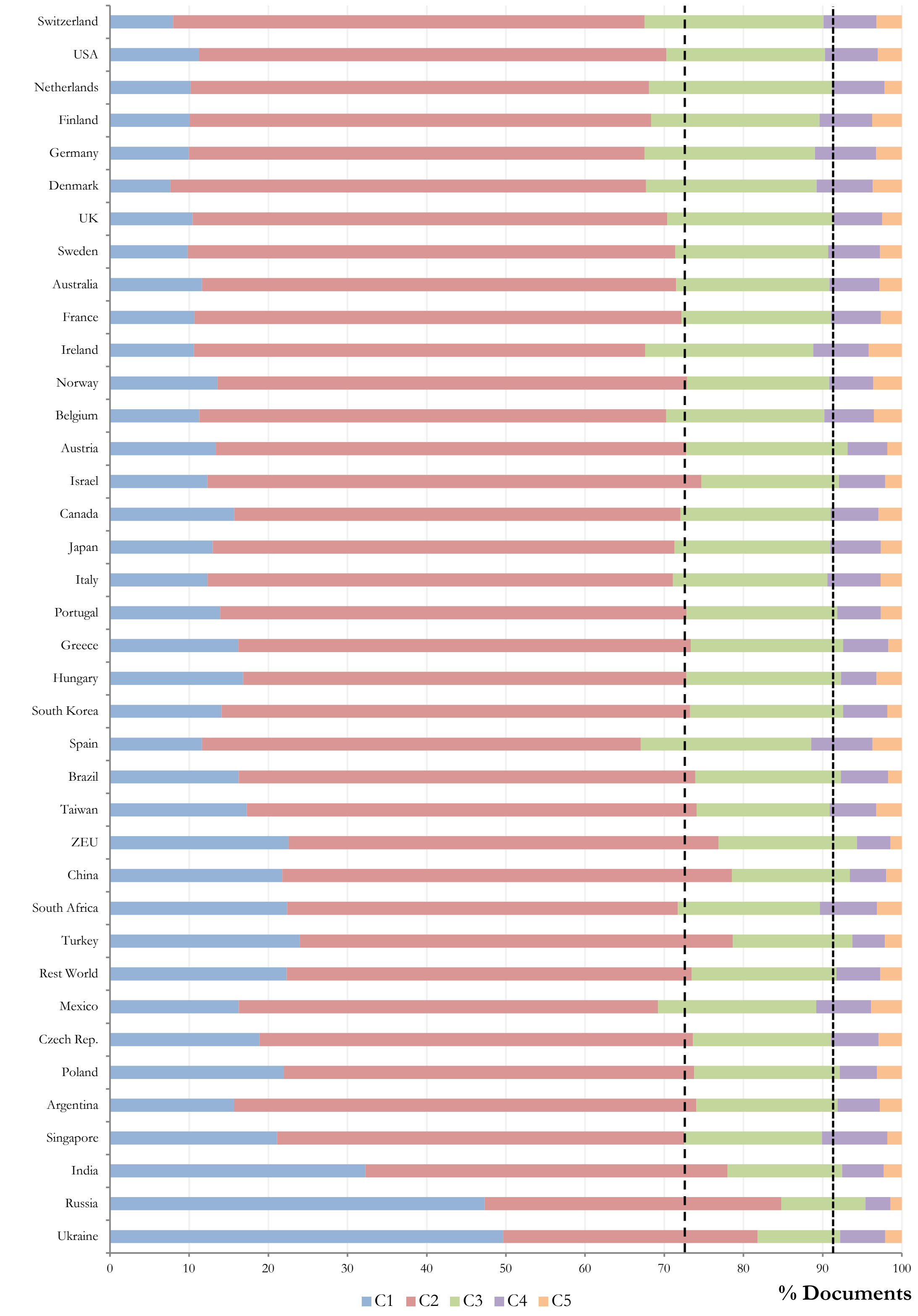

Figure 6. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Geosciences

Countries

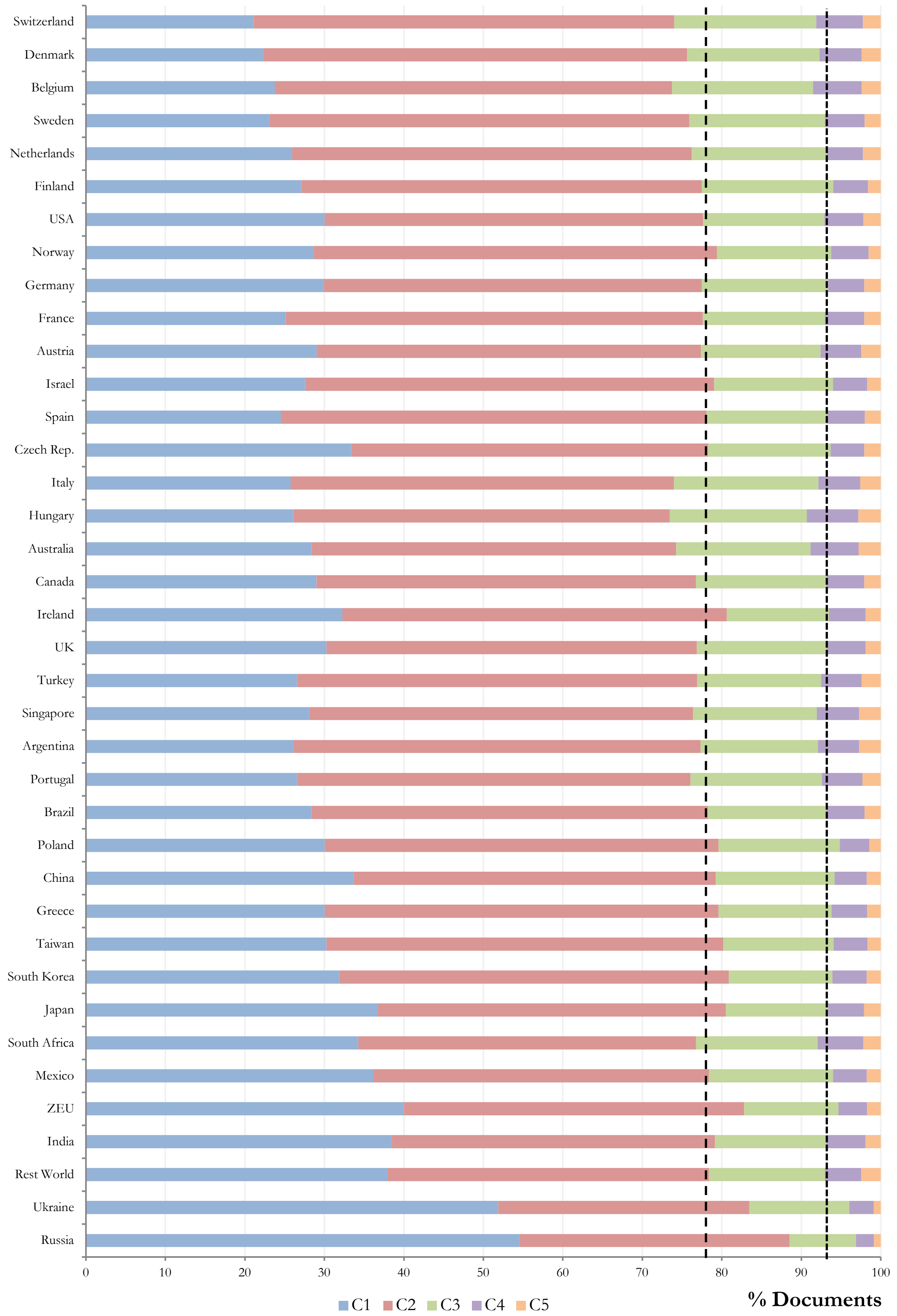

Figure 7. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Engineering

Countries

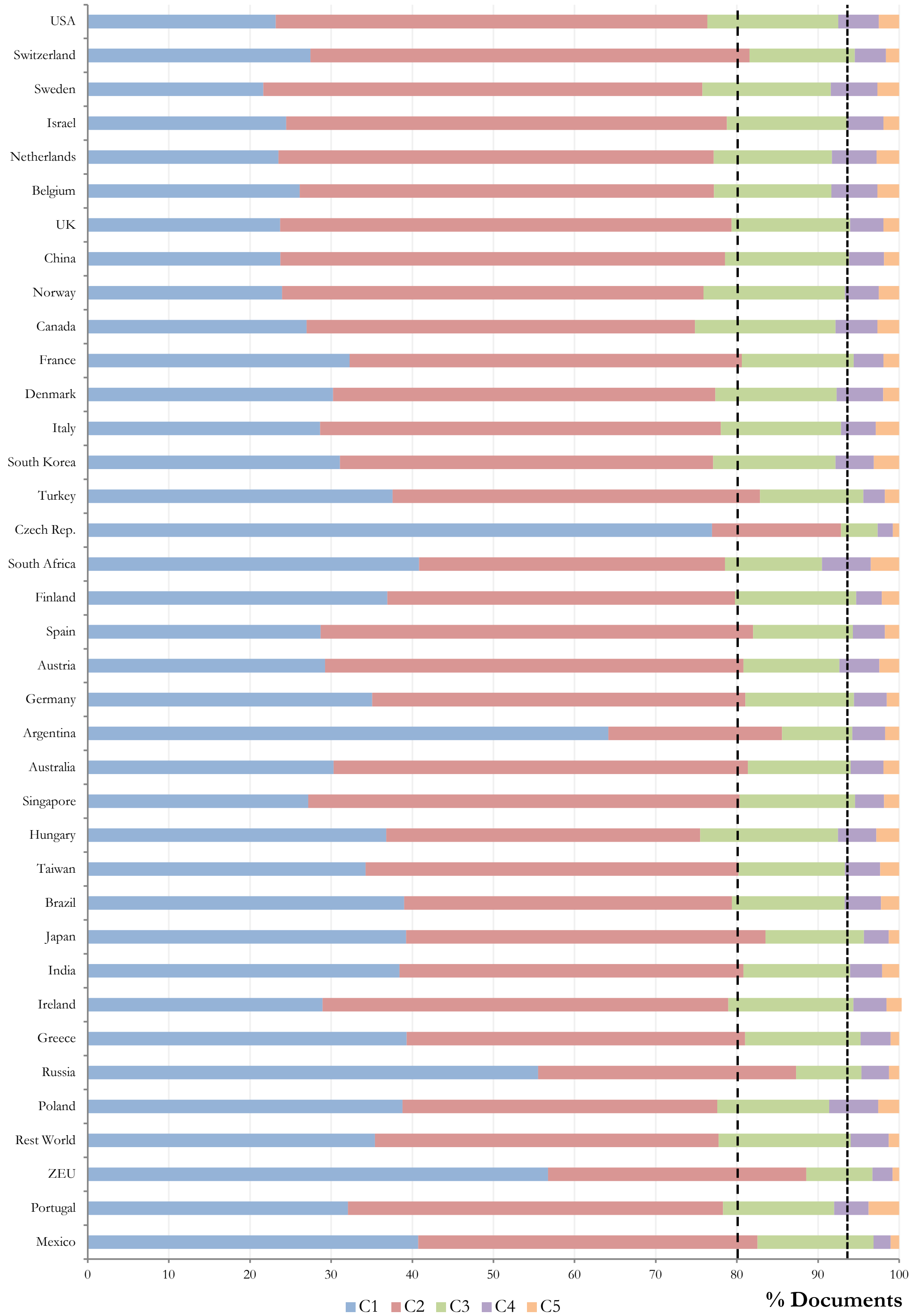

Figure 8. Partition of Citations Distributions Into Five Classes According to the CSS Technique. Economics & Business

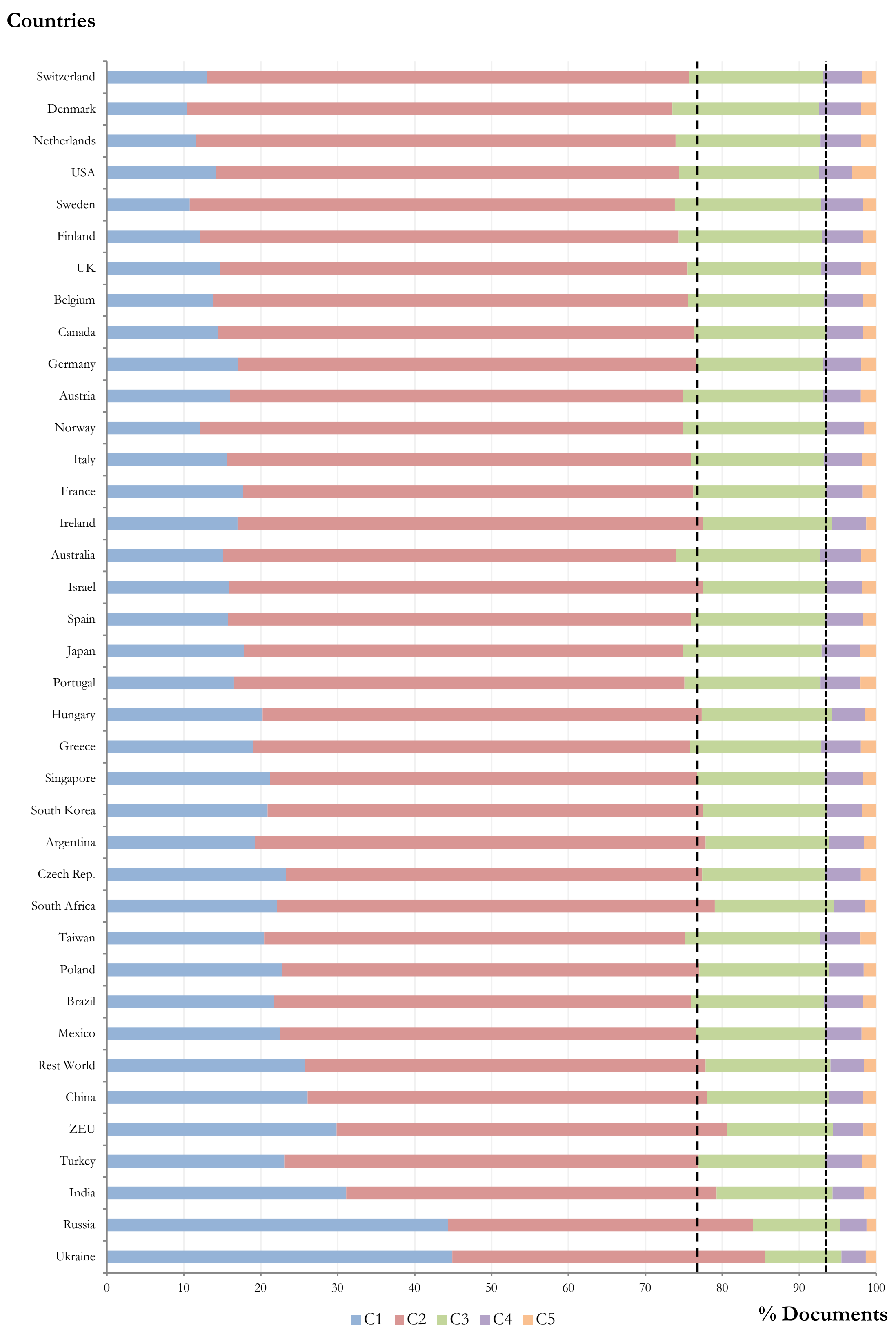

Figure 9. Partition of Citations Distributions Into Five Classes According to the CSS Technique. All-sciences Case

Table 1. The Skewness of Country Citation Distributions in Eight Fields and the All-fields Case. Average (Standard Deviation), and Coefficient of Variation over 38 Countries of the Percentages of Articles, and the Percentages of Total Citations by Category

	Percentage of Articles In Category:			Percentage of Total Citations Accounted For By Category:		
	1	**2**	**3**	**1**	**2**	**3**
PHYSICS	71.1 (2.3) 0.03	20.7 (1.5) 0.08	8.1 (1.2) 0.15	22.5 (2.3) 0.10	33.5 (1.1) 0.03	44.0 (2.2) 0.05
CHEMISTRY	72.0 (3.8) 0.05	19.1 (2.7) 0.14	9.0 (1.3) 0.15	29.7 (1.8) 0.06	32.3 (1.2) 0.04	38.0 (1.8) 0.05
CLINICAL MEDICINE	77.9 (3.6) 0.05	16.4 (2.6) 0.16	5.7 (1.2) 0.20	28.7 (1.6) 0.06	31.2 (1.0) 0.03	40.2 (1.5) 0.04
BIOLOGY & BIOCHEMISTRY	72.0 (3.2) 0.04	19.3 (2.1) 0.11	8.7 (1.6) 0.19	29.8 (1.5) 0.05	32.1 (1.4) 0.04	38.1 (1.5) 0.04
MATERIALS SCIENCE	77.0 (3.6) 0.05	15.8 (2.5) 0.16	7.2 (1.5) 0.20	30.2 (2.0) 0.07	31.4 (1.5) 0.05	38.5 (1.9) 0.05
GEOSCIENCES	72.7 (3.9) 0.05	18.6 (2.8) 0.15	8.7 (1.6) 0.18	29.8 (2.0) 0.07	32.2 (1.4) 0.04	38.0 (2.1) 0.06
ENGINEERING	78.1 (2.9) 0.04	15.2 (1.9) 0.12	6.8 (1.2) 0.18	31.0 (2.5) 0.08	31.9 (1.8) 0.06	37.1 (2.2) 0.06
ECONOMICS & BUSINESS	80.1 (3.8) 0.05	13.6 (2.7) 0.20	6.3 (1.6) 0.25	31.6 (3.4) 0.11	30.6 (3.0) 0.10	37.8 (3.9) 0.10
ALL SCIENCES	75.7 (2.6) 0.03	17.2 (1.9) 0.11	7.1 (0.8) 0.11	29.7 (0.7) 0.02	32.0 (0.5) 0.01	38.4 (0.8) 0.02

Category 1 = articles with a low number of citations, below μ_1
Category 2 = articles with a fair number of citations, above μ_1 and below μ_2
Category 3 = articles with a remarkable or outstanding number of citations, above μ_2

where:
μ_1 = mean citation of each country citation distribution
μ_2 = mean citation of articles with a number of citations above μ_1

The results are truly remarkable. A complex set of economic, sociological, political, and intellectual factors are influencing the research performance of each country in every field and, consequently, the shape of their citation distributions. The generally high percentages of articles in class I observed in Table 1 makes the variability over all countries of the small percentages in categories II and III to be relatively high in most fields. However, the small coefficients of variation in the percentage of articles in category I, as well as the percentage of citations accounted for by all categories indicate that country citation distributions in eight fields and the all-sciences case tend to share some fundamental characteristics. Barring some exceptional behavior in some small field, such as Economics & Business, we find that between 71% and 78% of all articles receive citations below the mean and account for, approximately, between 22% and 31% of all citations, while articles with a remarkable or outstanding number of citations represent about 6% or 9% of the total, and account for, approximately, between 37% and 44% of all citations. Thus, we can conclude that country citation distributions are all highly skewed in the sense that a large proportion of articles get no or few citations while a small percentage of them account for a disproportionate amount of all citations. These results closely resemble those concerning the shapes of citation distributions across a wide array of 219 sub-fields identified with the Web of Science subject categories distinguished by Thomson Reuters (as well as across a variety of categories at different aggregation levels).[3] However, judging from the size of coefficients of variation, citation distributions in that context appear to be more similar to each other.

We have also computed skewness indicators for all countries in all fields and the all-sciences case. The problem, of course, is that extreme observations with a very large number of citations are known to be prevalent in citation distributions (see Herranz and Ruiz-Castillo,

[3] In the multiplicative case, for example, approximately 69% of all articles receive citations below the mean and account for, at most, 21% of all citations, while articles with a remarkable or outstanding number of citations represent about 9% or 10% of the total, and account for approximately 44% of all citations (Albarrán *et al.*, 2011a, p. 391). For other aggregate levels, see Albarrán and Ruiz-Castillo (2010), and Li *et al.* (2013), while for the fractional case, see Herranz and Ruiz-Castillo (2012a).

2012a, and Li and Ruiz-Castillo, 2013b). This presents a challenge for conventional measures of skewness that are very sensitive to outliers. Fortunately, robust measures of skewness based on quartiles have been developed in the statistics literature (for a discussion in the context of the financial literature on stock market returns, see Kim and White, 2004). Here we use the measure suggested by Groeneveld and Meeden (1984) that improves upon the extension of Bowley's (1920) measure due to Hinkley (1975), and has better properties than the well known measure of Kendall and Stuart (1977).

Given a process $\{y_t\}$, $t = 1,\ldots, T$, where the y_t's are independent and identically distributed with a cumulative distribution function F, the conventional measure of skewness is

$$SK_1 = (1/T)\ \Sigma_t\ [(y_t - \mu)/\sigma]^3,$$

where μ is the mean, and σ is the variance, while the Groeneveld and Meeden robust measure is

$$SK_2 = (\mu - Q_2)/\mathrm{E}\left| y_t - Q_2 \right|,$$

where $Q_2 = \mathrm{F}^{-1}(0.5)$ is the second quartile of y_t. The results for SK_1 and SK_2 are in Table D in the Appendix. As expected, the presence of extreme observations cause SK_1 to take often very high values. In this context, the index SK_2 performs much better. Consequently, Table 2 reports the average, standard deviation, and coefficient of variation of SK_2 for all countries in our eight fields and the all-sciences case.

