

“THE COMPARISON OF CITATION IMPACT ACROSS SCIENTIFIC FIELDS”

Juan A. Crespo^a, Yunrong Li^b, and Javier Ruiz-Castillo^b

^a Departamento de Economía Cuantitativa, Universidad Autónoma de Madrid

^b Departamento de Economía, Universidad Carlos III

Abstract

This paper has two aims: (i) to introduce a novel method for measuring which part of overall citation inequality can be attributed to differences in citation practices across scientific fields, and (ii) to implement an empirical strategy for making meaningful comparisons between the number of citations received by articles in the 22 broad fields distinguished by Thomson Scientific. The paper is based on a model in which the number of citations received by any article is a function of the article’s scientific influence, and the field to which it belongs. The model includes a key assumption according to which articles in the same quantile of any field citation distribution have the same *degree* of citation impact in their respective field. Using a dataset of 4.4 million articles published in 1998-2003 with a five-year citation window, we find that differences in citation practices between the 22 fields account for about 14% of overall citation inequality. Our empirical strategy for making comparisons of citation counts across fields is based on the strong similarities found in the behavior of citation distributions over a large quantile interval. We obtain three main results. Firstly, we provide a set of *exchange rates* to express citations in any field into citations in the all-fields case. (This can be done for articles in the interval between, approximately, the 71st and the 99th percentiles of their citation distributions). The answer is very satisfactory for 20 out of 22 fields. Secondly, when the raw citation data is normalized with our exchange rates, the effect of differences in citation practices is reduced to, approximately, 2% of overall citation inequality in the normalized citation distributions. Thirdly, we provide an empirical explanation of why the usual normalization procedure based on the fields’ mean citation rates is found to be equally successful.

Acknowledgements

The authors acknowledge financial support by Santander Universities Global Division of *Banco Santander*. Crespo and Ruiz-Castillo also acknowledge financial help from the Spanish MEC through grants SEJ2007-67436 and ECO2010-19596. Conversations with Jesús Carro, Eduardo Ley, Ignacio Ortuño, Juan Romo, Carlos Velasco, Ludo Waltman, and, above all, Pedro Albarrán, are deeply appreciated. All shortcomings are the authors’ sole responsibility.

I. INTRODUCTION

The field dependence of reference and citation counts in scientific articles has been recognized since the beginning of Scientometrics as a field of study (see *inter alia* Pinski and Narin, 1976, Murugesan and Moravcsik, 1978, and Garfield, 1979). There are multiple reasons. Consider the differences across scientific disciplines in, for example, (i) size, measured by the number of publications in the periodical literature; (ii) the average number of authors per paper; (iii) the average paper length; (iv) the average number of papers per author in a given period of time; (v) the theoretical or experimental mix that characterizes each discipline; (vi) the average number of references per paper; (vii) the proportion of references that are made to other articles in the periodical literature; (viii) the percentage of internationally co-authored papers, or (ix) the speed at which the citation process evolves. For example, in the dataset used in this paper covering 22 broad fields, the mean citation in Mathematics is 2.4, about eight and a half times smaller than in Molecular Biology and Genetics where it is equal to 20.4 citations.

It would appear that differences in publication and citation practices pose insurmountable obstacles to direct comparisons of the absolute number of citations received by articles in different fields. However, this paper claims that some bold but reasonable assumptions, coupled with the striking similarity between citation distributions at different aggregation levels (documented in Albarrán and Ruiz-Castillo, 2011, Albarrán *et al.*, 2011, and Radicchi and Castellano, 2012a) opens up the possibility of meaningful comparison of citation counts across scientific fields.

We use a model in which the number of citations received by an article is a function of two variables: the article's underlying scientific influence, and the field to which it belongs. For our purposes, we may remain agnostic about the myriad of motives researchers have in their citation behavior as long as we are allowed to assume that citation impact varies monotonically with scientific influence (for a survey of the controversies concerning the meaning of citation counts, see Bornmann and Daniel, 2008). Thus, if one

article has greater scientific influence than another one in the same homogeneous field, then we expect the former to have also a greater citation impact than the latter.¹