Taking into account that SK_2 is bounded in the interval [-1, 1], it is clear that on average country citation distributions in all fields –but a little less so in the all-sciences case– are highly skewed. Relative high standard deviations and coefficients of variation indicate that country citation distributions within most fields differ from each other in the extent of skewness measured by the Groeneveld and Meeden index. Interestingly, the average skewness across the 22 scientific fields distinguished by Thomson Reuters is 0.58, and the coefficient of variation 0.16. That is to say, on average the skewness of country citation distributions in specific fields, as well as across field citation distributions is of the same order of magnitude. Moreover, the variability across countries within a field or across scientific fields according to the Groeneveld and Meeden index is also of the same order of magnitude.

Table 2. The Skewness of Country Citation Distributions in Eight Fields and the All-fields Case. Average (Standard Deviation), and Coefficient of Variation over 38 Countries of the Groeneveld and Meeden (1984) robust measure of skewness

	Average	Standard Deviation	Coefficient of Variation
PHYSICS	0.64	0.06	0.09
CHEMISTRY	0.54	0.08	0.15
CLINICAL MEDICINE	0.67	0.08	0.12
BIOLOGY & BIOCHEMISTRY	0.56	0.07	0.12
MATERIALS SCIENCE	0.61	0.13	0.21
GEOSCIENCES	0.54	0.09	0.16
ENGINEERING	0.60	0.14	0.24
ECONOMICS & BUSINESS	0.68	0.16	0.24
ALL-SCIENCES	0.38	0.09	0.23

Being as it may, the degree of similarity found across field and sub-field citation distributions has paved the way for meaningful comparisons of citation counts across heterogeneous scientific disciplines. In particular, as indicated in Section II.2, it has been observed that the use of field and sub-field mean citations as normalization factors dramatically reduces the citation inequality attributed to differences in publication and citation practices across them (Crespo *et al.*, 2013a,b, Li *et al.*, and Ruiz-Castillo, 2013). The results in Table 1 (illustrated in Figures 1 to 9), as well as in Table 2 suggest the possibility of investigating whether differences of all sorts between country citation distributions in the different fields and the all-sciences case are sufficiently small for normalization procedures to reduce the citation inequality attributable to them –a task pursued in the next Section.

III. THE EFFECT ON CITATION INEQUALITY OF DIFFERENCES IN CITATION IMPACT ACROSS COUNTRIES

III.1. A Measurement Framework

The fact just analyzed concerning the similarity between the shapes of country citation distributions, should not lead us to ignore the differences in citation impact across countries within any scientific discipline. Borrowing from Crespo *et al.* (2013a), this Section briefly presents a framework where these differences can be quantified in a given field using the notation introduced in Section II.1.

For any country p, let us partition the citation distribution $\mathbf{C_p}$ into Π quantiles of size N_p/Π. That is, let $\mathbf{C_p} = (\mathbf{C_p}^{1},\ldots, \mathbf{C_p}^{\pi},\ldots, \mathbf{C_p}^{\Pi})$, where $\mathbf{C_p}^{\pi} = \{c_{pj}{}^{\pi}\}$ is the vector of the number of citations received by the N_p/Π articles in the π-th quantile of distribution $\mathbf{C_p}$, with $j = 1,\ldots, N_p/\Pi$, and $c_{pj} = c_k$ for some article k in the original distribution $\mathbf{Q}$. Assume for a moment that we disregard the citation inequality within every vector $\mathbf{C_p}^{\pi}$ by assigning to every article in that vector the mean citation of the vector itself, $\mu_p{}^{\pi}$, defined by

$$\mu_p{}^{\pi} = (\Sigma_j\, c_{pj}{}^{\pi})/(N_p/\Pi).$$

The interpretation of the fact that, for example, $\mu_p{}^{\pi} = 2\ \mu_q{}^{\pi}$ is that, on average, country p receives twice the number of citations as country q to represent a common underlying phenomenon, namely, the same degree of citation impact in both countries. In other words, for any π, the distance between $\mu_p{}^{\pi}$ and $\mu_q{}^{\pi}$ is entirely attributable to the difference in the citation performance that prevails in the two countries for articles that represent the same degree of citation impact within each of them.

For any π, consider the distribution $(\boldsymbol{\mu_1}^{\pi},\ldots, \boldsymbol{\mu_p}^{\pi},\ldots, \boldsymbol{\mu_P}^{\pi})$ where, for each p, each article in vector $\mathbf{C_p}^{\pi}$ receives the mean citation of the vector itself, $\mu_p{}^{\pi}$. The citation inequality of this distribution according to any relative inequality index I, $I(\mu_1{}^{\pi},\ldots, \mu_P{}^{\pi})$, abbreviated $I(\pi)$, is entirely attributable to differences in citation impact across the P countries at quantile π. Hence, any weighted average of these quantities provides a good measure of the citation inequality due to such differences. In what follows, we introduce an appropriate citation inequality index, and its associated weighting system.

As in Crespo *et al.* (2013a), for any distribution $\boldsymbol{Z} = (z_1, \ldots, z_K)$ with K elements, indexed by $k = 1, \ldots, K$, it is useful to use an additively decomposable inequality index, denoted by I_1, defined as:

$$I_1(\boldsymbol{Z}) = (1/K)\, \Sigma_k\, (z_k/\mu) \log (z_k/\mu),$$

where μ is the mean of distribution $\boldsymbol{Z}$. For each π, define the vector $\boldsymbol{C}^{\pi} = (\boldsymbol{C}_1^{\pi}, \ldots, \boldsymbol{C}_p^{\pi}, \ldots, \boldsymbol{C}_P^{\pi})$ of size $(\Sigma_p N_p)/\Pi = N/\Pi$. Clearly, the set of vectors $\boldsymbol{C}^{\pi}$, $\pi = 1, \ldots, \Pi$, form a partition of $\boldsymbol{C}$. Apply the decomposability property of citation inequality index I_1 to the partition $\boldsymbol{C} = (\boldsymbol{C}^1, \ldots, \boldsymbol{C}^{\pi}, \ldots, \boldsymbol{C}^{\Pi})$:

$$I_1(\boldsymbol{C}) = \Sigma_{\pi} V^{\pi} I_1(\boldsymbol{C}^{\pi}) + I_1(\boldsymbol{\mu}^1, \ldots, \boldsymbol{\mu}^{\Pi}), \qquad (1)$$

where V^{π} is the share of total citations in $\boldsymbol{C}$ received by articles in $\boldsymbol{C}^{\pi}$, and $(\boldsymbol{\mu}^1, \ldots, \boldsymbol{\mu}^{\Pi})$ is the distribution where each article in sub-group $\boldsymbol{C}^{\pi}$ is assigned the citation mean of the sub-group, $\mu^{\pi} = \Sigma_p\, (N_p/\Pi)\ \mu_p^{\pi}$. Next, apply the decomposability property of I_1 to the partition $\boldsymbol{C}^{\pi} = (\boldsymbol{C}_1^{\pi}, \ldots, \boldsymbol{C}_p^{\pi}, \ldots, \boldsymbol{C}_P^{\pi})$:

$$I_1(\boldsymbol{C}^{\pi}) = \Sigma_p V_p^{\pi} I_1(\boldsymbol{C}_p^{\pi}) + I_1(\boldsymbol{\mu}_1^{\pi}, \ldots, \boldsymbol{\mu}_P^{\pi}), \qquad (2)$$

where V_p^{π} is the share of total citations in $\boldsymbol{C}^{\pi}$ received by articles in $\boldsymbol{C}_p^{\pi}$, and $(\boldsymbol{\mu}_1^{\pi}, \ldots, \boldsymbol{\mu}_P^{\pi})$ is the distribution where each article in quantile $\boldsymbol{C}_p^{\pi}$ is assigned the citation mean of the quantile, μ_p^{π}.

Substituting (2) into (1), we obtain that the overall citation inequality in the double partition of distribution $\boldsymbol{C}$ into P countries and Π quantiles can be decompose into the following three terms:

$$I_1(\boldsymbol{C}) = W + S + IDCC, \tag{3}$$

where:
$$W = \Sigma_\pi \Sigma_p V_p^{\pi} I_1(\boldsymbol{C_p}^{\boldsymbol{\pi}}),$$

$$S = I_1(\boldsymbol{\mu}^{\boldsymbol{1}},..., \boldsymbol{\mu}^{\boldsymbol{\Pi}}),$$

$$IDCC = \Sigma_\pi V^\pi I_1(\boldsymbol{\mu_1}^{\boldsymbol{\pi}},\ldots, \boldsymbol{\mu_P}^{\boldsymbol{\pi}}) = \Sigma_\pi V^\pi I_1(\pi). \tag{4}$$

The term W in Eq. 3 is a within-group term that captures the weighted citation inequality within each quantile in every country. For large Π, $I_1(\boldsymbol{C_p}^{\boldsymbol{\pi}})$, and hence W is expected to be small. The S term is the citation inequality of the distribution $(\boldsymbol{\mu}^{\boldsymbol{1}},..., \boldsymbol{\mu}^{\boldsymbol{\Pi}})$ in which each article in the vector $\boldsymbol{C}^{\boldsymbol{\pi}}$ is assigned the vector's citation mean, μ^π. Thus, S is a measure of citation inequality at different degrees of citation impact in the all-sciences case. Due to the high skewness of science prevalent in all citation distributions, S is expected to be large. Finally, $I_1(\pi)$ is the citation inequality according to I_1 attributable to differences in citation impact across countries at quantile π. Thus, the $IDCC$ term in Eq. 3, which is a weighted average of $I_1(\pi)$ for all π, provides a convenient measure of the citation inequality due to such differences over the entire support of citation distributions. Note that, again, due to the skewness of science the weights V^π are expected to increase dramatically with π.

III.2. The Importance of Differences in Citation Impact across Countries

The results concerning the decomposition of Eq. 3 for the eight fields and the all-science case are presented in Table 2 for the choice $\Pi = 100$, a practice maintained in the sequel. Two points should be emphasized. Firstly, as expected, the W and S terms are small and large, respectively. Secondly, relative to total citation inequality, the importance of the *IDCC* term ranges from 3.7% and 3.9% in Engineering and Clinical Medicine, to 8.4% and 10.1% in Biology & Biochemistry and Chemistry, with 5.4% in the all-sciences case (see column 7 in Table 3).

Table 3. Total Citation Inequality Decomposition for Eight Fields and the All-sciences Case

	Within-group Term, *W*	**Skew. of Science Term, *S***	***IDCC* Term**	**Total Citation Ineq., $I_1(C)$**	**In %: (1)/(4)**	**(2)/(4)**	**(3)/(4)**
FIELDS	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
PHYSICS	0.0427	0.8440	0.0435	0.9305	4.6	90.7	4.7
CHEMISTRY	0.0181	0.6128	0.0711	0.7021	2.6	87.3	10.1
CLINICAL MEDICINE	0.0318	0.8432	0.0351	0.9101	3.5	92.6	3.9
BIOLOGY & BIOCHEM	0.0166	0.5547	0.0527	0.6240	2.7	88.9	8.4
MATERIALS SCIENCE	0.0169	0.8293	0.0574	0.9036	1.9	91.8	6.4
GEOSCIENCES	0.0161	0.5900	0.0504	0.6566	2.5	89.9	7.7
ENGINEERING	0.0147	0.8345	0.0330	0.8822	1.7	94.6	3.7
ECONOMICS & BUSINESS	0.0098	0.8038	0.0426	0.8562	1.1	93.9	5.0
ALL SCIENCES	0.0275	0.8107	0.0477	0.8860	3.1	91.5	5.4

This order of magnitude can be compared with the importance of differences in citation practices across scientific disciplines in previous contributions. Firstly, there are three studies that, using the same dataset as this paper, differ in the classification system of articles into scientific disciplines. For the 22 fields that include the eight fields in this paper, Crespo *et al.* (2013a) find that an *IDCP* term (where *IDCP* stand for *I*nequality of *D*ifferences in *C*itation *P*ractices) represents about 14% of total citation inequality. For sub-fields identified with 219 Web of Science subject-categories, and its aggregation into 19 broad fields, Crespo *et al.* (2013b) and Li and Ruiz-Castillo (2013) find that the *IDCP* term represents 18% and 12.5% of total citation inequality, respectively. Secondly, Li *et al.* (2013) use publications in 172 Web of Science subject-categories that appeared in six different years, spanning a period of more than two decades from 1980 to 2004, with citation windows ranging from about seven years for the papers published in

2004, to 31 years for the 1980 subset.[4] In spite of the many differences between the six yearly datasets, the *IDCP* term represents about 13% of total citation inequality over the entire period.

Therefore, it would appear that the differences in citation impact across countries in the all-sciences case, for example, are smaller than the differences in citation practices across scientific disciplines. However, when the distinction between domestic and international articles is introduced in Section III.4 a somewhat different picture would emerge.

III. 3. The Consequences of Normalization

The importance of MCs in accounting for differences in country citation distributions can be measured by the extent of the reduction in the *IDCC* term as a consequence of this type of normalization.[5] The graphical evidence is in Figure 10 to 18. In each case, we represent the curves $V^{\pi} I_1(\pi)$ (recall Eq. 4 in Section III.1) as a function of π before and after normalization. The *IDCC* term is the integral below these curves. Since these expressions reach very high values at the lower tail of citation distributions, for clarity each Figure starts at the 33th percentile. Relative to the blue curve that represents the raw data, the red curve in all Figures illustrates the reduction of the *IDCC* term achieved by normalization.

---

[4] Among the 219 sub-fields in the first dataset, 184 correspond to the natural sciences, and 35 to the social sciences, while among the 172 sub-fields these two classes consist of 170 and 2 sub-fields.

[5] As a kind of robustness check, in Section III.5 we report the results of a second procedure that uses the so-called exchange rates introduced in Crespo *et al.* (2013a) as normalization factors.

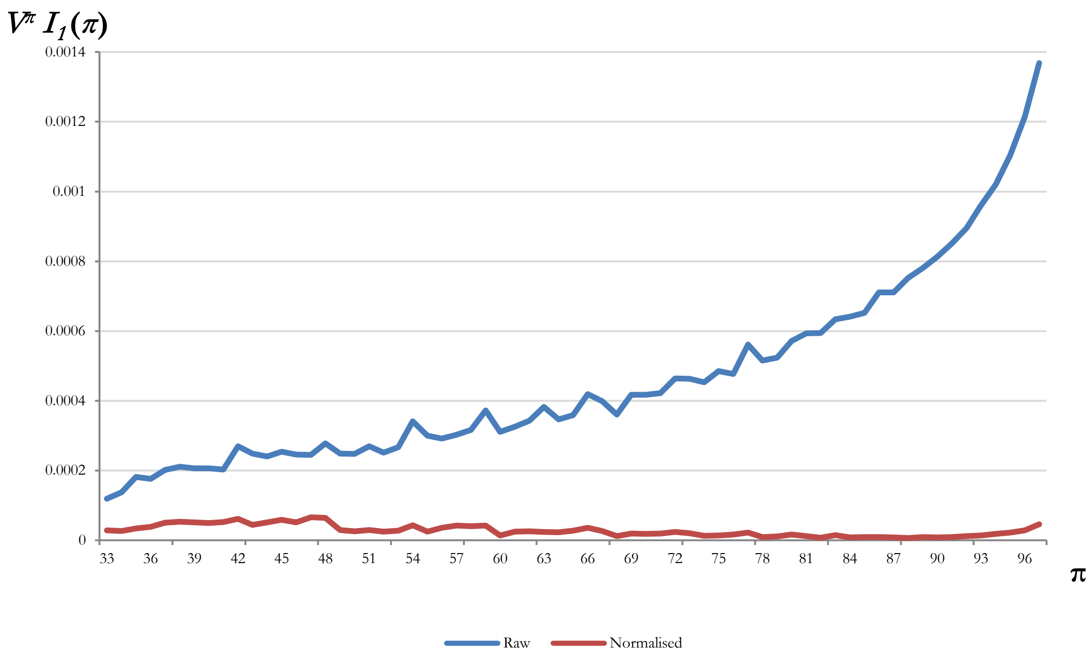

Figure 10. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Physics

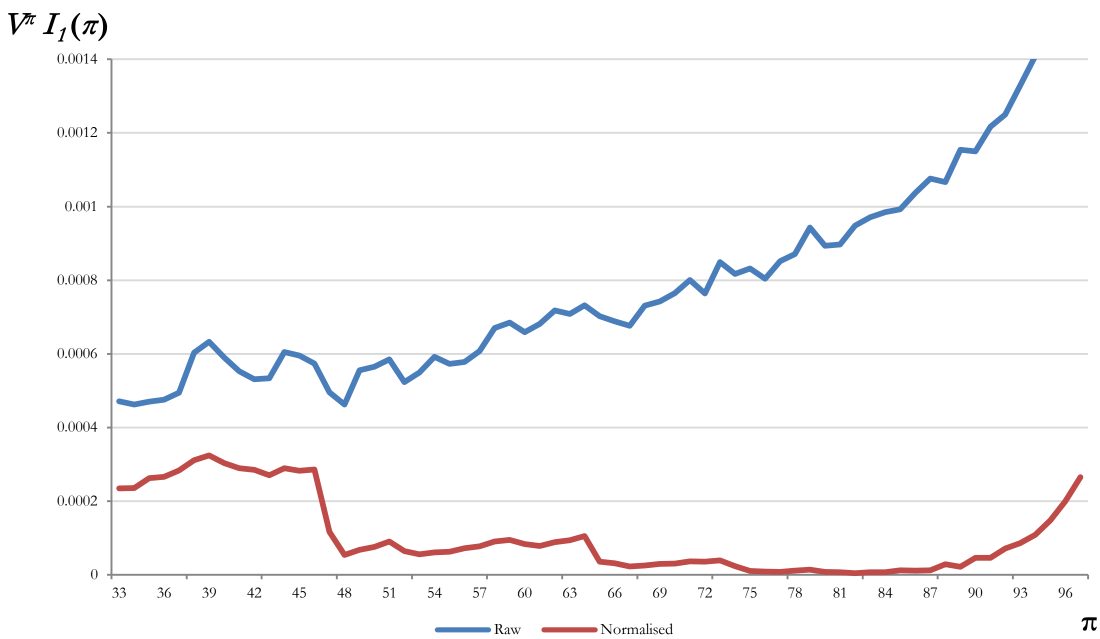

Figure 11. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Chemistry

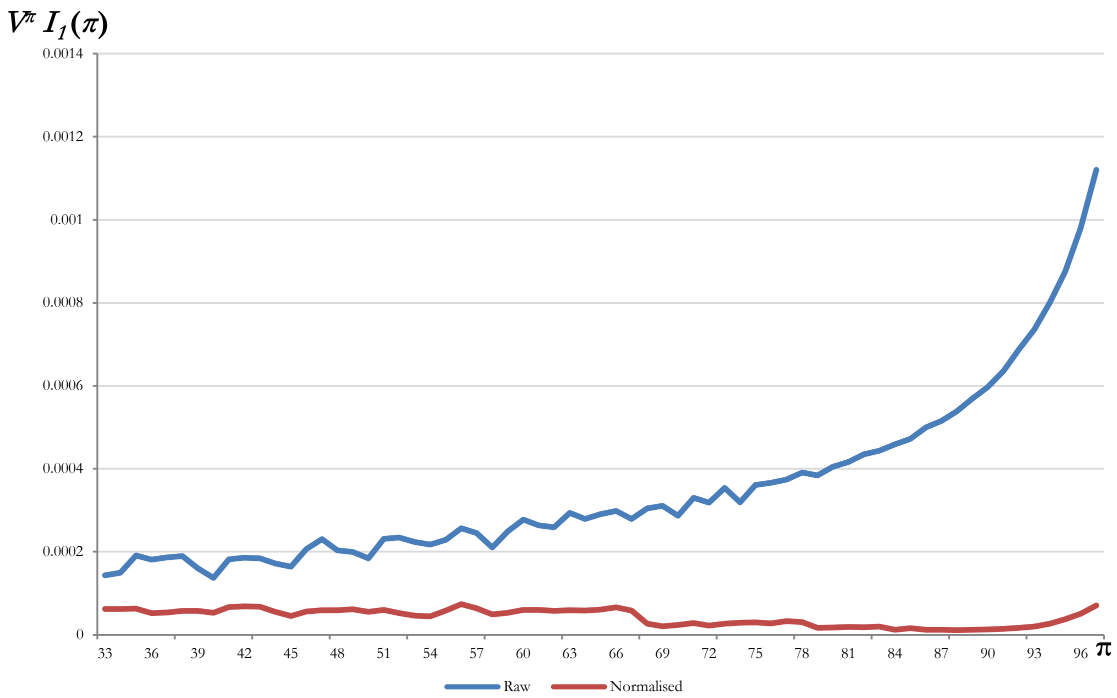

Figure 12. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Clinical Medicine

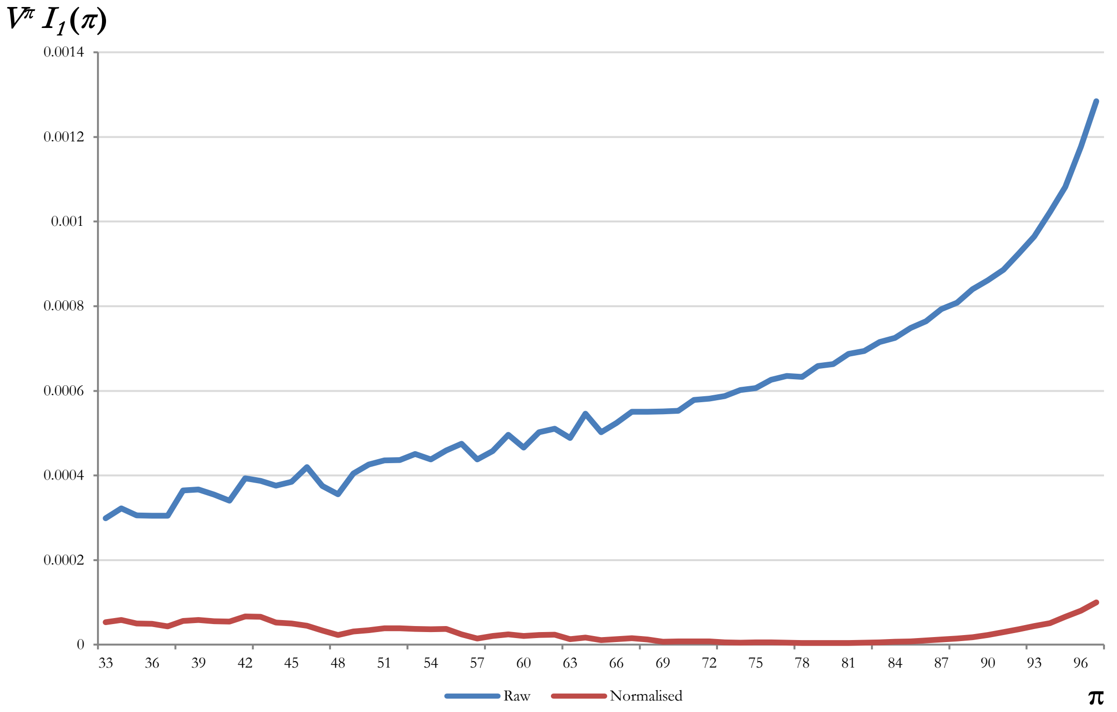

Figure 13. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Biology & Biochemistry

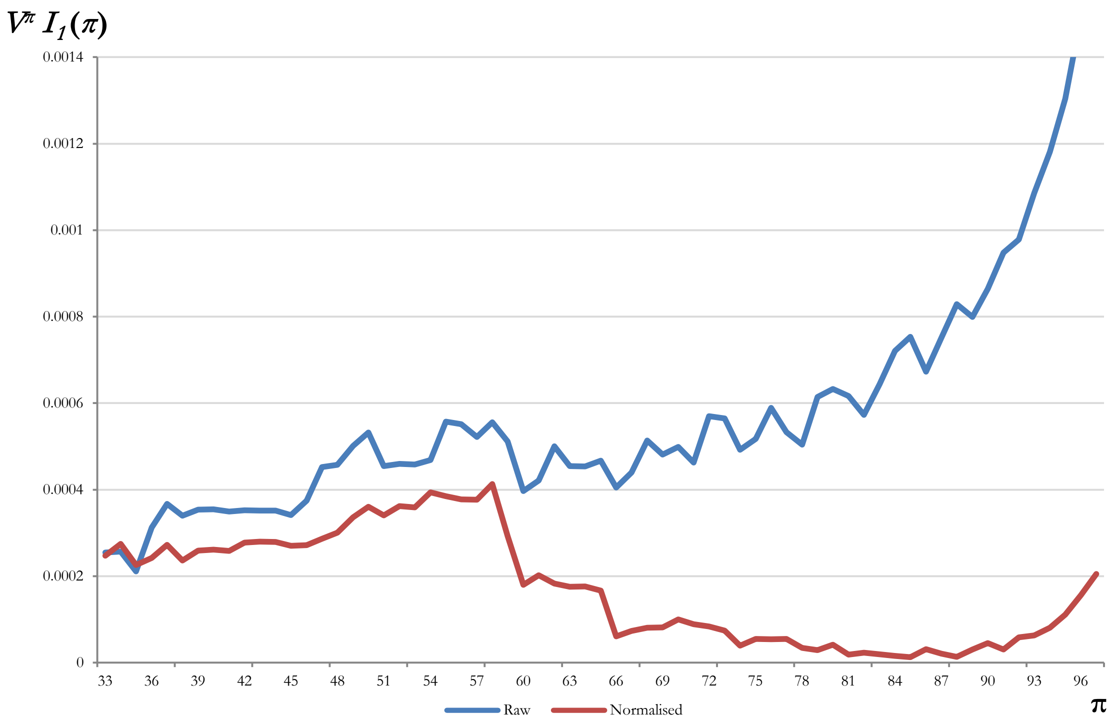

Figure 14. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Materials Science

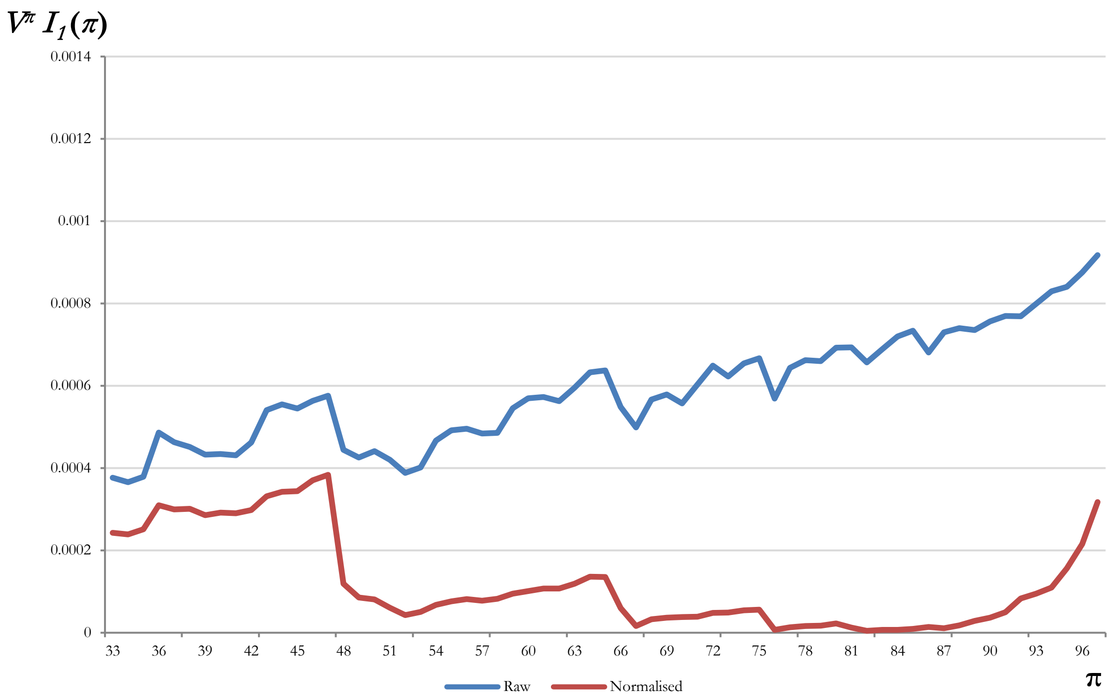

Figure 15. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Geosciences

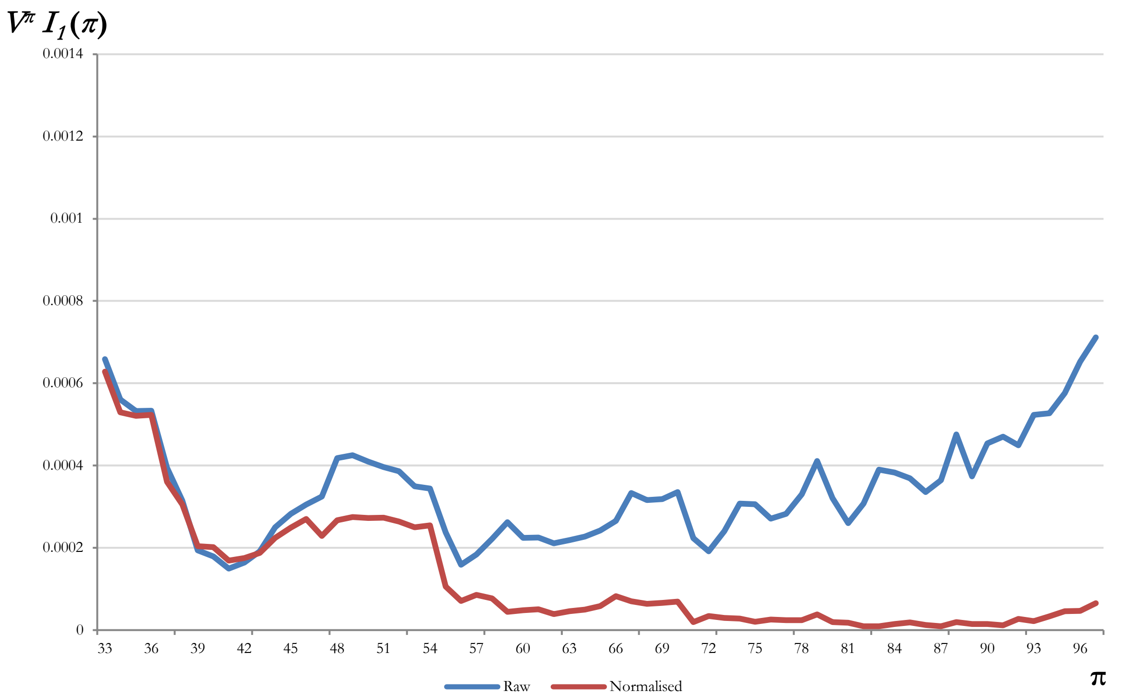

Figure 16. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Engineering

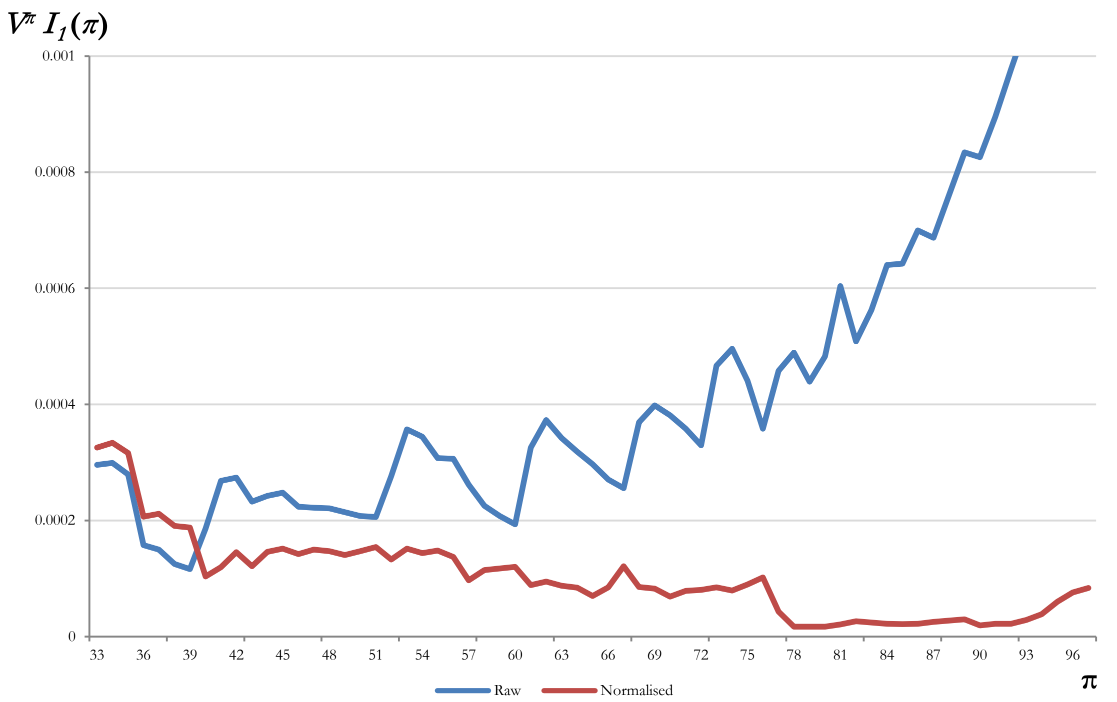

Figure 17. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for Economics & Business

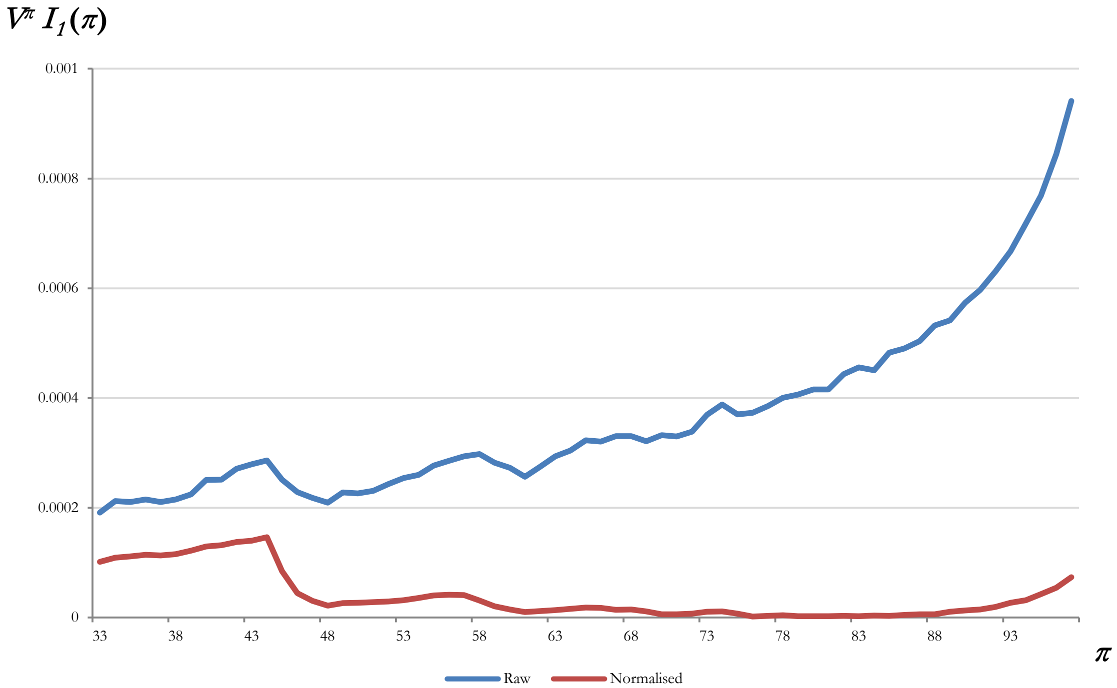

Figure 18. Expression V^{π} I(π) as a function of π. Raw and Country Normalized Data for the All-sciences Case

The results are mixed. For four fields –Physics, Chemistry, Clinical Medicine, and Biology & Biochemistry– and the all-sciences case, the evolution of the curves $V^{\pi} I_1(\pi)$ is rather smooth, and the impact of normalization seems important. Instead, for the remaining fields –Engineering, Materials Science, Geosciences, and Economics & Business– the curves evolve less smoothly and, more importantly, the distance between the two curves up to the 60th or even the 70th percentile is not that large.

More precisely, the consequences of normalization can be numerically measured in two complementary ways. Firstly, denote the *IDCC* term after normalization by *IDCC**. Then, we can simply compare *IDCC* and *IDCC** in absolute terms, expressing the result as the percentage that the difference [*IDCP* – *IDCP**] represents relative to the initial situation, *IDCC*. Secondly, we already know the relative importance of the phenomenon *before* normalization measured by the ratio $IDCP/I_1(\mathbf{C})$ in column 7 in Table 3. Thus, we can compare this percentage with the relative importance of differences in citation practices *after* normalization measured by the ratio

$IDCC^*/I_1(\mathbf{C}^*)$, where $\mathbf{C}^*$ is the citation distribution in each case after normalization. The results are presented in Table 4.

Table 4. The Consequences of Normalization

			Reduction in the differences in citation impact as a consequence of normalization in %		
	*IDCC**	*I(C*)*	In absolute terms:	In relative terms	
FIELDS	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
PHYSICS	0.0082	0.9123	81.0	4.7	0.9
CHEMISTRY	0.0158	0.6738	77.9	10.1	2.3
CLINICAL MEDICINE	0.0080	0.8935	77.1	3.9	0.9
BIOLOGY & BIOCHEMISTRY	0.0076	0.5967	85.7	8.4	1.3
MATERIALS SCIENCE	0.0200	0.8819	65.0	6.4	2.3
GEOSCIENCES	0.0172	0.6541	65.8	7.7	2.6
ENGINEERING	0.0154	0.8764	53.3	3.7	1.8
ECONOMICS & BUSINESS	0.0135	0.8345	68.4	5.0	1.6
ALL SCIENCES	0.0070	0.8596	85.2	5.4	0.8

Reduction measured in relative terms: (3) = 100 ***(IDCC - IDCC*)/IDCC***

Reduction measured in relative terms: (4) versus **(5)**, where **(4)** = 100 ***IDCC/I(C)***, taken from column 7 in Table 2, and **(5)** = 100 ***IDCC*/I(C*)*** = 100 **(1)/(2)**

Consistently with the graphical analysis, the reduction in the *IDCC* term for Engineering, Materials Science, Geosciences, and Economics & Business is somewhat limited, ranging from 53.3% to 68.3%.[6] Instead, for the remaining fields –Physics, Chemistry, Clinical Medicine, and Biology & Biochemistry– the reduction ranges from 77% to 86%, with the all-sciences case at 85.2%. In relative terms, the ratio $IDCC/I_1(C)$ for the first four fields is reduced by a factor ranging from 1.9 to 3 –not a small magnitude. For the second group, the reduction of this ratio ranges from 4.3 to 6.5, with 6.7 in the all-sciences case.[7]

---

6 Since the Economics & Business field is relatively small, we have tried a different aggregation into only fourteen countries and four residual geographical areas. Nevertheless, the results were only marginally better than with the previous aggregation scheme.

7 For the latter group, the reduction of the *IDCC* term after normalization is of the same order of magnitude as the corresponding *IDCP* term in the study of the effect on citation inequality of differences in citation practices across scientific fields and sub-fields. In this case, the reduction of the *IDCP* term in absolute terms after mean citation normalization is 86.6% and 84.3% for fields and sub-fields, while the importance of the term relative to overall citation inequality was reduced from 13.95% to 2.05%, and from 18.1% to 3.3% –a 6.8 and 5.5 reduction factor, respectively (Crespo *et al.*, 2013a, b).

III. 4. The Pattern of Differences in Citation Impact from Another Angle

As indicated in Section III.1, due to the skewness of science the weights V^{π} that enter into the definition of the *IDCC* term (see Eq. 4) tend to increase dramatically with π. Consequently, as in Crespo *et al.* (2013a, b) and Li and Ruiz-Castillo (2013a), it is illuminating to graphically observe the evolution of $I_1(\pi)$ as a function of π in Figures 19 to 27 (for percentiles $\pi > 33$). As before, the blue and the red curves represent the situation before and after normalization. This information deserves three comments. Firstly, the distance between two curves illustrates again the consequences of normalization. These are clearly more important for Physics, Chemistry, Clinical Medicine, Biology & Biochemistry and the all-sciences case than for the remaining four fields. Secondly, for later reference, note that, except in Physics and Engineering, the $I_1(\pi)$ curve after normalization tends to raise quite dramatically when we reach the last few percentiles. Thirdly, except for Engineering the effect of differences in citation impact across countries in the raw data (blue curve) tends to decline as we advance towards higher percentiles. This partially offsets the rapidly increasing pattern of the weights V^{π}, leading –as we have saw in column 7 in Table 2– to a relatively low value for the *IDCC* term in comparison with the behavior of the analogous *IDCP* term measuring differences in citation practices across scientific disciplines.