In this context, the citation inequality of the distribution consisting of all articles in all fields –the all-fields case– is the result of two forces: differences in scientific influence, and differences in citation practices across fields. Given the field, the monotonicity assumption imposed on the function that links citation impact with scientific influence ensures that the quantiles of the (unobservable) scientific influence distribution for that field coincide with the quantiles of the corresponding (observable) citation distribution. More importantly, the model includes a second, strong assumption according to which articles in the same quantile have the same *degree* of scientific influence independently of the field to which they belong. Therefore, citation inequality at any given quantile can be solely attributed to differences in citation practices across fields. The aggregation of this measure over all quantiles provides a distribution-free method of quantifying the effect of these differences (This is, essentially, John Roemer’s, 1998, model for the study of inequality of opportunities in an economic or sociological context). We implement this model by using an additively decomposable inequality index, in which case the citation inequality attributed to differences in citation practices can be measured as a between-group inequality term in the double partition by field and citation quantile (This idea is taken from Ruiz-Castillo, 2003). The usefulness of our approach lies in the fact that any normalization procedure can be evaluated by the reduction it induces in the between-group term just described.

Using a dataset of 4.4 million articles published in 1998-2003 with a five-year citation window, we find that the citation inequality attributable to differences in citation practices across the 22 fields distinguished by Thomson Scientific is approximately constant over a wide range of quantiles. This allows the effect of idiosyncratic citation practices to be rather well estimated over that interval. Consequently, we provide a set of *exchange rates* and their standard deviations (SDs hereafter) that serve to address the following two empirical

¹ The idea that citations is an observable indicator for a latent concept of scientific or scholarly influence, as well as the monotonicity assumption, are also found in Ravallion and Wagstaff (2011) in a different scenario: the construction of bibliometric measures of research impact.

questions. Firstly, how many citations in a given field are equivalent to, say, 10 citations in the all-fields case? For example, in *Clinical Medicine* the answer is 12.1 with a SD of 0.6, while in *Mathematics* the answer is 3.3 with a SD of 0.2. Secondly, how much can we reduce the effect of different citation practices by normalizing the raw citation data with the exchange rates? We find that this normalization procedure reduces this effect from, approximately, 14% to 2% of overall citation inequality.

It should be emphasized that the difficulty of comparing citation counts across scientific fields is a very well known issue that has worried practitioners of Scientometrics since its inception. Differences in citation practices are usually taken into account by choosing the world mean citation rates as normalization factors (see *inter alia* Moed *et al.*, 1985, 1988, 1995, Braun *et al.*, 1985, Schubert *et al.*, 1983, 1987, 1988, Schubert and Braun, 1986, 1996, and Vinkler 1986, 2003). More recently, other contributions support this traditional procedure on different grounds (Radicchi *et al.*, 2008, Radicchi and Castellano, 2012a, 2012b). In our last contribution to the literature, we find that using field mean citations as normalization factors leads practically to the same large reduction of the effect of differences in citation practices as our exchange rates. We show how our model helps explaining why the traditional model is so successful.²

The rest of the paper consists of three Sections. Section II introduces the model for the measurement of the effect of differences in citation practices. Section III presents the estimation of average-based exchange rates and its SDs over a large quantile interval, and discusses the consequences of using field exchange rates and mean citations as normalization factors. Section IV contains some concluding comments.

II. THE MODEL

II. 1. Notation

From an operational point of view, a scientific field is a collection of papers published in a set of closely related professional journals. A field is said to be *homogeneous* if the number of citations received by its papers

² Methods that use mean citations or exchange rates as normalization factors belong to the class of target or “cited side” normalization procedures. Following an idea in Small and Sweeney (1985), source or “citing side” procedures have been recently suggested (see *inter alia* Zitt and Small, 2008, Moed, 2010, and Leydesdorff and Opthof, 2010). Since our dataset lacks citing side information, applying this type of procedure is beyond the scope of this paper.

is comparable independently of the journal where each has been published. Let N_f be the total number of articles in a homogeneous field f , and let $\mathbf{c}_f = (c_{f1}, \dots, c_{fN_f})$ be the citation distribution for that field where, for each $i = 1, \dots, N_f$, c_{fi} is the number of citations received by the i -th article. Assume that there are F homogeneous fields, indexed by $f = 1, \dots, F$. The total number of articles in the all-fields case is $N = \sum_f N_f$.