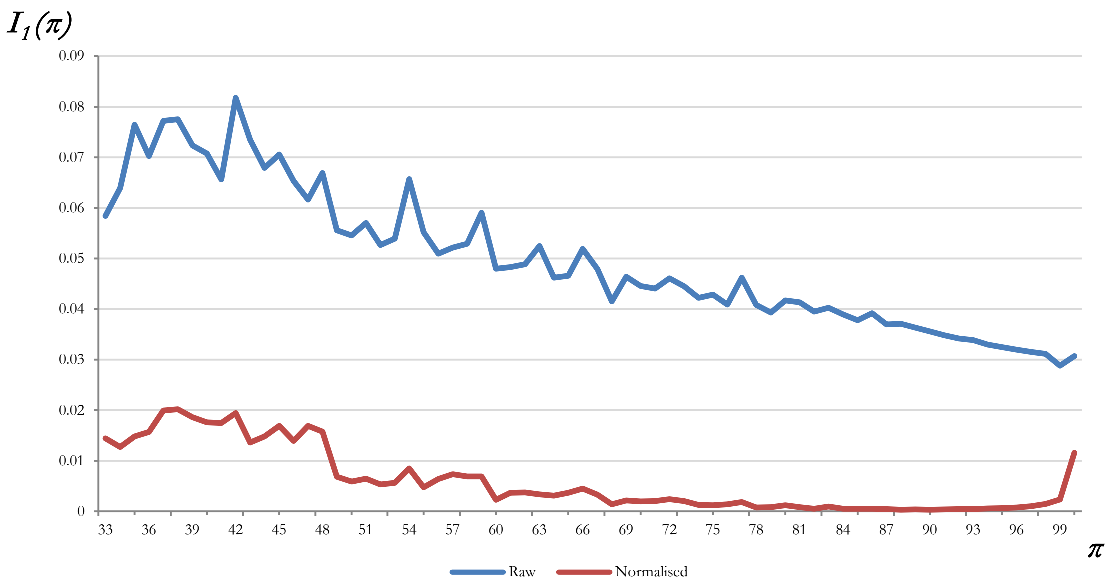

Figure 19. Citation Inequality Due to Differences in Citation Impact Across Countries, $I_1(\pi)$, as a function of π. Raw and Country Normalized Data for Physics

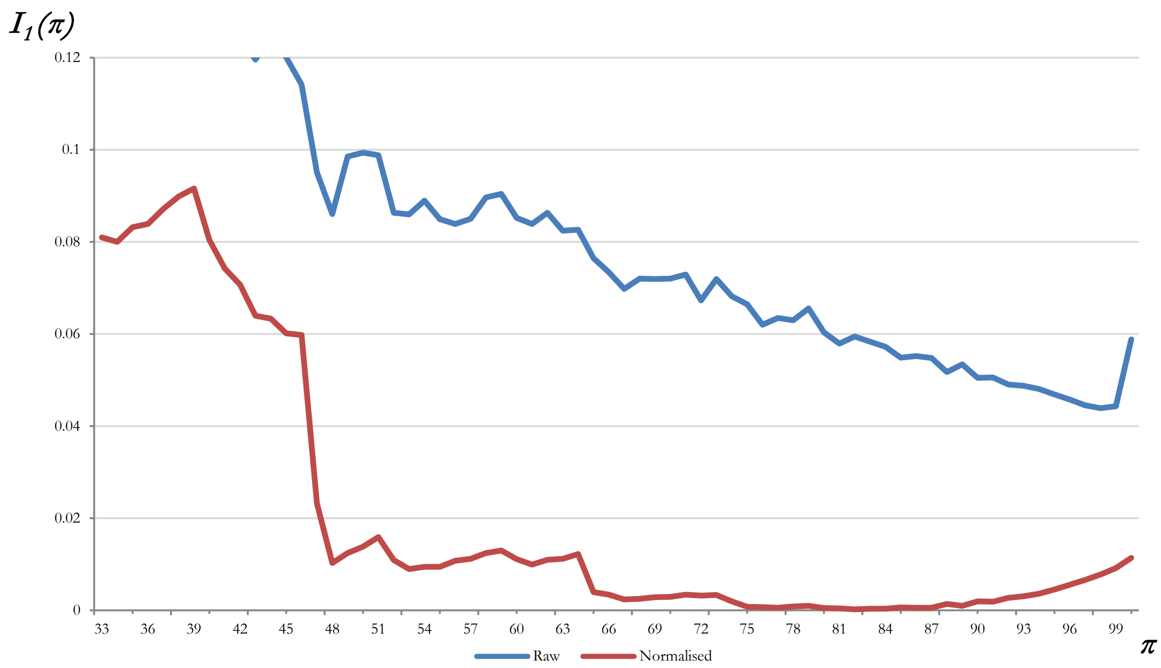

Figure 20. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Chemistry

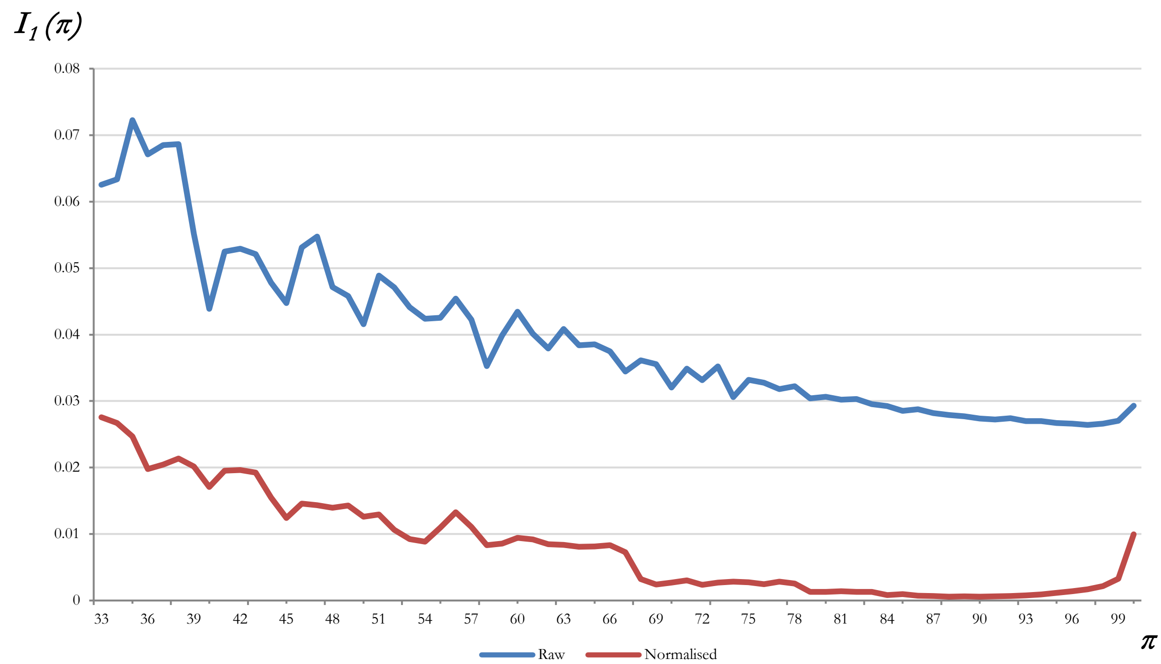

Figure 21. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Clinical Medicine

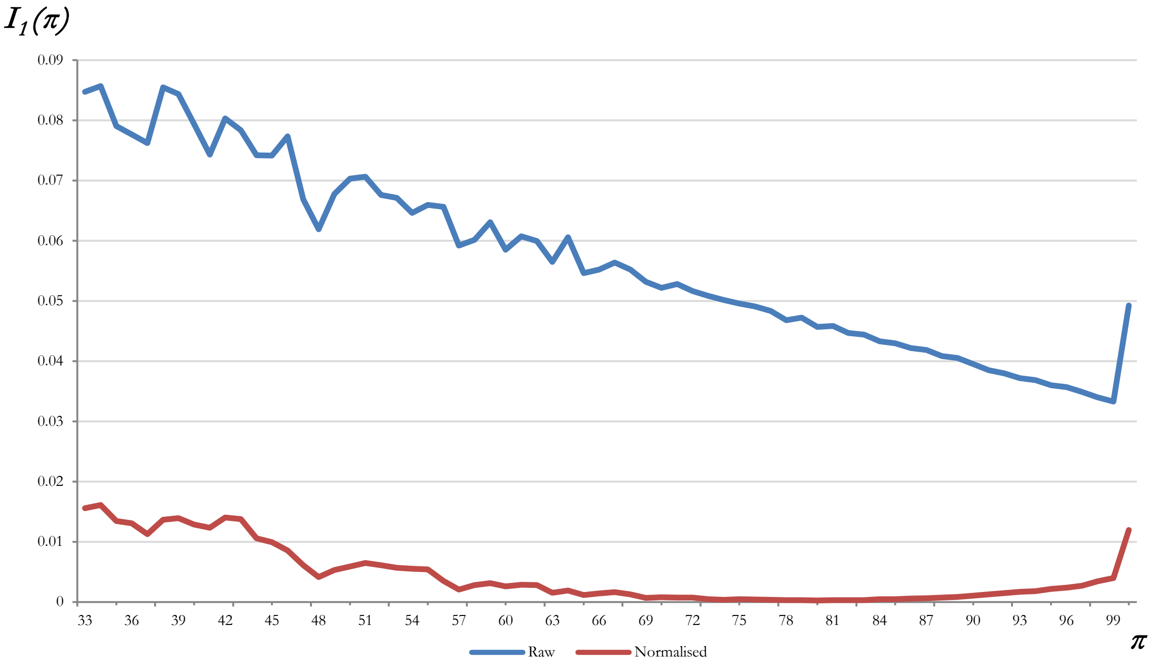

Figure 22. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Biology & Biochemistry

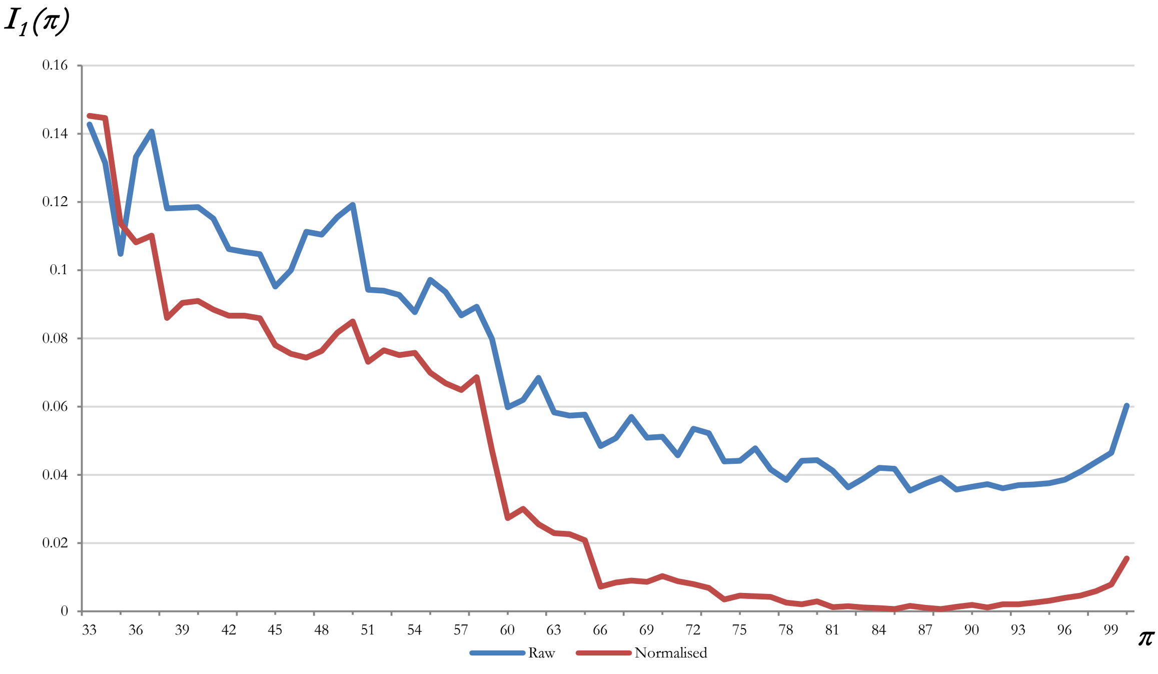

Figure 23. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Materials Science

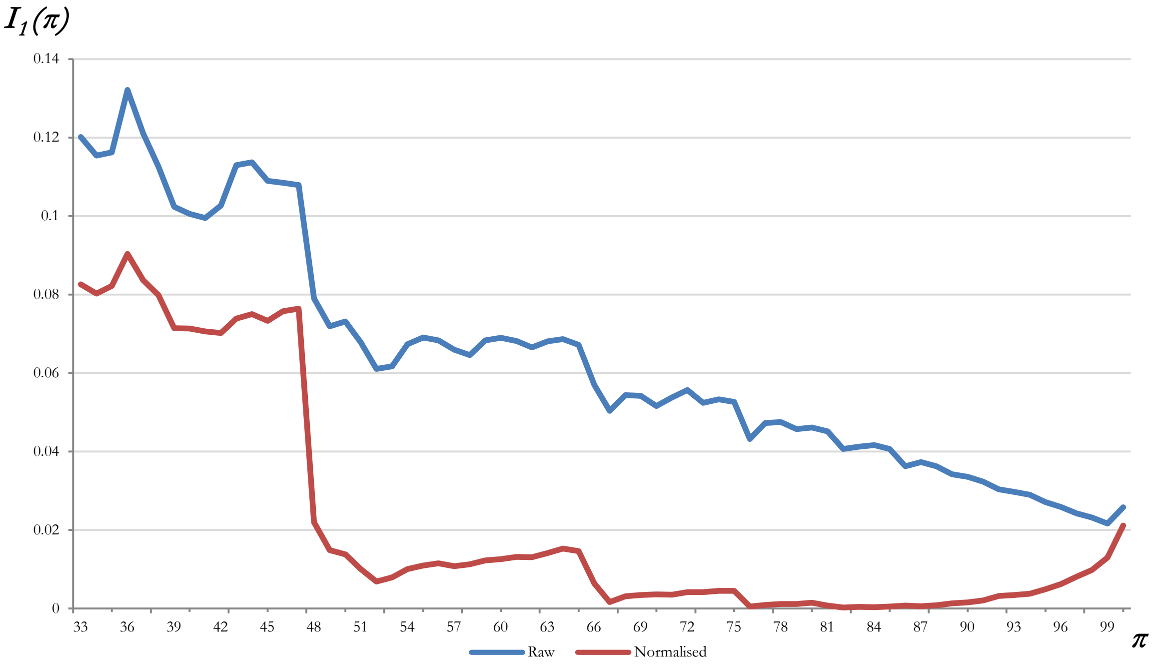

Figure 24. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Geosciences

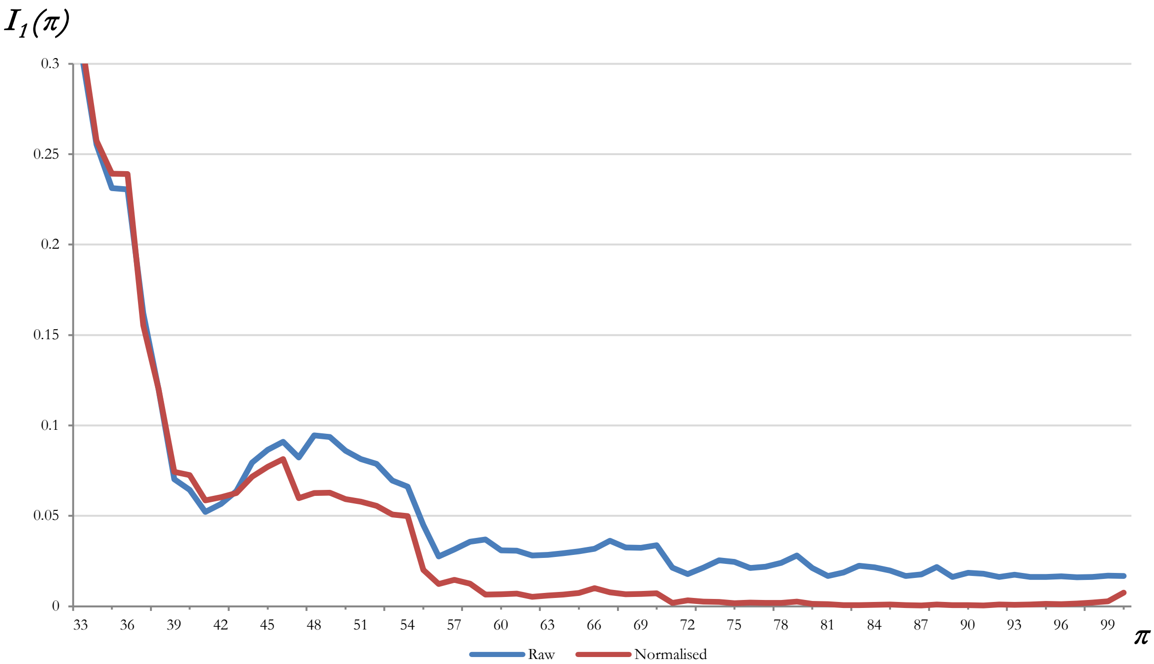

Figure 25. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Engineering

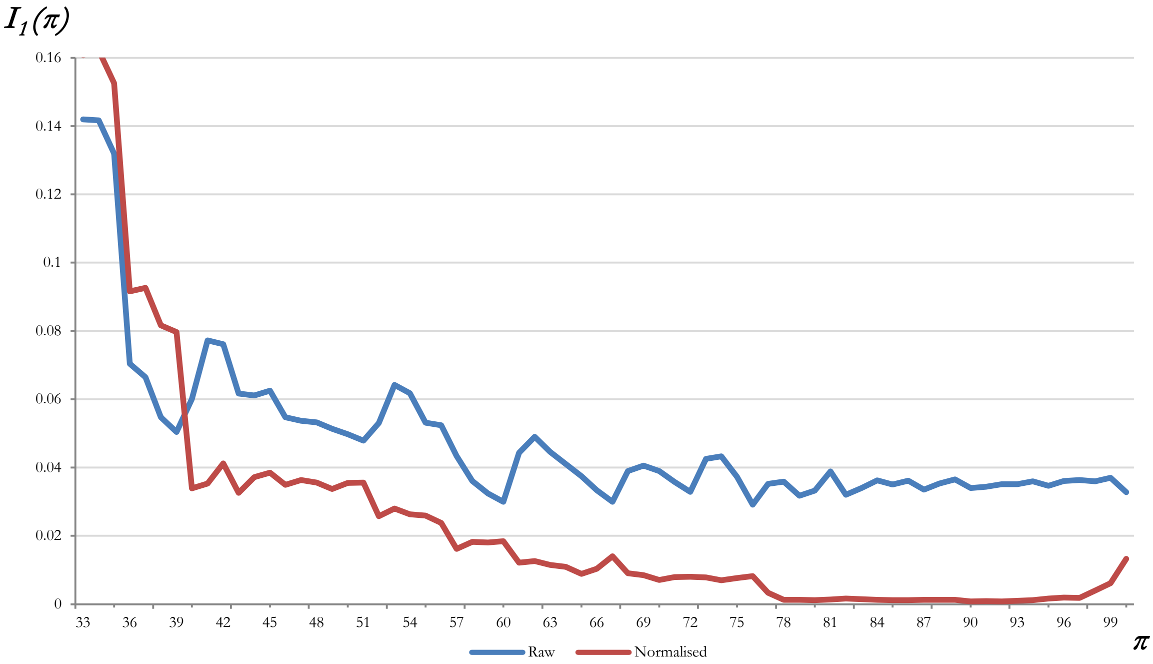

Figure 26. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for Economics & Business

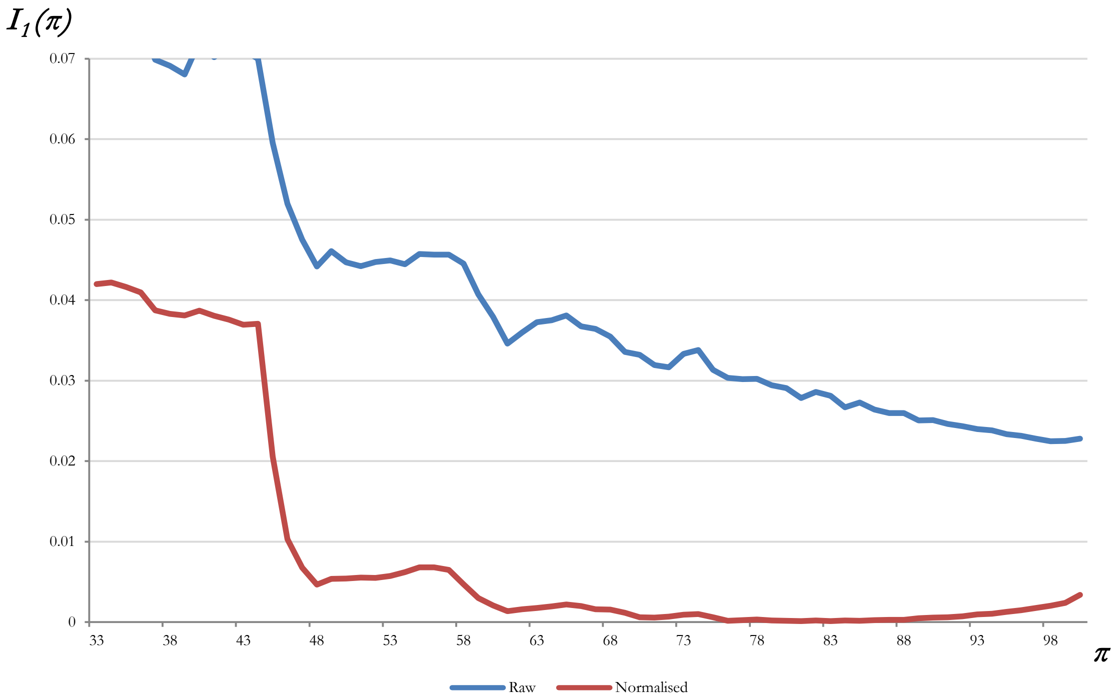

Figure 27. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Raw and Country Normalized Data for the All-sciences Case

To better understand this declining pattern, we have constructed Figures 28 to 36 where the expression $I_1(\pi)$ is represented as a function of π (for values of $\pi > 33$) for two types of articles: domestic articles (green curve), and international articles (orange curve). Three points should be noted. Firstly, independently of which are the cooperating countries, the presence of international articles necessarily decreases the effect on citation inequality of differences between countries (regardless of whether a multiplicative or a fractional approach is followed to deal with international articles). Thus, the fact that the percentage of international articles per percentile increases as we move towards more cited articles (see Figure 37) contributes to the declining pattern of $I_1(\pi)$ in Figures 19 to 27.[8] Secondly, except for Materials Science, the curve $I_1(\pi)$ for international articles tends to decrease with π (see Figures 28 to 36), further contributing to the decline of $I_1(\pi)$ in Figures 19 to 27. Thirdly, at least in five fields –Chemistry, Biology & Biochemistry, Materials Science, Engineering, and Economics & Business– the curve $I_1(\pi)$ for domestic articles shows also a declining pattern as a function of π. A possible explanation is that top authors, and hence highly cited articles in every country, are more alike than lesser authors and less cited articles.[9]

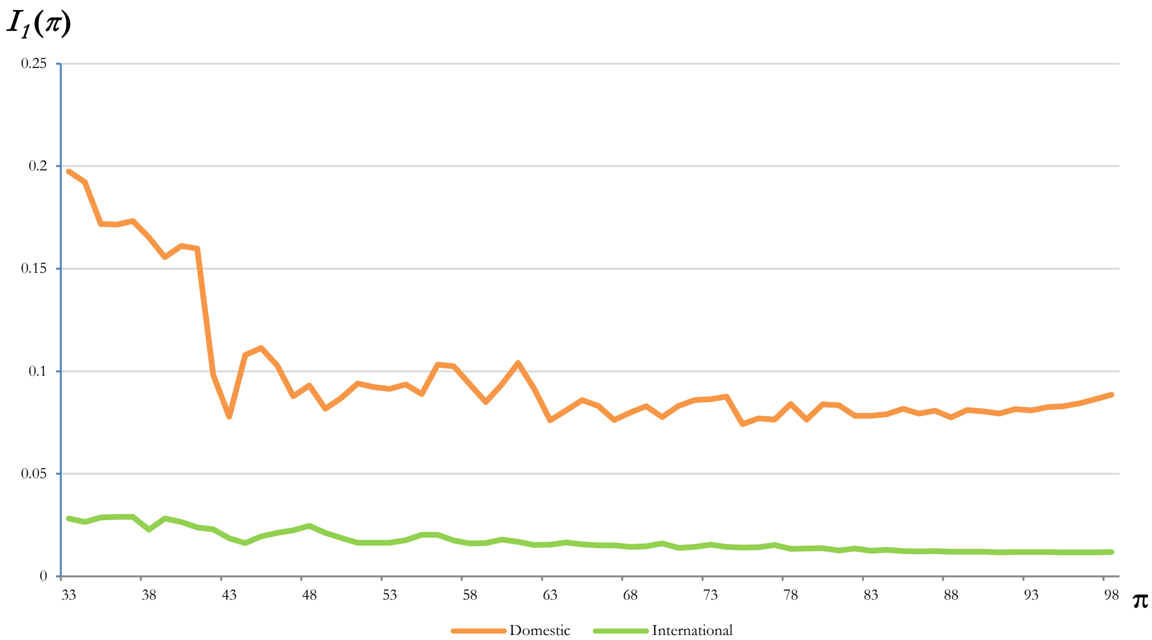

Figure 28. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Physics

[8] For Physics, Geoscience, Economics & Business, and Engineering the percentage of international articles reaches 20% at the second, 15th, 55th, and 64th percentiles, while for the all-sciences case and the remaining four fields international articles reach this percentage between the 72th and the 90th percentiles.

[9] In sports, for example, the teams that reach the quarterfinals in an international tournament are more alike than those who play in previous rounds.

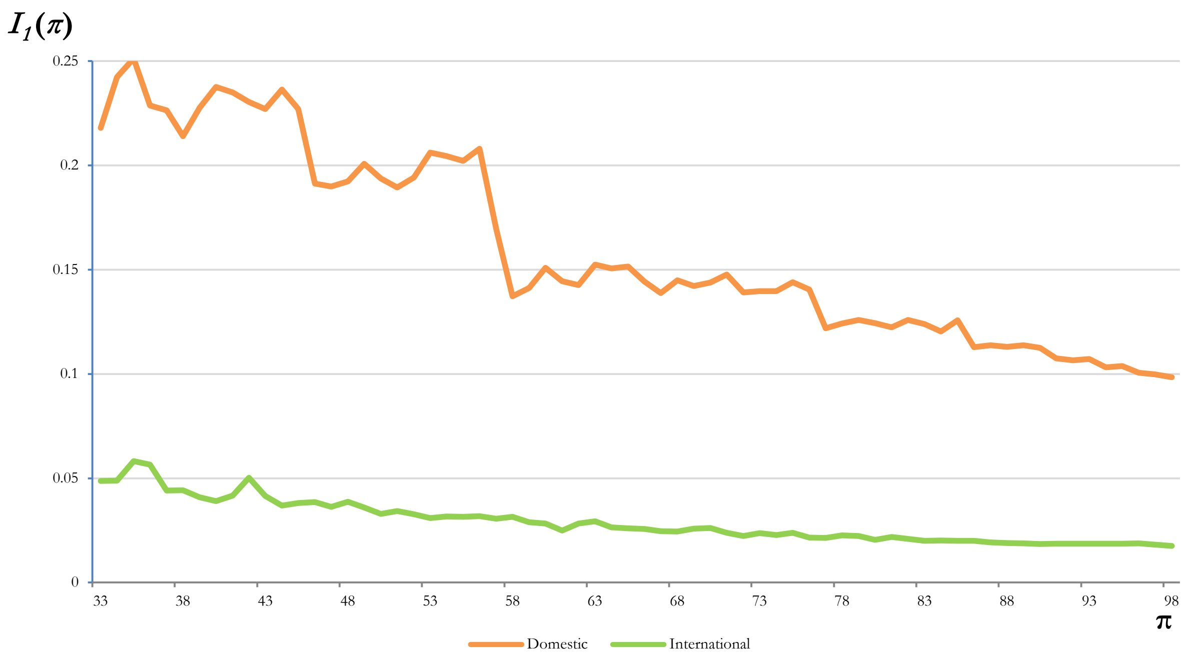

Figure 29. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Chemistry

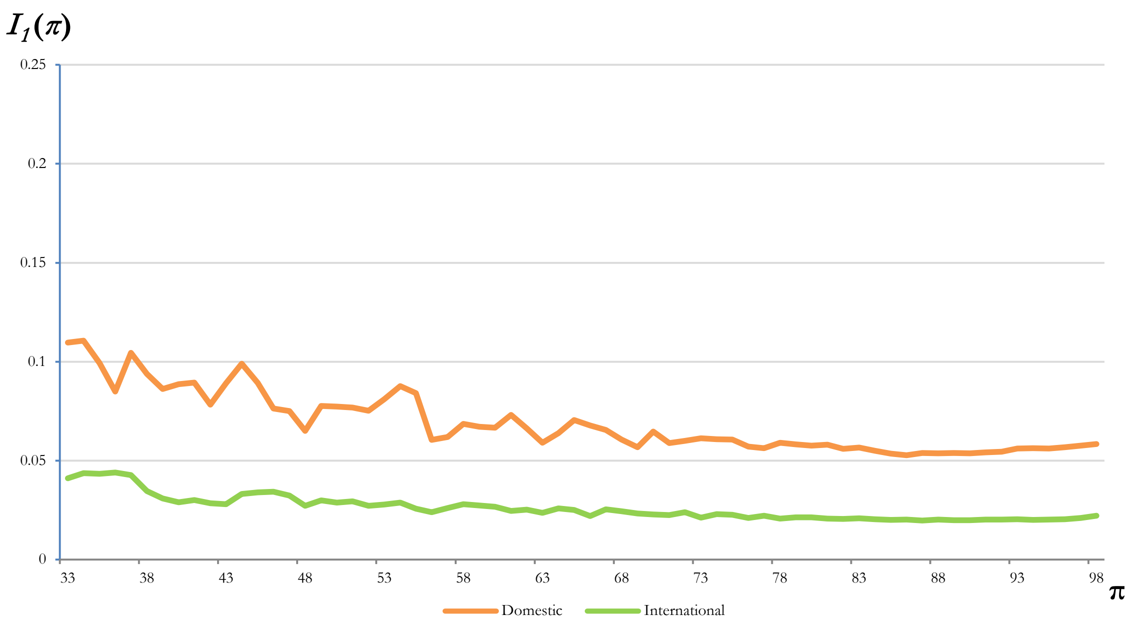

Figure 30. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Clinical Medicine

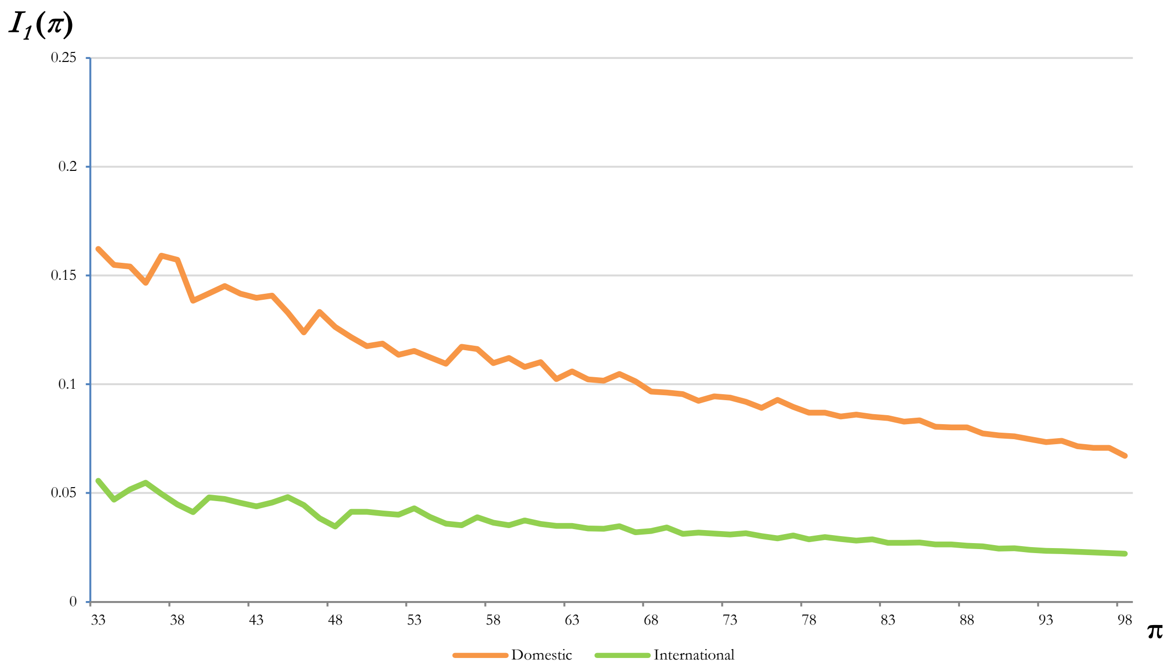

Figure 31. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Biology & Biochemistry

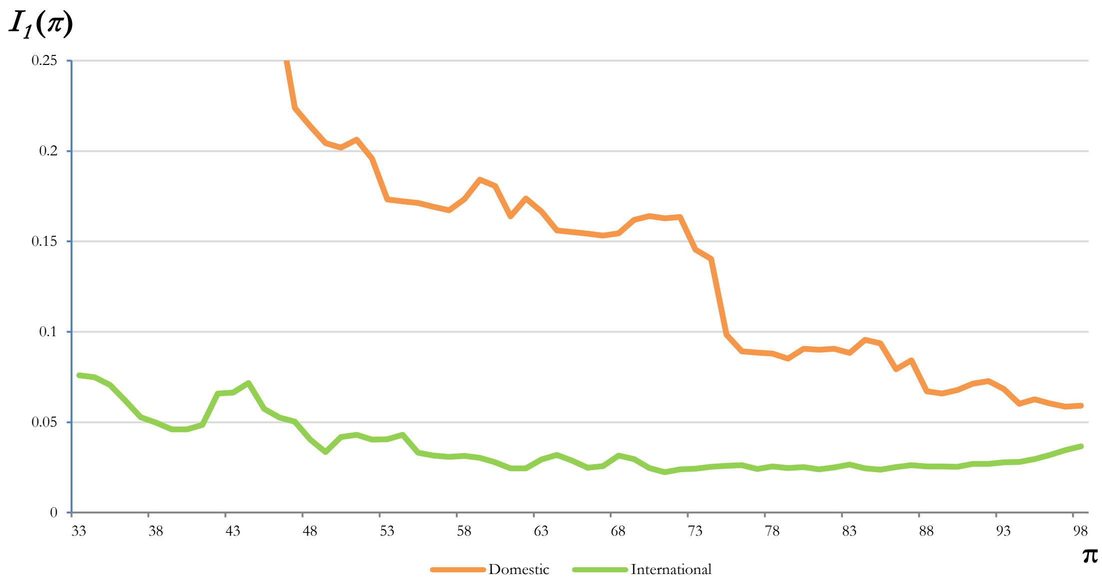

Figure 32. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Materials Science

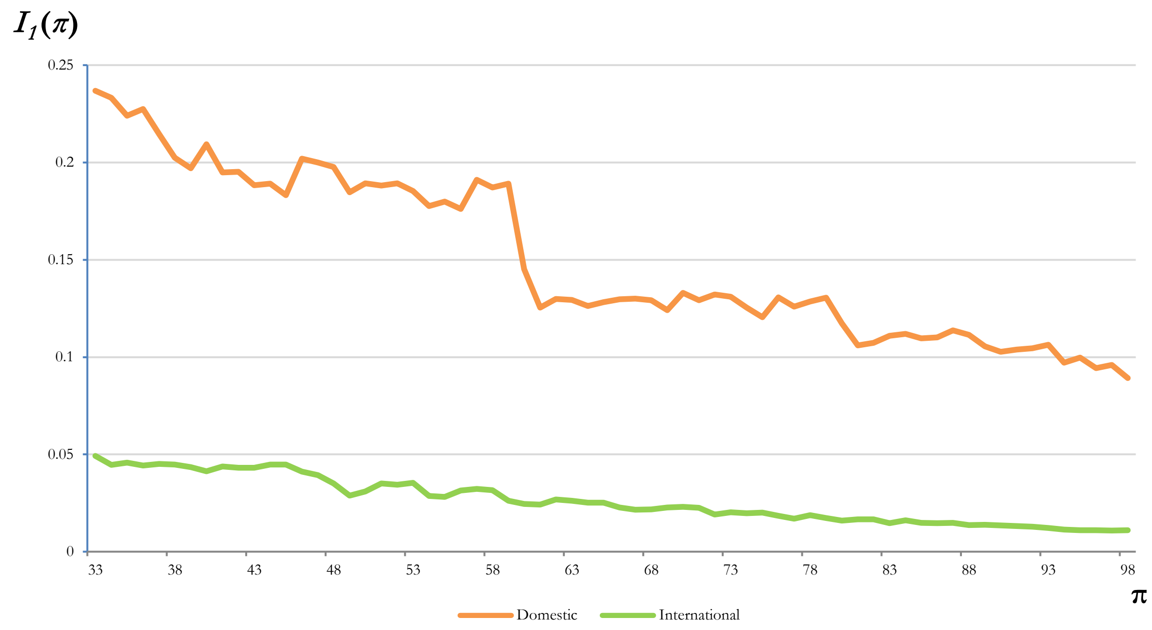

Figure 33. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Geosciences

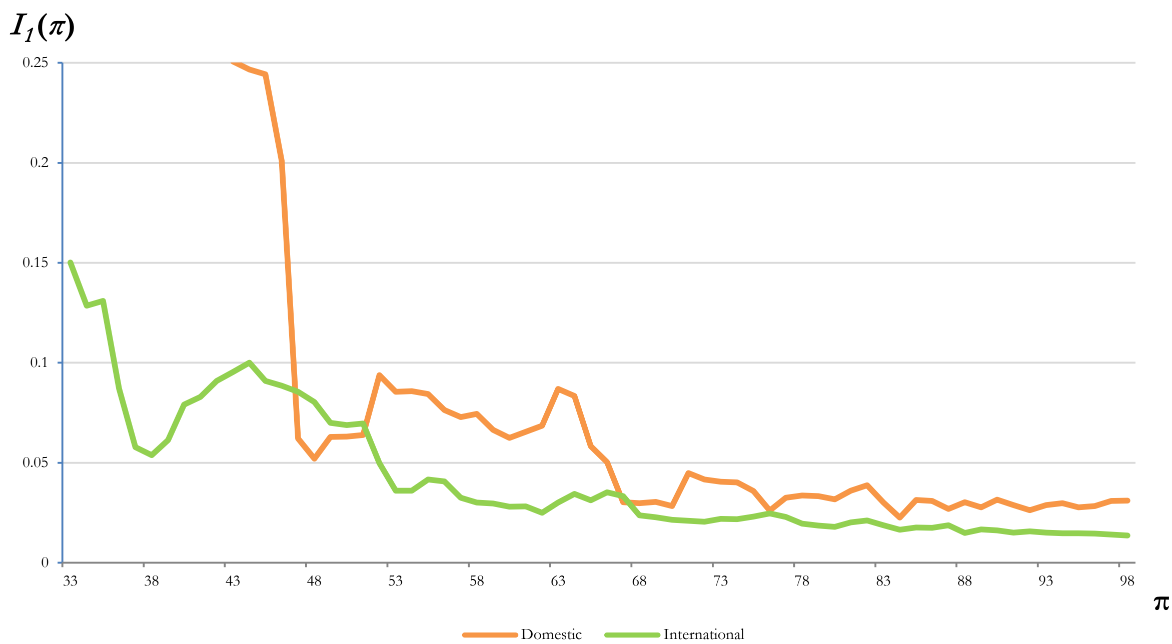

Figure 34. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Engineering

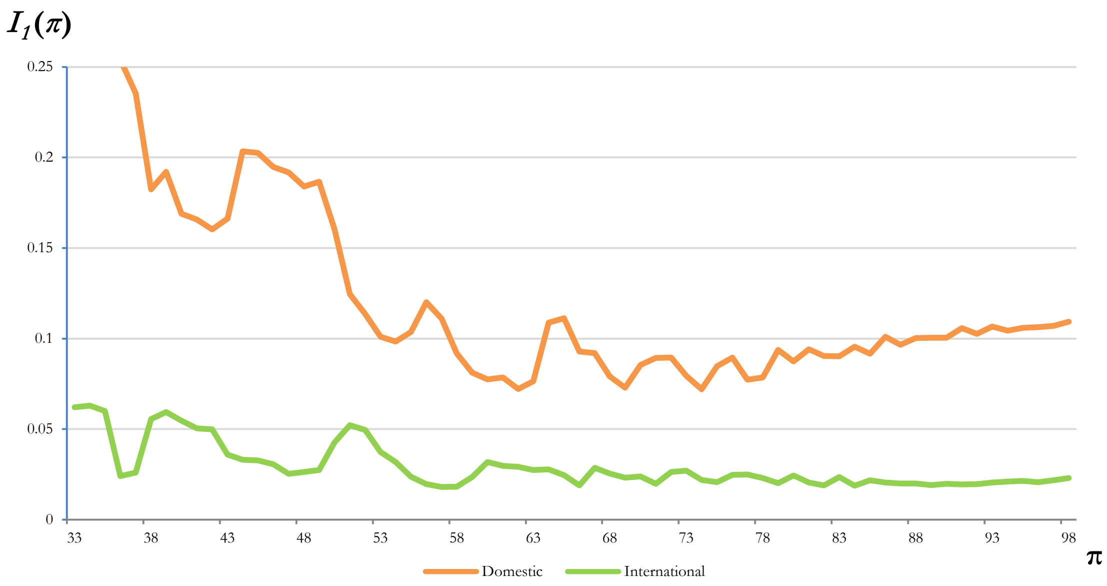

Figure 35. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in Economics & Business

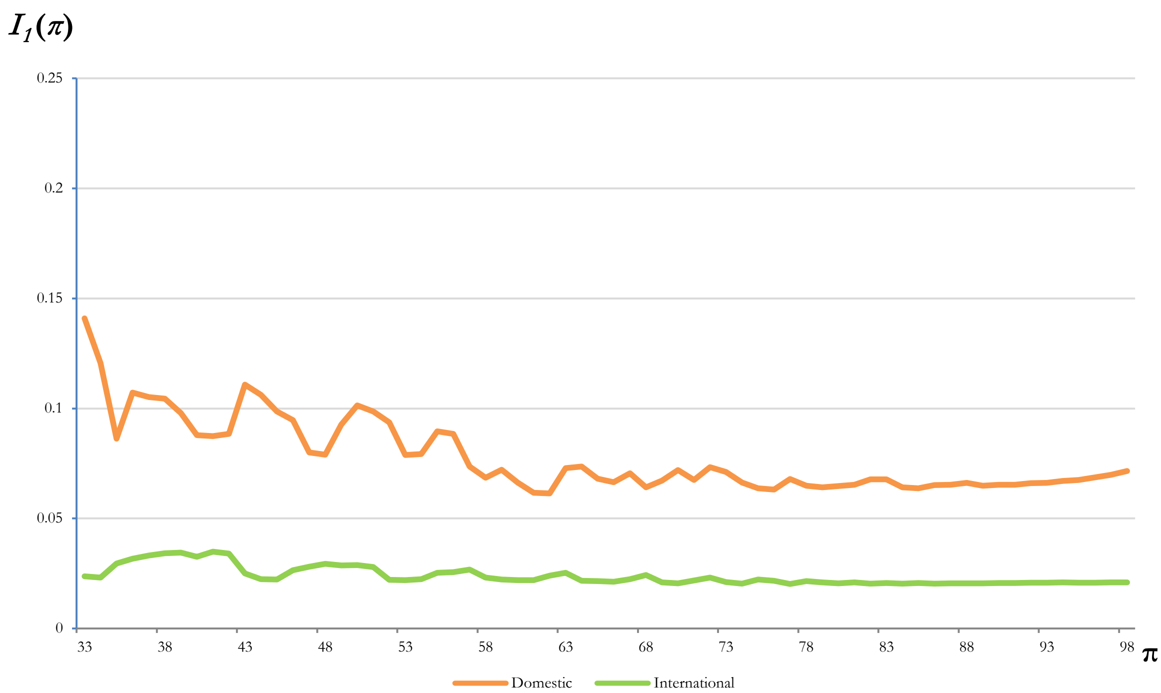

Figure 36. Citation Inequality Due to Differences in Citation Impact Across Countries, I(π), as a function of π. Domestic and International Articles in All-sciences case

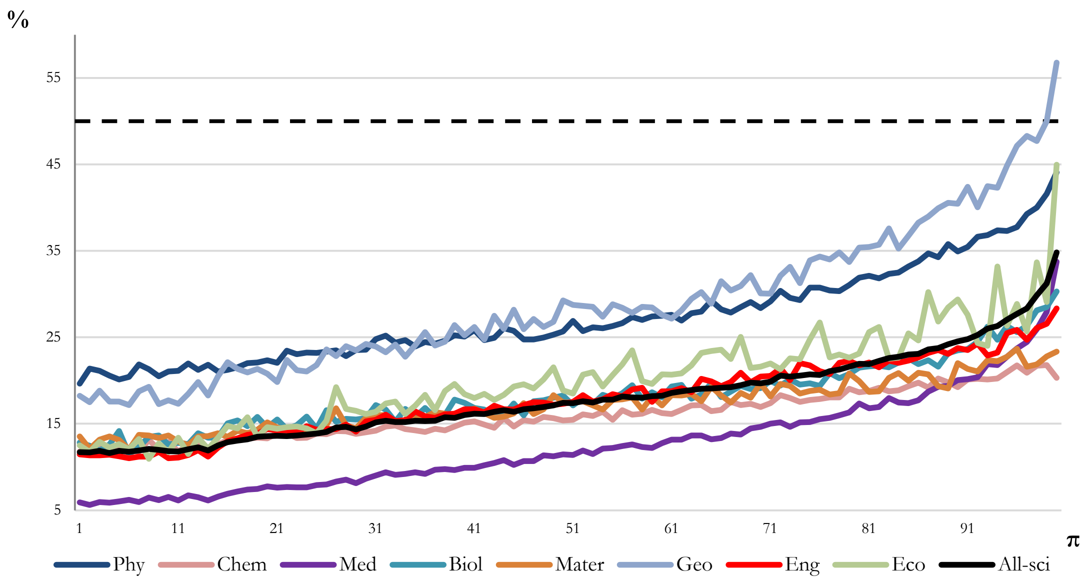

Figure 37. Percentage of international articles by percentile. Physics, Chemistry, Clinical Medicine, Biology & Biochemistry, Materials Science, Geosciences, Engineering, Economics & Business, and the All-sciences case

Finally, it is worthwhile to study separately the importance of differences in citation impact across countries, as well as the consequences of normalization for domestic and international articles. The results in Table 4 deserve the following two comments. Firstly, the effect on citation inequality of differences in citation impact across countries in domestic articles for the raw data (column 3 in Table 5) is much larger than for all articles taken together (column 7 in Table 3). In the all-sciences case, for example, the *IDCC* term represents 8.7% of total citation inequality, a figure much closer to those we reviewed in Section III.2 for the effect of differences in citation practices across scientific fields and sub-fields.