The number of citations of any article, c_{fi} is assumed to be a function of two variables: the field f to which the article belongs, and the scientific influence of the article in question, q_{fi} , which is assumed for simplicity to be a single-dimensional variable. Thus, for every f we write:

$$c_{fi} = \phi(f, q_{fi}), \quad i = 1, \dots, N_f \quad (1)$$

Let $\mathbf{q}_f = (q_{f1}, q_{f2}, \dots, q_{fN_f})$ with $q_{f1} \leq q_{f2} \leq \dots \leq q_{fN_f}$ be the ordered distribution of scientific influence in every field. The distribution \mathbf{q}_f is assumed to be a characteristic of the field. Consequently, for any two articles i and j in two different fields f and g , the values q_{fi} and q_{gj} cannot be compared. No restriction is *a priori* imposed on distributions $\mathbf{q}_f, f = 1, \dots, F$. The form of ϕ is unknown, but we adopt the following assumption about it:

Assumption 1 (A1). *The function ϕ in expression (1) is assumed to be monotonic in scientific influence, that is, for every pair of articles i and j in field f , if $q_{fi} \leq q_{fj}$ then $c_{fi} \leq c_{fj}$.*

For each f , consider the partition of \mathbf{q}_f into Π quantiles of size N_f/Π , indexed by $\mathbf{q}_f^\pi, \pi = 1, \dots, \Pi$, so that $\mathbf{q}_f = (\mathbf{q}_f^1, \dots, \mathbf{q}_f^\pi, \dots, \mathbf{q}_f^\Pi)$. Typically, scientific influence is an unobservable variable. However, given A1, the degree of scientific influence uniquely determines the location of an article in its field citation distribution. In other words, for every f , the partition of the scientific influence distribution \mathbf{q}_f into Π quantiles induces a corresponding partition of the citation distribution $\mathbf{c}_f = (c_f^1, \dots, c_f^\pi, \dots, c_f^\Pi)$ into Π quantiles, where c_f^π is the

vector of the citations received by the N_f/Π articles in the π -th quantile of field f . For later reference, let μ_f be the mean citation of field f .

Consider the overall citation distribution in the all-fields case, $\mathbf{C} = (c_1, \dots, c_b, \dots, c_N)$, where, for each l , there exists some article i in some field f such that $c_i = c_{\hat{f}}$. Note that distribution \mathbf{C} can be organized as an $(F \times \Pi)$ array consisting of F rows, $\mathbf{c}_f = (c_f^1, \dots, c_f^\pi, \dots, c_f^\Pi)$, $f = 1, \dots, F$, or Π columns, $\mathbf{c}^\pi = (c_1^\pi, \dots, c_f^\pi, \dots, c_F^\pi)$, $\pi = 1, \dots, \Pi$. Assume for a moment that we disregard the citation inequality within every vector \mathbf{c}_f^π by assigning to every article in that vector the mean citation of the vector itself, namely, μ_f^π . Consider the corresponding $(F \times \Pi)$ array, denoted by \mathbf{M} :

$$\mathbf{M} = \begin{bmatrix} \mu_1^1 & \dots & \mu_1^\pi & \dots & \mu_1^\Pi \\ \mu_f^1 & \dots & \mu_f^\pi & \dots & \mu_f^\Pi \\ \mu_F^1 & \dots & \mu_F^\pi & \dots & \mu_F^\Pi \end{bmatrix} \quad (2)$$

Of course, the citation inequality of each row $\mathbf{m}_f = (\mu_f^1, \dots, \mu_f^\Pi)$ of \mathbf{M} is due to differences in the underlying scientific influence distribution \mathbf{q}_f . The citation inequality of each column $\mathbf{m}^\pi = (\mu_1^\pi, \dots, \mu_F^\pi)$ of \mathbf{M} require some explanation that will be provided in the next Sub-section with the help of Assumption 2.

II. 2. Comparability Conditions

As we know, the number of citations $c_{\hat{f}_i}$ and $c_{\hat{f}_j}$ for any two articles i and j in two fields f and g cannot be directly compared. We do not even know how to rank the scientific influences of the two articles in question. Similarly, mean citations μ_f and μ_g are not comparable either. To overcome this difficulty, in this paper we introduce some structure into the comparability problem by means of the following key assumption.