However, as a consequence of international co-authorship, the *IDCC* term for all articles as a whole is reduced to 5.4%. Secondly, the reduction in absolute terms of the *IDCC* term for the two types of articles is in column 7 of Table 5, while the reduction in relative terms can be appreciated by comparing columns 3 and 6. Note that the relative worse results achieved by normalization in four fields are due to different reasons. In Materials Science and Engineering these results can be attributed mostly to domestic articles, while in Geosciences and Economics & Business they rely on the weak impact of normalization in international articles.

Table 5. Total Citation Inequality Decomposition for Domestic and International Articles

PHYSICS	*IDCC* (1)	$I_1(C)$ (2)	100(1)/(2) (3)	*IDCC** (4)	$I_1(C^*)$ (5)	100 (1)/(2) (6)	100 [(1) – (4)]/(1) (7)
A. Domestic articles	0.094	0.93	10.1	0.011	0.85	1.3	88.2
B. International articles	0.019	0.89	2.1	0.005	0.87	0.6	70.6
CHEMISTRY							
A. Domestic articles	0.138	0.79	17.5	0.025	0.73	3.4	81.9
B. International articles	0.028	0.62	4.6	0.006	0.60	1.0	77.8
CLINICAL MEDICINE							
A. Domestic articles	0.069	0.90	7.6	0.014	0.85	1.7	79.2
B. International articles	0.026	0.89	2.9	0.004	0.87	0.5	82.8
BIOLOGY & BIOCHEMISTRY							
A. Domestic articles	0.097	0.62	15.7	0.016	0.58	2.8	83.6
B. International articles	0.035	0.62	5.6	0.005	0.60	0.9	84.7
MATERIALS SCIENCE							
A. Domestic articles	0.111	0.97	11.5	0.057	0.98	5.9	42.8
B. International articles	0.038	0.83	4.6	0.011	0.80	1.3	72.4
GEOSCIENCES							
A. Domestic articles	0.131	0.69	18.9	0.033	0.65	5.0	74.9
B. International articles	0.024	0.63	3.9	0.008	0.62	1.4	65.0
ENGINIEERING							
A. Domestic articles	0.050	0.98	5.1	0.021	0.96	2.2	57.8
B. International articles	0.030	0.84	3.6	0.015	0.84	1.8	50.9
ECONOMICS & BUSINESS							
A. Domestic articles	0.109	0.97	11.3	0.025	0.89	2.8	77.0
B. International articles	0.028	0.79	3.6	0.012	0.78	1.5	58.2
ALL SCIENCES							
A. Domestic articles	0.077	0.88	8.7	0.011	0.82	1.3	85.8
B. International articles	0.024	0.83	2.9	0.002	0.81	0.3	89.3

Reduction measured in relative terms: (3) = 100 *(IDCC - IDCC*)/IDCC*

Reduction measured in relative terms: (4) versus **(5)**, where **(4)** = 100 *IDCC/I(C)*, taken from column 7 in Table 2, and **(5)** = 100 *IDCC*/I(C*)* = 100 **(1)/(2)**

III. 5. Extensions

We complete this paper with two extensions: the use of country exchange rates as normalization factors, and the use of the fractional approach to the treatment of internationally co-authored articles.

III.5.i. Country Exchange Rates

One of the key findings in Crespo *et al.* (2013a, b) is that the analogue of expression $I_1(\pi)$ in Eq. 4 –which in that context captured the effect on citation inequality of differences in citation practices across scientific disciplines– was essentially constant over a large, intermediate quantile interval. This suggested the possibility of estimating an average based measure and its standard deviation (*SD* hereafter) over that interval for every discipline with the purpose of answering the following two questions. Firstly, how many citations in a given discipline are equivalent to, say, 10 citations in the all-fields case? Secondly, how much can we reduce the effect of different citation practices by normalizing the raw citation data with the exchange rates?

In our context, this idea would work as follows. First, mean citations of comparable articles belonging to the same quantile can be used to express the citations in any country in terms of the citations in the field as a whole. Thus, if for a given field we let μ^{π} be the mean citation of all articles in quantile π, then the *exchange rate at quantile* π *for country* p, $e_p(\pi)$, defined by

$$e_p(\pi) = \mu_p^{\pi} / \mu^{\pi},$$

can be seen to answer the following question: how many citations for an article at the degree π of citation impact in country p are equivalent on average to one citation in this field? Next, we use a certain quantile interval $[\pi_m, \pi^M]$ to estimate an *exchange rate* (*ER* hereafter) for every country in that field as

$$e_p = [1/(\pi^M - \pi_m)] [\Sigma_{\pi} e_p(\pi)]. \quad (5)$$

An advantage of this definition is that we can easily compute the associated *SD*, denoted by σ_p. Ideally, one would obtain a small σ_p, and hence a small coefficient of variation $CV_p = \sigma_p / e_p$.

For every field and the all-sciences case we use the interval [50, 97]. The percentage of total citations covered on average over all countries ranges, approximately, from 76% in Economics &

Business to 81% in Engineering and the all-sciences case. The results for *ER*s and *CV*s are in Table 6. For convenience, *ER*s are multiplied by 10. Thus, for example, while the world gets on average a citation impact of 10 citations in Physics Argentina only gets 8.8. In other words, the citation impact of Argentina in Physics is 12% below the average citation impact of the world.

The problem, of course, is that, as we have seen in Section III.4, except for Engineering the effect on citation inequality of differences in citation impact across countries in the raw data tends to decline as we advance towards higher percentiles, forcing *SD*s and *CV*s to be higher than if the curve $I_1(\pi)$ had been relatively constant over a long quantile interval. The results, however, are generally satisfactory. We suggest that countries can be conveniently classified into three groups according to the *CV*. In the first group of highly reliable *ER*s, the *CV* is less than or equal to 0.05, indicating that the *SD* of the *ER* is less than to equal to five per cent of the *ER* itself. In the second group of fairly reliable *ER*s, the *CV* is between 0.05 and 0.10, while in the third group of unreliable *ER*s, the *CV* is grater than 0.10. In four fields –Physics, Chemistry, Clinical Medicine, and Biology & Biochemistry– there are between 22 to 28 reliable country *ER*s, and between three to five unreliable ones. Except in Economics & Business, where only the *ER* of the U.S. is reliable, in the remaining three fields the situation is only somewhat worse, while in the all-sciences case 36 out of 38 *ER*s can be considered reliable.[10]

Furthermore, when we use country exchange rates as normalization factors we achieve practically the same results (available on request) as with country MCs. A possible explanation is that, as we saw in Section II.3, country MCs are reached at approximately the 75th percentile. Since the *ER*s have been estimated for the [50, 97] interval, and the curve $I_1(\pi)$ is moderately declining with π, *ER*s are not very far of country MC ratios, i. e. for any field f and any country p,

[10] These figures can be compared with previous results for exchange rates in fields and sub-fields. In Crespo *et al.* (2013a, b) *CV*s were below 0.05 in 10 out 22 fields, and in 69 out of 219 sub-fields, and between 0.05 and 0.10 in another 10 fields and 118 sub-fields.

the e_p defined in Eq. 5 is close to the ratio $e_p(75) = \mu_p^{75}/\mu^{75}$, which in turn is close to the ratio μ_p/μ, where μ_p and μ are the mean citation of country p and the entire field, respectively.

Table 6. Country Exchange Rates by Field

Country	Physics		Chemistry		Clinical Medicine		Biology & Biochemistry		Materials Sci.		Geosciences		Engineering		Economics & Business		All Sciences	
	ERc	CV	ERc	CV	ERc	CV	ERc	CV	ERc	CV	ERc	CV	ERc	CV	ERc	CV	ERc	CV
ARGENTINA	8.8	0.04	7.1	0.04	8.2	0.12	6.2	0.05	8.2	0.13	6.6	0.06	9.5	0.08	9.6	0.29	7.8	0.03
AUSTRALIA	10.6	0.06	10.9	0.04	12.2	0.02	12.6	0.04	10.0	0.08	12.4	0.03	10.6	0.06	9.6	0.07	11.5	0.02
AUSTRIA	13.1	0.03	11.9	0.03	11.1	0.05	12.1	0.03	10.8	0.25	10.9	0.10	11.4	0.04	10.2	0.06	12.4	0.02
BELGIUM	11.1	0.06	11.8	0.04	14.3	0.01	13.2	0.04	11.8	0.03	11.9	0.05	13.2	0.04	13.8	0.07	13.2	0.02
BRAZIL	7.9	0.03	7.8	0.02	7.4	0.05	6.2	0.05	8.4	0.06	9.4	0.06	9.1	0.05	8.2	0.10	7.3	0.03
CANADA	12.0	0.05	12.5	0.03	14.3	0.02	13.8	0.06	10.8	0.05	11.0	0.03	9.9	0.06	12.5	0.07	12.9	0.02
CHINA	6.9	0.04	7.5	0.11	9.2	0.03	6.2	0.07	8.3	0.10	8.7	0.12	9.0	0.06	12.9	0.09	6.9	0.05
CZECH REPUBLIC	8.5	0.05	8.7	0.02	9.3	0.07	6.9	0.05	9.4	0.04	7.7	0.05	11.3	0.06	12.0	0.12	8.0	0.03
DENMARK	15.0	0.07	15.2	0.02	14.8	0.04	13.3	0.07	12.7	0.03	13.1	0.08	14.8	0.04	11.7	0.06	14.9	0.04
FINLAND	11.7	0.11	10.6	0.05	14.4	0.04	13.1	0.08	10.4	0.07	13.8	0.04	11.6	0.05	10.5	0.06	13.5	0.03
FRANCE	11.0	0.02	10.9	0.02	10.7	0.06	11.9	0.01	11.4	0.03	12.1	0.03	11.3	0.04	12.2	0.07	11.9	0.01
GERMANY	12.5	0.04	11.8	0.02	10.9	0.03	13.5	0.03	10.9	0.06	13.6	0.05	11.6	0.03	10.0	0.06	12.7	0.02
GREECE	10.5	0.06	9.8	0.05	8.3	0.05	8.0	0.06	9.7	0.09	9.8	0.07	8.7	0.06	7.2	0.10	8.4	0.03
HUNGARY	10.5	0.04	8.2	0.04	9.8	0.06	8.9	0.05	8.0	0.07	9.8	0.10	10.8	0.07	8.6	0.08	9.1	0.04
INDIA	7.2	0.12	6.9	0.07	5.0	0.08	5.0	0.04	7.7	0.05	5.3	0.13	7.0	0.08	7.4	0.09	5.5	0.05
IRELAND	10.8	0.05	12.7	0.05	10.8	0.02	12.5	0.42	13.3	0.06	12.1	0.05	10.4	0.05	7.1	0.09	11.4	0.04
ISRAEL	11.9	0.04	13.3	0.03	9.0	0.03	13.2	0.05	14.7	0.05	10.6	0.04	11.0	0.05	13.1	0.07	11.5	0.03
ITALY	11.4	0.02	11.2	0.03	12.2	0.02	10.0	0.03	10.6	0.06	9.9	0.03	10.8	0.06	11.4	0.07	11.9	0.02
JAPAN	9.1	0.04	10.4	0.02	8.9	0.06	10.7	0.03	10.3	0.03	10.2	0.03	8.6	0.06	7.0	0.10	10.1	0.02
MEXICO	6.6	0.04	7.5	0.03	7.4	0.08	6.1	0.04	7.7	0.06	7.8	0.04	7.4	0.08	1.9	0.79	7.2	0.03
NETHERLANDS	13.0	0.04	14.9	0.11	15.4	0.05	13.5	0.05	14.8	0.06	13.2	0.06	12.1	0.04	13.5	0.08	14.9	0.02
NORWAY	11.5	0.05	10.1	0.04	13.2	0.04	11.2	0.04	10.7	0.05	12.1	0.05	11.5	0.07	12.3	0.07	12.1	0.03
POLAND	8.6	0.11	7.0	0.03	8.7	0.07	6.1	0.05	6.7	0.08	7.2	0.06	8.8	0.06	6.7	0.15	7.2	0.04
PORTUGAL	11.2	0.04	9.5	0.08	12.6	0.05	9.4	0.03	11.1	0.08	9.9	0.03	9.6	0.08	5.2	0.20	9.7	0.04
REST OF WORLD	6.3	0.07	6.3	0.06	7.1	0.03	6.8	0.06	6.3	0.07	8.3	0.04	7.1	0.07	5.8	0.17	6.9	0.04
RUSSIA	6.5	0.11	3.6	0.24	2.3	0.56	5.5	0.14	4.0	0.29	3.9	0.27	5.4	0.23	6.8	0.24	4.5	0.18
SINGAPORE	7.5	0.08	12.0	0.02	8.7	0.06	13.0	0.11	12.1	0.04	5.5	0.11	9.8	0.05	9.0	0.12	8.3	0.03
SOUTH AFRICA	6.9	0.08	7.2	0.05	8.3	0.08	8.0	0.04	7.8	0.10	8.8	0.05	7.6	0.07	10.3	0.07	7.6	0.03
SOUTH KOREA	8.4	0.07	8.8	0.09	8.4	0.10	8.6	0.05	10.8	0.03	9.4	0.08	8.4	0.05	11.4	0.08	8.0	0.03
SPAIN	11.6	0.02	11.0	0.07	9.6	0.05	9.9	0.03	10.9	0.05	9.4	0.07	11.2	0.06	9.7	0.07	10.5	0.03
SWEDEN	11.5	0.05	13.5	0.05	13.4	0.05	13.0	0.04	10.9	0.06	12.7	0.02	12.3	0.04	14.3	0.09	13.7	0.04

Country	Physics		Chemistry		Clinical Medicine		Biology & Biochemistry		Materials Sci.		Geosciences		Engineering		Economics & Business		All Sciences	
	ERc	**CV**	**ERc**	**CV**	**ERc**	**CV**	**ERc**	**CV**	**ERc**	**CV**	**ERc**	**CV**	**ERc**	**CV**	**ERc**	**CV**	**ERc**	**CV**
SWITZERLAND	16.7	0.01	15.4	0.02	14.3	0.03	16.2	0.06	15.2	0.04	14.6	0.06	15.3	0.04	14.4	0.12	16.6	0.01
TAIWAN	7.7	0.07	9.0	0.03	7.6	0.06	7.8	0.06	10.2	0.03	9.5	0.05	8.5	0.06	7.8	0.09	7.5	0.03
TURKEY	6.8	0.04	6.8	0.04	4.5	0.11	5.8	0.07	8.3	0.10	8.6	0.09	9.7	0.05	11.2	0.11	5.9	0.05
UK	12.0	0.02	12.3	0.02	12.5	0.02	14.5	0.02	11.3	0.03	12.8	0.03	9.8	0.05	12.8	0.06	13.4	0.02
UKRAINE	4.8	0.06	4.3	0.21	3.2	0.19	5.9	0.12	3.0	0.27	4.4	0.22	5.8	0.12			4.0	0.13
USA	14.3	0.01	15.9	0.02	14.4	0.04	16.3	0.03	14.3	0.08	14.1	0.04	11.8	0.04	17.3	0.03	15.1	0.02
ZEU	7.7	0.04	5.9	0.07	7.5	0.06	5.0	0.10	6.4	0.08	8.7	0.10	7.5	0.08	4.7	0.19	6.2	0.05

[1]**ZEU** = European countries outside the European Union (excluding Luxembourg) before the 2004 accession.

[2]**RW** = Rest of the World = Central America, and South America except Argentina, and Brazil; Asia, except China, India, Japan, Singapore, South Korea, and Taiwan; Armenia, Azerbaijan, Kazakhstan, Kyrgyzstan, Republic of Georgia, Tajikistan, Turkmenistan, and Uzbekistan; remaining Islamic countries; remaining African countries except South Africa; Oceania except Australia.

III.5.ii. The Fractional Approach

The treatment of internationally co-authored articles according to a fractional strategy is discussed in the Appendix II. To save space, the main results for Physics and the all-sciences case are included in Table 7. The only difference is that the importance of the *IDCC* term relative to overall citation inequality for the raw data in the fractional case is slightly greater than in the multiplicative case. In Physics, 5.1% versus 4.7%, and in all-sciences, 5.6% versus 5.4%. Country MC normalization leads to practically the same graphical and numerical results (similar results for the remaining fields are available on request). This shows that the choice of strategies for dealing with internationally co-authored articles does not essentially alter the findings of the paper.

Table 7. Total Citation Inequality Decomposition Before and After Country Mean Normalization. The Fractional Case

	IDCC	$I_1(C)$	100(1)/(2)	*IDCC**	$I_1(C^*)$	100 (1)/(2)	100 [(1) – (4)]/(1)
PHYSICS	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
A. All articles	0.046	0.90	5.1	0.008	0.87	1	81.3
B. Domestic articles	0.094	0.93	10.1	0.011	0.85	1.3	88.2
C. International articles	0.017	0.83	2.1	0.005	0.81	0.6	73.0
ALL SCIENCES							
A. All articles	0.048	0.88	5.6	0.008	0.83	0.9	83.8
B. Domestic articles	0.077	0.88	8.7	0.011	0.82	1.3	85.8
C. International articles	0.024	0.79	3.1	0.003	0.77	0.3	88.9

Reduction measured in relative terms: (3) = 100 *(IDCC - IDCC*)/IDCC*

Reduction measured in relative terms: (4) versus **(5)**, where **(4)** = 100 ***IDCC/I(C)***, taken from column 7 in Table 2, and **(5)** = 100 ***IDCC*/I(C*)*** = 100 **(1)/(2)**

IV. CONCLUSIONS, DISCUSSION, AND EXTENSIONS

IV. 1. Conclusions

This paper, which has adopted a multiplicative strategy in the treatment of internationally co-authored articles, has achieved the following three aims.

Firstly, using the CSS technique and a robust index of skewness, we have presented convincing evidence concerning some stylized features shared by the shape of country citation distributions within eight broad fields and the all-sciences case for articles published in 1998-2003 with a five-year citation window.

Secondly, we have used the measurement framework introduced in Crespo *et al.* (2013a) to assess the effect on total citation inequality of differences in citation impact across countries in a given field and the all-sciences case. Using an additively decomposable citation inequality index, we have shown how this effect can be measured by a between-group term, denoted *IDCC*, in a certain partition of the dataset into countries and percentiles. The *IDCC* term represents, approximately, between 3.7% and 10.1% of total citation inequality in the different fields, and 5.4% in the all-sciences case.

Thirdly, we have studied the consequences of normalizing raw citation counts by country mean citations. The results vary by field. In absolute terms, the reduction of the *IDCC* term ranges from 53.3% to 68.3% in four fields, from 77% to 86% in the other four, and it is of 79.5% in the all-sciences case. In relative terms, the importance of the *IDCC* term with respect total citation inequality goes down after normalization by a factor ranging from 1.9 to 6.5, with 4.5 in the all-sciences case.

Additionally, we have found that the effect on citation inequality of differences in citation impact across countries exhibits in most cases a declining pattern as we proceed towards higher percentiles. Generally, this is partly explained by the increasing presence of international articles in higher percentiles; the same declining pattern of the effect on citation inequality of differences in citation impact across countries in international articles in most fields, and even the existence of a similar declining pattern among domestic articles in five fields.

Exchange rates (*ER*s) for most countries are rather reliably estimated in the all-sciences case and in seven out of eight fields. Using country *ER*s as normalization factors leads to very similar results as those achieved with country MC normalization. Finally, treating international articles according to a fractional approach does not essentially alter the conclusions obtained with a multiplicative approach.

IV. 2. Implications

As we have seen, the shape of country citation distributions in the all-sciences case and in the eight fields studied in this paper country citation distributions are highly skewed. In spite of the fact that this skewness is somewhat different across countries in each field, there is enough similarity for normalization by country MCs (or *ER*s) to generate important reductions in the *IDCC* term –albeit in different degrees across fields. These facts can be interpreted as indicating that, to a large extent, country citation distributions behave as if they differ by a relatively constant scale factor over a large part of their support.[11] A convenient practical consequence is that country differences in citation impact can be well summarized by average-based indicators.

However, the skewness of citation distributions have lead many authors –including ourselves in Albarrán *et al.* (2011a, b), for example– to suggest that the mean, or any central-tendency statistic, may not provide a good representation of citation distributions. Consistently with this view, Albarrán and Ruiz-Castillo (2012) used the same dataset analyzed in this paper to make international comparisons according to the MC and an indicator that focus on the upper tail of citation distributions –the $PP_{top\ 10\%}$ indicator, defined as the percentage of an institution's scientific output included in the set formed by the 10% of the most highly cited papers in their respective scientific fields.[12] In principle, before knowing about the similarity of country citation distributions unveiled in this paper, we expected large differences between the two rankings for a partition of the world (into 39 countries and eight geographical areas) very similar to the one used in this paper. However, it was found that that the rank correlation coefficient between MC and $PP_{top\ 10\%}$ taking together the results for the 22 fields distinguished by Thomson and Reuters is 0.93. As a matter of fact, except for Engineering for which this correlation coefficient is 0.70, in

[11] Of course, the fact that countries have identical citation distributions, with each pair differing by a common scale factor over their entire support, implies that after normalization by country MCs the *IDCC* term would be reduced to zero. However, it would be easy to construct an example establishing that the opposite is not true. Thus, even a very large reduction of the *IDCC* term is only a necessary but not a sufficient condition for country differences to be mostly attributed to constant scale factors over the entire support.

[12] For a defense of the $PP_{top\ 10\%}$ indicator within the rank percentile approach, as well as other references to this literature, see Bornmann and Marx, 2013.

the remaining cases it ranges from 0.95 in Materials Science and Economics & Business to 0.99 for Chemistry and the all-sciences case. We claim that the two key findings of this paper, namely, (i) the similarity between the shapes of country citation distributions where country MCs are reached at approximately the 75th percentile, and (ii) the evidence in favor of the fact that a good part of the differences between country citation distributions can be accounted by scale factors, describe an empirical scenario in which it is not surprising that the MC and the $PP_{top\ 10\%}$ focusing in the last 10% of the upper tail of citation distributions provide very similar rankings.

Quite apart from the fact that our results do not apply with the same strength to the eight fields studied in this paper, it should be noted that before the implications of our analysis become acceptable for international comparisons our methods must be applied to other scientific fields different from those considered here, and the robustness of the results must be investigated with other datasets: other publication years; other citation windows, and other data sources different from Thomson Reuters.

IV. 3. Extensions

Before closing this paper, we would like to comment in two possible extensions of our work. Firstly, the Centre for Science and Technology Studies at Leiden University in the Netherlands considers at present the $PP_{top\ 10\%}$ indicator as the most important impact indicator in the influential Leiden Ranking.[13] However, as in Albarrán and Ruiz-Castillo (2012), it has been found that there is a strong, more or less linear relationship between the $PP_{top\ 10\%}$ and the MNCS (see Figure 2 in Waltman *et al.*, 2012) for the 500 universities in the 2011/2012 edition of the Leiden Ranking. Therefore, it is worthwhile to apply the methods used in this paper to the case of the 500 universities considered in the Leiden Ranking. The conjecture is that the similarity across highly skewed university citation distributions, and the important role of scale factors in

[13] SCImago, a research group from the *Consejo Superior de Investigaciones Científicas*, University of Granada, Extremadura, Carlos III (Madrid) and Alcalá de Henares in Spain, also uses the $PP_{top\ 10\%}$ indicator. The *Leiden Ranking 2011/2012* (http://www.leidenranking.com/methodology.aspx) is based on publications in the sciences and the social sciences in Thomson Reuters' Web of Science database in the period 2005-2009, while the *SCImago Institutions Rankings* (SIR) 2011 World Report (http://www.scimagoir.com/pdf/sir_2011_world_report.pdf) is based on the Scopus® database (Elsevier B.V.).

accounting for their differences in citation impact may go a long way towards understanding why the university rankings by the MNCS and the $PP_{top\ 10\%}$ is very similar indeed.

Secondly, there is another finding of the paper that needs to be addressed. As we have seen in Section III.4, differences in citation impact across countries after normalization within most fields and in the all-sciences case tend to raise 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 $PP_{top\ 10\%}$ indicators with other instruments that highlight the behavior of citation distributions over the last few percentiles. Given the important role of extreme observations in citation distributions, the robustness of alternative high-impact indicators to these extreme situations will be an important element in the discussion.[14]

[14] See Li and Ruiz-Castillo (2013) for the large impact that extreme observations have on citation inequality in certain fields in the dataset used in this paper. See also Waltman *et al.* (2012) for the case of the University of Göttingen in the 2010/2011 Leiden Ranking, where a single observation dramatically affects an average-based indicator such as the MNCS.

REFERENCES

Aksnes, D., Schneider, J., and Gunnarsson, M. (2012), "Ranking National Research Systems by Citation Indicators. A Comparative Analysis Using Whole and Fractionalised Counting Methods", *Journal of Informetrics*, **6**: 36-43.

Albarrán, P., J. Crespo, I. Ortuño, and J. Ruiz-Castillo (2010), "A Comparison of the Scientific Performance of the U.S. and Europe at the Turn of the 21st Century", *Scientometrics*, **85**: 329-344.

Albarrán, P. and J. Ruiz-Castillo (2011), "References Made and Citations Received By Scientific Articles", *Journal of the American Society for Information Science and Technology*, **62**: 40-49.

Albarrán, P., J. Crespo, I. Ortuño, and J. Ruiz-Castillo (2011), "High- and Low-impact Citation Measures: Empirical Applications", *Journal of Informetrics*, **5**: 122-145.

Albarrán, P. and J. Ruiz-Castillo (2012), "The Measurement of Scientific Excellence Around the World", Working Paper, Economic Series 12-08, Universidad Carlos III (http://hdl.handle.net/10016/13896).

Albarrán, P., J. Crespo, I. Ortuño, and J. Ruiz-Castillo (2011a), "The Skewness of Science In 219 Sub-fields and A Number of Aggregates", *Scientometrics*, **88**: 385-397.

Albarrán, P., I. Ortuño, and J. Ruiz-Castillo (2011b). "The Measurement of Low- and High-impact In Citation Distributions: Technical Results", *Journal of Informetrics*, **5**: 48-63.

Albarrán, P., I. Ortuño and J. Ruiz-Castillo (2011c), "High- and Low-impact Citation Measures: Empirical Applications", *Journal of Informetrics*, **5**: 122-145.

Anderson, J., P. Collins, J. Irvine, P. Isard, B. Martin, F. Narin, and K. Stevens (1988), "On-line Approaches to Measuring National Scientific Output: A Cautionary tale", *Science and Public Policy*, **15**: 153-161.

Bornmann, L., and Mutz, R. (2011). "Further Steps Towards an Ideal Method of Measuring Citation Performance: The Avoidance of Citation (Ratio) Averages in Field-normalization", *Journal of Informetrics*, **5**: 228-230.

Bornmann L. & Marx, W. (2013), "How good is research really?", *EMBO reports*, **14**: 226-230.

Bowley, A.L., 1920. *Elements of Statistics*. Scribner's, New York.

Carrasco, R., and Ruiz-Castillo, J. (2012), "The Evolution of the Scientific Productivity of Highly Productive Economists", in press in *Economic Inquiry*; Working Paper 12-16, Universidad Carlos III, (http://hdl.handle.net/10016/14527).

Crespo, J. A., Li, Y., and Ruiz-Castillo, J. (2013a), "Differences in Citation Practices Across Scientific Fields", *PLoS ONE*, **7**(3): e33833.

Crespo, J. A., Li, Yunrong, Herranz, N., and Ruiz-Castillo, J. (2012b), "Field Normalization at Different Aggregation Levels", in press in *Journal of the American Society for Information Science and Technology*; Working Paper 12-022, Universidad Carlos III (htpp:/hdl.handle.net/10016/15344).

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

Herranz, N., and Ruiz-Castillo, J. (2012), "Multiplicative and Fractional Strategies When Journals Are Assigned to Several Sub-fields", in press, *Journal of the American Society for Information Science and Technology*, **63**: 2195–2205 (2012).

Hinkley, D.V. (1975), "On power transformations to symmetry", *Biometrika*, **62**: 101–111.

Kendall, M.G., and Stuart, A. (1977), *The Advanced Theory of Statistics*, vol. 1. Griffin, London.

Kim, T-H, and White, A. (2004), "On More Robust Estimation of Skewness and Kurtosis", *Finance Research Letters*, **1**: 56-73.

King, D. (2004), "The Scientific Impact of Nations", *Nature*, **430**: 311-316.

Leydesdorff, L., and Bornmann, L. (2011), "Integrated Impact Indicators (I3) compared with Impact Factors (Ifs): An alternative research design with policy implications", *Journal of the American Society for Information Science and Technology*, **62**: 2133-2146.

Leydesdorff, L., Bornmann, L., Mutz, R., and Opthof, T. (2011), "Turning the tables on citation analysis one more time: Principles for comparing sets of documents", *Journal of the American Society for Information Science and Technology*, **62**: 1370-1381.

Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., and de Nooye, W. (2012), "Field-normalized Impact Factors: A Comparison of Rescaling *versus* Fractionally Counted IFs", in press, *Journal of the American Society for Information Science and Technology.*

Li, Y., Radicchi, F., Castellano, C., and Ruiz-Castillo, J. (2013), "Quantitative Evaluation of Alternative Field Normalization Procedures", Working Paper 13-5, Universidad Carlos III, April 2013, (http://hdl.handle.net/10016/16741).

Li, Y., and Ruiz-Castillo, J. (2013), "The Impact of Extreme Observations in Citation Distributions", Working Paper 13-8, Universidad Carlos III, May 2013, (http://hdl.handle.net/10016/16919).

May, R. (1997), "The Scientific Wealth of Nations", *Science*, **275**: 793-796.

National Science Board (2010), *Science and Engineering Indicators 2010*, Appendix Tables, Arlington, VA, U.S.: National Science Foundation.

Radicchi, F., and Castellano, C. (2012a), "A Reverse Engineering Approach to the Suppression of Citation Biases Reveals Universal Properties of Citation Distributions", *Plos One*, **7**, e33833, 1-7.

Radicchi, F., and Castellano, C. (2012b), "Testing the fairness of citation indicators for comparisons across scientific domains: The case of fractional citation counts", *Journal of Informetrics*, **6**: 121-130.

Rousseau (2011), "Basic Properties of Both Percentile Rank Scores and the I3 Indicator", in press *Journal of the American Society for Information Science* (DOI: 10.10002/asi.2168).

Schubert, A., W. Glänzel and T. Braun (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.

Tijssen, R. M., Visser, M., & van Leeuwen, T. (2002), "Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference?", *Scientometrics*, **54**: 381–397.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., van Leeuwen, T. H., van Raan, A. F. J., Visser, M. S.. and Wouters, P. (2012), The Leiden Ranking 2011/2012: Data Collection, Indicators, and Interpretation, *Journal of the American Society for Information Science*, **63**: 2419-2432.

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

APPENDIX I

Table A. Number of Articles and Mean Citations of the Original Dataset and the Geographically Extended Count

		Original Dataset			Geographically Extended Count		
		Number of Articles (1)	% (2)	MC (3)	Number of Articles (4)	% (5)	MC (6)
A.	**LIFE SCIENCES**	**1,806,398**	**40.4**		**2,156,080**	**39.6**	
(1)	Biology & Biochemistry	275,568	6.2	12.5	338,858	6.2	13.1
(2)	Clinical Medicine	947,261	21.2	9.7	1,102,367	20.2	11.1
(3)	Immunology	60,875	1.4	16.0	78,715	1.4	16.8
(4)	Microbiology	73,039	1.6	11.4	91,874	1.7	12.0
(5)	Molecular Biology & Genetics	122,233	2.7	20.4	159,038	2.9	22.0
(6)	Neuroscience & Behav. Science	140,686	3.1	13.7	171,280	3.1	14.5
(7)	Pharmacology & Toxicology	76,728	1.7	8.0	89,933	1.7	8.3
(8)	Psychiatry & Psychology	110,008	2.5	7.0	124,015	2.3	7.4
B.	**PHYSICAL SCIENCES**	**1,282,919**	**28.7**		**1,644,936**	**30.2**	
(9)	Chemistry	550,147	12.3	7.6	651,956	12.0	7.9
(10)	Computer Science	98,727	2.2	3.0	117,843	2.2	3.2
(11)	Mathematics	117,496	2.6	2.4	149,174	2.7	2.6
(12)	Physics	456,144	10.2	6.9	626,304	11.5	7.8
(13)	Space Science	60,405	1.4	11.0	99,659	1.8	12.8
C.	**OTHER NATURAL SCIENCES**	**1,150,428**	**25.7**		**1,390,734**	**25.5**	
(14)	Agricultural Sciences	82,837	1.9	4.9	94,141	1.7	5.1
(15)	Engineering	356,269	8.0	3.2	421,332	7.7	3.4
(16)	Environment & Ecology	109,826	2.5	7.1	134,942	2.5	7.6
(17)	Geoscience	120,059	2.7	6.7	162,952	3.0	7.5
(18)	Materials Science	199,364	4.5	4.5	236,156	4.3	4.7
(19)	Multidisciplinary	20,672	0.5	3.2	23,563	0.4	3.3
(20)	Plant & Animal Science	261,401	5.8	5.1	317,648	5.8	5.5
D.	**SOCIAL SCIENCES**	**232,587**	**5.2**		**258,559**	**4.7**	
(21)	Economics & Business	63,380	1.4	3.9	75,687	1.4	4.1
(22)	Social Sciences, General	169,207	3.8	3.3	182,872	3.4	3.5
ALL SCIENCES		**4,472,332**	**100.0**		**5,450,309**	**100.0**	
Average Values		203,288		11.1	247,741		12.1
Standard Deviation		214,385		4.7	254,804		5.1

Table B. Number of Articles In Physics and the All-fields Case By Country and Geographical Area

	Number of Articles, All Fields				Number of Articles, Physics			
	Extended Count		Original Dataset		Extended Count		Original Dataset	
Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ARGENTINA	25939	0.5	20550.6	0.5	3367	0.5	2376.2	0.5
AUSTRALIA	126072	2.3	100870.4	2.3	7616	1.2	5432.4	1.2
AUSTRIA	43009	0.8	31202.2	0.7	4767	0.8	2728.6	0.6
BELGIUM	60038	1.1	42086.1	0.9	6382	1.0	3785.0	0.8
BRAZIL	66556	1.2	54335.9	1.2	10372	1.7	7745.3	1.7
CANADA	195938	3.6	154408.5	3.5	11861	1.9	7851.0	1.7
CHINA	197462	3.6	172272.0	3.9	35193	5.6	30548.3	6.7
CZECH REPUBLIC	26542	0.5	19430.7	0.4	4208	0.7	2411.7	0.5
DENMARK	45908	0.8	32547.6	0.7	4401	0.7	2444.3	0.5
FINLAND	43769	0.8	33036.7	0.7	3694	0.6	2149.8	0.5
FRANCE	282729	5.2	218546.5	4.9	39253	6.3	26291.7	5.8
GERMANY	390873	7.2	305166.7	6.8	56148	9.0	37186.7	8.2
GREECE	30917	0.6	24175.6	0.5	3516	0.6	2122.6	0.5
HUNGARY	24398	0.4	17257.4	0.4	3293	0.5	1748.2	0.4
INDIA	107025	2.0	96628.7	2.2	13847	2.2	11498.1	2.5
IRELAND	16005	0.3	11765.9	0.3	1280	0.2	800.8	0.2
ISRAEL	55837	1.0	43346.5	1.0	7574	1.2	5168.4	1.1
ITALY	190078	3.5	150122.5	3.4	25702	4.1	17247.4	3.8
JAPAN	431828	7.9	387796.2	8.7	63314	10.1	54488.4	11.9
MEXICO	29858	0.5	22664.0	0.5	5004	0.8	3574.0	0.8
NETHERLANDS	111959	2.1	83484.5	1.9	9118	1.5	5763.7	1.3
NORWAY	29511	0.5	21545.7	0.5	1568	0.3	838.1	0.2
POLAND	61172	1.1	47036.3	1.1	12877	2.1	8203.8	1.8
PORTUGAL	20173	0.4	14398.5	0.3	2601	0.4	1558.9	0.3
RW[2]	178240	3.3	131966.1	3.0	14516	2.3	10247.7	2.2
RUSSIA	157349	2.9	129175.9	2.9	45144	7.2	33558.5	7.4
SINGAPORE	22834	0.4	18505.1	0.4	2914	0.5	2360.2	0.5
SOUTH AFRICA	21994	0.4	17217.0	0.4	1150	0.2	742.0	0.2
SOUTH KOREA	89445	1.6	77457.9	1.7	15219	2.4	12424.0	2.7
SPAIN	135317	2.5	108655.6	2.4	14137	2.3	9431.4	2.1
SWEDEN	89902	1.6	66662.6	1.5	8326	1.3	5057.7	1.1
SWITZERLAND	80669	1.5	55295.6	1.2	11580	1.8	6404.0	1.4
TAIWAN	62928	1.2	56568.1	1.3	7906	1.3	6551.7	1.4
TURKEY	40018	0.7	35816.1	0.8	2454	0.4	1937.2	0.4
UK	397488	7.3	316042.9	7.1	33366	5.3	22861.2	5.0
UKRAINE	24631	0.5	18990.2	0.4	7638	1.2	5399.0	1.2
U.S.	1463587	26.8	1279285.2	28.6	112340	17.9	87246.3	19.1
ZEU[1]	74445	1.4	56017.3	1.3	12658	2.0	7959.4	1.7
TOTAL	**5,452,443**	**100**	**4,472,331**	**100**	**626,304**	**100**	**456,144**	**100**

[1]**ZEU** = European countries outside the European Union (excluding Luxembourg) before the 2004 accession.

[2]**RW** = Rest of the World = Central America, and South America except Argentina, and Brazil; Asia, except China, India, Japan, Singapore, South Korea, and Taiwan; Armenia, Azerbaijan, Kazakhstan, Kyrgyzstan, Republic of Georgia, Tajikistan, Turkmenistan, and Uzbekistan; remaining Islamic countries; remaining African countries except South Africa; Oceania except Australia.