Assumption 2 (A2). *Articles at the same quantile π of any field citation distribution have the same degree of citation impact in their respective field.*³

Let μ_f^π and μ_g^π be the mean citations of articles in quantile π in two fields f and g . Under A2, the interpretation of the fact that, for example, $\mu_f^\pi = 2 \mu_g^\pi$ is that, on average, field f uses twice the number of citations as field g to represent the same underlying phenomenon, namely, the same degree of scientific influence in both fields. Consequently, for any π , the difference between μ_f^π and μ_g^π for articles with the same degree of scientific influence is entirely attributable to differences in citation practices between the two fields.⁴

II.3. The Measurement of the Effect of Differences in Citation Practices

For any population partition, we are interested in expressing the overall citation inequality as the sum of two terms: a weighted sum of *within-group* inequalities, plus a *between-group* inequality component. An inequality index is said to be *decomposable by population subgroup*, if the decomposition procedure of overall inequality into a within-group and a between-group term is valid for any arbitrary population partition. In the relative, or scale-invariant inequality case it is customary to calculate the between-group component by applying the inequality index to a citation vector in which each article in a given subgroup is assigned the

³ Since the quantiles of citation impact correspond –as we have already seen– to quantiles of the underlying scientific influence distribution, holding constant the degree of citation impact at any level is equivalent to holding constant the degree of scientific influence at that level.

⁴ Welfare economists would surely recognize the above as Roemer’s (1998) model for the inequality of opportunities where individual incomes (or other indicators of performance, such as educational outcomes) are assumed to be a function of two types of factors: a set of variables outside an individual’s responsibility –the *circumstances*, mainly inherited from our parents–, and *effort*, an unobservable single dimensional variable entirely within the sphere of each individual’s responsibility. Circumstances allow a partition of the population into *types*. In this model, income inequality holding constant the degree of effort by every type is seen to be entirely due to differences in circumstances, or to the *inequality of opportunities* at this degree of effort. Income inequality due to differences in effort is not worrisome from a social point of view. It is income inequality due to differences in circumstances, namely, the inequality of opportunities, what society might attempt to compensate for. Individuals are articles, the equivalent of income is citations, types are fields, and effort is scientific influence.

subgroup's citation mean. Under this convention, it is well known that the Generalized Entropy (GE hereafter) family of inequality indices are the only measures of relative inequality that satisfy the usual properties required from any inequality index⁵ and, in addition, are decomposable by population subgroup (Bourguignon, 1978, and Shorrocks, 1980, 1984).

Without loss of generality, it is useful to develop the following measurement framework in terms of only one member of this family, the first Theil index, denoted by I_T , and defined as:

$$I_T(\mathbf{C}) = (1/N) \sum_l (c_l/\mu) \log (c_l/\mu), \quad (3)$$

where μ is the mean of distribution \mathbf{C} . In the Working Paper version of this article, Crespo *et al.* (2012), it is shown that, in this case, overall citation inequality in the double partition of distribution \mathbf{C} into Π quantiles and F fields is seen to be:

$$I_T(\mathbf{C}) = W + S + IDCP, \quad (4)$$

where:

$$W = \sum_{\pi} \sum_f v^{\pi,f} I_T(\mathbf{c}_f^{\pi})$$

$$S = I_T(\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^{\Pi})$$

$$IDCP = \sum_{\pi} v^{\pi} I_T(\boldsymbol{\mu}_1^{\pi}, \dots, \boldsymbol{\mu}_F^{\pi}) = \sum_{\pi} v^{\pi} I_T(\pi),$$

where $v^{\pi,f}$ is the share of total citations in quantile π of field f , and $v^{\pi} = \sum_f v^{\pi,f}$ is the share of total citations in quantile π . The term W in equation (4) is a within-group term, which captures the weighted citation inequality within each quantile in every field. Obviously, since all articles in each vector \mathbf{c}_f^{π} belong to the same field, there is no difficulty in computing the expression $I_T(\mathbf{c}_f^{\pi})$. The term S is the citation inequality of the distribution $\mathbf{m} = (\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^{\Pi})$ in which each article in a given quantile π is assigned the quantile's citation

⁵ Namely, continuity; scale invariance; invariance to population replications, or size-invariance, and S-convexity that ensures that transfers from an article with more citations to another with fewer citations without altering their ranking reduces citation inequality.

mean, $\mu^\pi = \sum_f [(N_f/N)\mu_f^\pi]$. Thus, S is a measure of citation inequality at different degrees of citation impact that captures well the skewness of science in the all-fields case. Finally, for any π , under A2 the citation inequality of distribution m^π according to any relative inequality index $I, I_1(\mu_1^\pi, \dots, \mu_F^\pi)$, abbreviated $I(\pi)$, is entirely attributable to differences in citation practices across the F fields at quantile π . Thus, the weighted average of these quantities that constitute the third term in expression (4), denoted by *IDCP* (*Inequality due to Differences in Citation Practices*), provides a good measure of the citation inequality due to such differences.