Table C. Number of Articles In the Extended Count

Countries	Chemistry		Clinical Medicine		Biology & Biochemistry		Materials Sci		Geosciences		Engineering		Economics & Business	
	Articles	%	Articles	%	Articles	%	Articles	%	Articles	%	Articles	%	Articles	%
ARGENTINA	3763	0.6	3589	0.3	2248	0.7	842	0.4	1033	0.6	1033	0.2	173	0.2
AUSTRALIA	9018	1.4	26520	2.4	7324	2.2	3408	1.4	6167	3.8	8257	2.0	2508	3.3
AUSTRIA	4206	0.6	13026	1.2	2567	0.8	1550	0.7	1196	0.7	2457	0.6	448	0.6
BELGIUM	6632	1.0	14638	1.3	3826	1.1	2042	0.9	1153	0.7	4114	1.0	968	1.3
BRAZIL	8046	1.2	9832	0.9	3996	1.2	2674	1.1	1506	0.9	4088	1.0	223	0.3
CANADA	15970	2.4	39400	3.6	13003	3.8	5435	2.3	9107	5.6	15193	3.6	3748	5.0
CHINA	48597	7.5	14524	1.3	7367	2.2	22883	9.7	5659	3.5	21974	5.2	1393	1.8
CZECH REP.	4978	0.8	2357	0.2	1946	0.6	1587	0.7	784	0.5	1580	0.4	377	0.5
DENMARK	3789	0.6	11183	1.0	3998	1.2	782	0.3	1604	1.0	2154	0.5	711	0.9
FINLAND	3338	0.5	12050	1.1	2588	0.8	1537	0.7	989	0.6	2742	0.7	509	0.7
FRANCE	35875	5.5	56390	5.1	18756	5.5	12328	5.2	10695	6.6	18351	4.4	2424	3.2
GERMANY	52100	8.0	85795	7.8	22521	6.6	18829	8.0	11090	6.8	24225	5.8	2731	3.6
GREECE	3286	0.5	7445	0.7	1253	0.4	1149	0.5	997	0.6	3731	0.9	379	0.5
HUNGARY	4811	0.7	2909	0.3	1810	0.5	896	0.4	404	0.2	1559	0.4	106	0.1
INDIA	24399	3.7	9360	0.8	5374	1.6	7915	3.4	3257	2.0	9539	2.3	380	0.5
IRELAND	1359	0.2	3745	0.3	929	0.3	585	0.2	358	0.2	1147	0.3	318	0.4
ISRAEL	4584	0.7	13064	1.2	3342	1.0	1400	0.6	853	0.5	3895	0.9	932	1.2
ITALY	21291	3.3	45375	4.1	12072	3.6	5273	2.2	5428	3.3	15692	3.7	1452	1.9
JAPAN	66148	10.1	91200	8.3	33671	9.9	28783	12.2	7109	4.4	36121	8.6	1147	1.5
MEXICO	2897	0.4	3404	0.3	1768	0.5	1484	0.6	952	0.6	1973	0.5	189	0.3
NETHERLANDS	9961	1.5	29856	2.7	6418	1.9	2579	1.1	3020	1.9	6937	1.7	2331	3.1
NORWAY	1940	0.3	7363	0.7	1640	0.5	652	0.3	2029	1.2	1548	0.4	597	0.8
POLAND	13987	2.1	4734	0.4	4108	1.2	3779	1.6	942	0.6	4539	1.1	116	0.2
PORTUGAL	3245	0.5	2075	0.2	1259	0.4	1662	0.7	416	0.3	2146	0.5	212	0.3
RW[2]	23018	3.5	33766	3.1	7209	2.1	6271	2.7	6491	4.0	14425	3.4	1520	2.0
RUSSIA	33416	5.1	6542	0.6	6080	1.8	9811	4.2	9864	6.1	14460	3.4	236	0.3
SINGAPORE	2569	0.4	2862	0.3	786	0.2	2478	1.0	109	0.1	5405	1.3	478	0.6
SOUTH AFRICA	1664	0.3	3971	0.4	938	0.3	520	0.2	1321	0.8	1170	0.3	284	0.4
SOUTH KOREA	15364	2.4	10385	0.9	5533	1.6	9116	3.9	864	0.5	12316	2.9	637	0.8
SPAIN	21950	3.4	25100	2.3	7782	2.3	5307	2.2	2971	1.8	8315	2.0	1509	2.0

Countries	Chemistry		Clinical Medicine		Biology & Biochemistry		Materials Sci		Geosciences		Engineering		Economics & Business	
	Articles	%	Articles	%	Articles	%	Articles	%	Articles	%	Articles	%	Articles	%
SWEDEN	7722	1.2	23961	2.2	7164	2.1	3563	1.5	2269	1.4	5183	1.2	1150	1.5
SWITZERLAND	9311	1.4	19053	1.7	5227	1.5	2280	1.0	2798	1.7	4939	1.2	732	1.0
TAIWAN	8522	1.3	11244	1.0	2469	0.7	4450	1.9	1053	0.6	11682	2.8	596	0.8
TURKEY	4566	0.7	14737	1.3	1634	0.5	1426	0.6	1022	0.6	3906	0.9	338	0.4
UK	35852	5.5	93513	8.5	24924	7.4	13008	5.5	13404	8.2	28192	6.7	9205	12.2
UKRAINE	4765	0.7	483	0.0	468	0.1	4256	1.8	423	0.3	2983	0.7		
U.S.	114067	17.5	339325	30.8	100065	29.5	37433	15.9	41894	25.7	104536	24.9	33727	44.7
ZEU[1]	14950	2.3	6604	0.6	4770	1.4	6183	2.6	1721	1.1	7009	1.7	603	0.8
TOTAL	651956		1101380		338833		236156		162952		419516		75387	

[1]**ZEU** = European countries outside the European Union (excluding Luxembourg) before the 2004 accession.

[2]**RW** = Rest of the World = Central America, and South America except Argentina, and Brazil; Asia, except China, India, Japan, Singapore, South Korea, and Taiwan; Armenia, Azerbaijan, Kazakhstan, Kyrgyzstan, Republic of Georgia, Tajikistan, Turkmenistan, and Uzbekistan; remaining Islamic countries; remaining African countries except South Africa; Oceania except Australia.

Table D. The Skewness of Country Citation Distributions in Eight Fields and the All-fields Case. Average (Standard Deviation), and Coefficient of Variation over 38 Countries of the conventional measure of skewness, SK_1, and the Groeneveld and Meeden (1984) robust measure of skewness, SK_2

PHYSICS	SK1	SK2
USA	26.61	0.63
GERMANY	49.61	0.54
JAPAN	55.04	0.65
FRANCE	55.74	0.60
UK	46.97	0.66
RUSSIA	74.05	0.64
ITALY	51.33	0.61
CHINA	20.81	0.68
SWITZERLAND	43.07	0.60
SPAIN	56.50	0.65
CANADA	46.35	0.67
SOUTH KOREA	26.08	0.61
NETHERLANDS	7.60	0.57
POLAND	23.51	0.62
INDIA	61.58	0.72
SWEDEN	51.61	0.65
ZEU	6.67	0.73
ISRAEL	7.53	0.65
BRAZIL	8.83	0.60
AUSTRALIA	5.69	0.58
BELGIUM	16.07	0.61
DENMARK	7.69	0.55
TAIWAN	15.96	0.75
AUSTRIA	7.46	0.69
FINLAND	40.33	0.68
GREECE	4.95	0.58
UKRAINE	8.75	0.79
HUNGARY	7.18	0.58
CZECH REPUBLIC	5.48	0.57
MEXICO	5.13	0.66
ARGENTINA	6.79	0.63
PORTUGAL	6.10	0.61
SINGAPORE	3.25	0.51
NORWAY	6.92	0.65
TURKEY	8.45	0.71
IRELAND	4.30	0.59
SOUTH AFRICA	8.41	0.72
REST OF WORLD	8.00	0.64
Average	23.59	0.64
Standard Deviation	21.37	0.06
Coefficient of Variation	0.91	0.09

CLINICAL MEDICINE	SK1	SK2
USA	25.01	0.67
GERMANY	16.37	0.69

JAPAN	11.64	0.61
FRANCE	15.98	0.77
UK	18.79	0.68
RUSSIA	13.47	1.00
ITALY	14.81	0.68
CHINA	20.39	0.63
SWITZERLAND	12.70	0.65
SPAIN	19.64	0.76
CANADA	17.14	0.60
SOUTH KOREA	7.20	0.59
NETHERLANDS	12.77	0.66
POLAND	10.90	0.75
INDIA	17.91	0.66
SWEDEN	15.68	0.58
ZEU	19.86	0.67
ISRAEL	15.90	0.65
BRAZIL	25.58	0.71
AUSTRALIA	19.13	0.69
BELGIUM	14.17	0.66
DENMARK	16.18	0.64
TAIWAN	26.98	0.54
AUSTRIA	11.78	0.62
FINLAND	16.07	0.63
GREECE	9.29	0.57
UKRAINE	3.95	0.63
HUNGARY	18.09	0.70
CZECH REPUBLIC	10.16	0.76
MEXICO	23.49	0.68
ARGENTINA	15.39	0.73
PORTUGAL	5.38	0.67
SINGAPORE	22.96	0.65
NORWAY	17.77	0.68
TURKEY	13.37	0.62
IRELAND	16.62	0.61
SOUTH AFRICA	17.56	0.74
REST OF WORLD	22.55	0.62
Average	16.12	0.67
Standard Deviation	5.30	0.08
Coefficient of Variation	0.33	0.12

ECONOMICS & BUSINESS	**SK1**	**SK2**
USA	7.02	0.71
GERMANY	10.38	0.73
JAPAN	6.45	0.58
FRANCE	5.19	0.80
UK	8.05	0.58
RUSSIA	5.95	1.00
ITALY	3.37	0.47
CHINA	4.37	0.58

SWITZERLAND	7.21	0.62
SPAIN	7.73	0.77
CANADA	3.58	0.53
SOUTH KOREA	2.98	0.79
NETHERLANDS	3.12	0.59
POLAND	1.93	0.51
INDIA	2.68	0.58
SWEDEN	3.77	0.63
ZEU	6.52	1.00
ISRAEL	4.34	0.60
BRAZIL	3.18	0.62
AUSTRALIA	3.86	0.75
BELGIUM	3.02	0.57
DENMARK	4.47	0.48
TAIWAN	11.39	0.66
AUSTRIA	3.51	0.77
FINLAND	3.64	0.73
GREECE	6.53	0.57
UKRAINE	3.39	1.00
HUNGARY	2.78	0.74
CZECH REPUBLIC	6.72	1.00
MEXICO	8.55	0.69
ARGENTINA	4.12	1.00
PORTUGAL	3.08	0.72
SINGAPORE	5.87	0.52
NORWAY	3.95	0.50
TURKEY	3.87	0.58
IRELAND	4.96	0.77
SOUTH AFRICA	1.73	0.38
REST OF WORLD	5.97	0.68
Average	4.98	0.68
Standard Deviation	2.25	0.16
Coefficient of Variation	0.45	0.24

BIOLOGY & BIOCHEMISTRY	**SK1**	**SK2**
USA	23.89	0.51
GERMANY	27.86	0.51
JAPAN	9.10	0.60
FRANCE	35.80	0.57
UK	14.76	0.55
RUSSIA	9.10	0.69
ITALY	4.06	0.55
CHINA	10.67	0.59
SWITZERLAND	17.18	0.54
SPAIN	6.65	0.56
CANADA	4.39	0.44
SOUTH KOREA	3.51	0.56
NETHERLANDS	6.58	0.51
POLAND	6.03	0.66

INDIA	4.87	0.50
SWEDEN	21.35	0.50
ZEU	8.21	0.66
ISRAEL	4.87	0.56
BRAZIL	3.63	0.63
AUSTRALIA	11.46	0.54
BELGIUM	3.75	0.55
DENMARK	3.96	0.52
TAIWAN	3.72	0.52
AUSTRIA	5.15	0.51
FINLAND	3.27	0.52
GREECE	3.73	0.50
UKRAINE	2.38	0.70
HUNGARY	3.83	0.57
CZECH REPUBLIC	4.33	0.54
MEXICO	5.73	0.61
ARGENTINA	3.88	0.64
PORTUGAL	4.41	0.54
SINGAPORE	10.74	0.48
NORWAY	7.14	0.57
TURKEY	5.61	0.60
IRELAND	21.96	0.69
SOUTH AFRICA	3.84	0.48
REST OF WORLD	9.38	0.66
Average	8.97	0.56
Standard Deviation	7.77	0.07
Coefficient of Variation	0.87	0.12

CHEMISTRY	**SK1**	**SK2**
USA	38.16	0.53
GERMANY	91.08	0.55
JAPAN	5.11	0.61
FRANCE	110.75	0.51
UK	73.50	0.45
RUSSIA	5.42	0.61
ITALY	24.79	0.56
CHINA	5.31	0.73
SWITZERLAND	17.69	0.51
SPAIN	3.24	0.39
CANADA	91.70	0.49
SOUTH KOREA	5.74	0.65
NETHERLANDS	67.75	0.53
POLAND	4.78	0.51
INDIA	4.36	0.68
SWEDEN	3.53	0.43
ZEU	4.13	0.60
ISRAEL	3.07	0.50
BRAZIL	4.95	0.60
AUSTRALIA	5.84	0.51

BELGIUM	7.38	0.42
DENMARK	4.75	0.51
TAIWAN	3.67	0.52
AUSTRIA	3.69	0.56
FINLAND	3.41	0.51
GREECE	3.56	0.45
UKRAINE	4.36	0.70
HUNGARY	3.47	0.41
CZECH REPUBLIC	3.76	0.46
MEXICO	5.58	0.57
ARGENTINA	3.43	0.52
PORTUGAL	3.16	0.41
SINGAPORE	5.97	0.56
NORWAY	3.50	0.60
TURKEY	3.33	0.50
IRELAND	2.86	0.47
SOUTH AFRICA	4.71	0.57
REST OF WORLD	4.96	0.64
Average	17.01	0.54
Standard Deviation	28.92	0.08
Coefficient of Variation	1.70	0.15

ENGINEERING	**SK1**	**SK2**
USA	7.74	0.57
GERMANY	8.04	0.56
JAPAN	9.61	0.71
FRANCE	11.59	0.58
UK	10.54	0.46
RUSSIA	17.12	1.00
ITALY	4.36	0.54
CHINA	9.99	0.75
SWITZERLAND	4.69	0.52
SPAIN	4.42	0.57
CANADA	8.82	0.48
SOUTH KOREA	14.61	0.74
NETHERLANDS	4.24	0.60
POLAND	18.26	0.40
INDIA	23.95	0.64
SWEDEN	5.56	0.64
ZEU	5.36	0.65
ISRAEL	6.16	0.55
BRAZIL	7.36	0.43
AUSTRALIA	13.77	0.50
BELGIUM	4.40	0.66
DENMARK	3.81	0.49
TAIWAN	16.54	0.76
AUSTRIA	4.12	0.55
FINLAND	6.89	0.59
GREECE	6.57	0.37

UKRAINE	30.18	1.00
HUNGARY	6.49	0.52
CZECH REPUBLIC	4.96	0.80
MEXICO	8.69	0.67
ARGENTINA	3.48	0.47
PORTUGAL	5.96	0.46
SINGAPORE	3.75	0.47
NORWAY	14.11	0.57
TURKEY	4.50	0.48
IRELAND	4.17	0.79
SOUTH AFRICA	5.67	0.68
REST OF WORLD	4.74	0.63
Average	8.82	0.60
Standard Deviation	6.01	0.14
Coefficient of Variation	0.68	0.24

MATERIALS SCIENCE	**SK1**	**SK2**
USA	9.08	0.64
GERMANY	6.95	0.66
JAPAN	12.62	0.67
FRANCE	10.40	0.51
UK	9.74	0.51
RUSSIA	11.74	1.00
ITALY	14.01	0.46
CHINA	7.68	0.81
SWITZERLAND	4.97	0.66
SPAIN	7.00	0.47
CANADA	4.59	0.69
SOUTH KOREA	6.44	0.71
NETHERLANDS	16.77	0.66
POLAND	2.93	0.74
INDIA	12.57	0.49
SWEDEN	18.68	0.50
ZEU	10.11	0.71
ISRAEL	4.22	0.66
BRAZIL	3.57	0.54
AUSTRALIA	3.48	0.67
BELGIUM	18.55	0.53
DENMARK	5.89	0.59
TAIWAN	10.76	0.44
AUSTRIA	17.85	0.71
FINLAND	3.94	0.63
GREECE	5.04	0.64
UKRAINE	8.20	1.00
HUNGARY	4.92	0.53
CZECH REPUBLIC	5.23	0.63
MEXICO	6.84	0.51
ARGENTINA	2.74	0.55
PORTUGAL	3.20	0.47

SINGAPORE	6.66	0.56
NORWAY	3.27	0.46
TURKEY	2.76	0.54
IRELAND	4.29	0.56
SOUTH AFRICA	3.86	0.49
REST OF WORLD	5.49	0.71
Average	7.82	0.61
Standard Deviation	4.67	0.13
Coefficient of Variation	0.60	0.21

GEOSCIENCES	**SK1**	**SK2**
USA	15.02	0.47
GERMANY	30.34	0.44
JAPAN	6.39	0.52
FRANCE	33.10	0.51
UK	27.26	0.55
RUSSIA	10.81	0.61
ITALY	6.10	0.52
CHINA	7.07	0.75
SWITZERLAND	5.19	0.54
SPAIN	2.97	0.48
CANADA	4.13	0.55
SOUTH KOREA	8.55	0.47
NETHERLANDS	34.75	0.44
POLAND	3.57	0.67
INDIA	10.82	0.78
SWEDEN	5.00	0.54
ZEU	8.89	0.55
ISRAEL	5.82	0.58
BRAZIL	7.61	0.63
AUSTRALIA	4.36	0.52
BELGIUM	2.93	0.46
DENMARK	3.24	0.42
TAIWAN	2.68	0.63
AUSTRIA	8.88	0.59
FINLAND	2.94	0.58
GREECE	9.82	0.66
UKRAINE	2.94	0.59
HUNGARY	3.36	0.46
CZECH REPUBLIC	3.46	0.50
MEXICO	7.12	0.52
ARGENTINA	3.14	0.39
PORTUGAL	4.40	0.48
SINGAPORE	2.81	0.52
NORWAY	2.77	0.47
TURKEY	8.57	0.73
IRELAND	2.73	0.50
SOUTH AFRICA	2.95	0.55
REST OF WORLD	7.75	0.51

Average	8.43	0.54
Standard Deviation	8.54	0.09
Coefficient of Variation	1.01	0.16

ALL-SCIENCES CASE	**SK1**	**SK2**
USA	3.89	0.42
GERMANY	3.09	0.44
JAPAN	3.60	0.44
FRANCE	3.65	0.35
UK	3.45	0.45
RUSSIA	2.70	0.51
ITALY	3.16	0.44
CHINA	2.57	0.44
SWITZERLAND	1.91	0.38
SPAIN	2.79	0.39
CANADA	2.82	0.39
SOUTH KOREA	2.47	0.45
NETHERLANDS	3.10	0.41
POLAND	1.93	0.40
INDIA	2.49	0.40
SWEDEN	3.26	0.37
ZEU	2.65	0.38
ISRAEL	2.20	0.45
BRAZIL	3.16	0.39
AUSTRALIA	3.08	0.43
BELGIUM	2.63	0.42
DENMARK	3.15	0.38
TAIWAN	1.89	0.36
AUSTRIA	1.75	0.41
FINLAND	2.23	0.35
GREECE	1.67	0.37
UKRAINE	1.57	0.42
HUNGARY	1.99	0.28
CZECH REPUBLIC	1.36	0.36
MEXICO	1.90	0.39
ARGENTINA	2.09	0.35
PORTUGAL	2.22	0.34
SINGAPORE	0	0
NORWAY	2.19	0.16
TURKEY	1.59	0.42
IRELAND	1.82	0.43
SOUTH AFRICA	1.73	0.32
REST OF WORLD	3.06	0.45
Average	2.44	0.38
Standard Deviation	0.78	0.09
Coefficient of Variation	0.32	0.23

APPENDIX II

THE FRACTIONAL STRATEGY

In the fractional strategy, country p's citation distribution can be described by $\boldsymbol{c_p} = \{w_{pi}\, c_{pi}\}$, where $w_{pi} = (1/x_l)$ for all $p \in X_l$ and some article l in the initial distribution for which $c_{pi} = c_l$. Therefore, $\Sigma_{p \in Xl}\, w_{pi} = 1$. The fractional number of articles in country p is $n_p = \Sigma_i\, w_{pi}$, the citations received by each fractional article are $w_{pi}\, c_{pi}$, and the fractional number of citations in country p is $\Sigma_p\, w_{pi}\, c_{pi}$. It should be noted that $\Sigma_p\, n_p = \Sigma_p\, \Sigma_i\, w_{pi} = \Sigma_l \Sigma_{p \in Xl}\, w_{pi} = N$, and $\Sigma_p\, \Sigma_i\, w_{pi}\, c_{pi} = \Sigma_l\, c_l = \gamma$; that is, in the fractional strategy the total number of articles and citations in the original dataset, and hence the mean citation, are preserved at the country level.

For any p, let us partition the citation distribution $\boldsymbol{c_p}$ into Π quantiles of size n_p/Π. That is, let $\boldsymbol{c_p} = (\boldsymbol{c_p}^{\boldsymbol{1}}, \ldots, \boldsymbol{c_p}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c_p}^{\boldsymbol{\Pi}})$, where $\boldsymbol{c_p}^{\boldsymbol{\pi}} = \{w_{pj}^{\boldsymbol{\pi}}\, c_{pj}^{\boldsymbol{\pi}}\}$ is the vector of the citations received by the n_p/Π articles in the π-th quantile of distribution $\boldsymbol{c_p}$, with $j = 1, \ldots, n_p/\Pi$ and $c_{pj} = c_k$ for some article k in distribution $\boldsymbol{Q}$. Assume that we disregard the citation inequality within every vector $\boldsymbol{c_p}^{\boldsymbol{\pi}}$ by assigning to every article in that vector the (fractional) mean citation of the vector itself, $m_p^{\boldsymbol{\pi}}$, defined by $m_p^{\boldsymbol{\pi}} = (\Sigma_j\, w_{pj}^{\boldsymbol{\pi}}\, c_{pj}^{\boldsymbol{\pi}})/\Sigma_i\, w_{pj}^{\boldsymbol{\pi}} = (\Sigma_j\, w_{pj}^{\boldsymbol{\pi}}\, c_{pj}^{\boldsymbol{\pi}})/(n_p/\Pi)$.

For any π, define the vector $\boldsymbol{c}^{\boldsymbol{\pi}} = (\boldsymbol{c_1}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c_p}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c_P}^{\boldsymbol{\pi}})$. Consider the citation distribution $\boldsymbol{c} = \cup_p\, \boldsymbol{c_p}$, whose total number of articles is $\Sigma_p\, n_p = N$. Applying the decomposability property of the citation inequality index I_1 first to the partition $\boldsymbol{c} = (\boldsymbol{c}^{\boldsymbol{1}}, \ldots, \boldsymbol{c}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c}^{\boldsymbol{\Pi}})$, and then to the partition $\boldsymbol{c}^{\boldsymbol{\pi}} = (\boldsymbol{c_1}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c_p}^{\boldsymbol{\pi}}, \ldots, \boldsymbol{c_P}^{\boldsymbol{\pi}})$, we obtain a decomposition of the overall citation inequality in three terms equivalent to expression (4), $I_1(\boldsymbol{c}) = W + S + IDCC$, where the $IDCC$ term captures the citation inequality according to I_1 attributable to differences in citation impact across countries in the fractional case.