The following three points should be emphasized. Firstly, for large $\pi, I_1(\mu_f^\pi)$, and hence term W is expected to be small. Secondly, due to the high skewness of all citation distributions (see *inter alia* Albarrán and Ruiz-Castillo, 2011, and Albarrán *et al.*, 2011), the term S is expected to be large. Thirdly, overall citation inequality due to differences in scientific influence –captured by the W and S terms– is not worrisome. Instead, we would like to eliminate as much as possible the citation inequality attributable to differences in citation practices. Thus, the impact of any normalization procedure can be evaluated by the reduction in the term *IDCP* before and after normalization.

III. EMPIRICAL RESULTS

In this paper only research articles or, simply, articles, are studied. Our dataset consists of 4.4 million articles published in 1998-2003, and the 35 million citations they receive after a common five-year citation window for every year. Since we must work with partitions of the N articles, we identify the set of homogeneous fields with the 20 broad fields for the natural sciences and two fields for the social sciences distinguished by Thomson Scientific (Table A in the Appendix in Crespo *et al.*, 2012, presents the number of articles and mean citation rates by field).

This Section analyzes two empirical problems: (i) how to compare the citations received by two articles in any pair of the 22 fields in our dataset by using exchange rates that are approximately constant over a large quantile interval, and (ii) how much the effect of differences in citation practices is reduced when these exchange rates, or the field mean citations are used as normalization factors.

III. 1. The Comparison of Citation Counts Across Different Fields

Mean citations of comparable articles belonging to the same quantile can be used to express the citations in any field in terms of the citations in a reference situation. For example, if we let μ^π be the mean citation of all articles in quantile π , then the *exchange rates at quantile π* , $e_f(\pi)$, defined by

$$e_f(\pi) = \mu_f^\pi / \mu^\pi, \tag{5}$$

can be seen to answer the following question: how many citations for an article at the degree π of scientific influence in field f are equivalent on average to one citation in the all-fields case? In the metaphor according to which a field's citation distribution is like an income distribution in a certain currency, the exchange rates $e_f(\pi)$ permit to express all citations in the same reference currency for that π : since c_{fi} is the number of citations received by article i in quantile π of field f , the ratio $c_{fi}^*(\pi) = c_{fi}/e_f(\pi)$ is the equivalent number of citations in the reference currency at that quantile. Naturally, if for many fields $e_f(\pi)$ were to drastically vary with π , then we might not be able to claim that differences in citation practices have a common element that can be precisely estimated. However, we next establish that exchange rates are sufficiently constant over a wide range of quantiles.

It is very instructive to have a graphical representation of how the effect of differences in citation practices, measured by $I(\pi)$, changes with π when $\Pi = 1,000$ (since $I(\pi)$ is very high for $\pi < 600$, for clarity

these quantiles are omitted from Figure 1).⁶ It is observed that $I(\pi)$ is particularly high until $\pi \approx 700$, as well as for a few quantiles at the very upper tail of citation distributions. However, $I(\pi)$ is strikingly similar for a wide range of intermediate values.⁷ In this situation, it is reasonable to define an average-based *exchange rate* (ER hereafter) over some interval $[\pi_m, \pi^M]$ in that range as

$$e_f = [1/(\pi^M - \pi_m)] [\sum_{\pi} e_f(\pi)]. \quad (6)$$

The advantage of this definition is that we can easily compute the associated SD, denoted by σ_f . The fact that, for each f , the $e_f(\pi)$ defined in (5) are very similar for all π in the interval $[\pi_m, \pi^M]$ would manifest itself in a small σ_f and hence in a small coefficient of variation $CV_f = \sigma_f/e_f$.

Figure 1 around here

We find that the choice $[\pi_m, \pi^M] = [706, 998]$ —where $I(\pi)$ for most π is equal to $I(\pi_m) = 0.1081$ and $I(\pi^M) = 0.1084$ —is a good one. The ERs e_f as well as the σ_f and CV_f are in columns 1 to 3 in Table 1. For convenience, ERs are multiplied by 10. Thus, for example, the first row indicates that 15.8 citations with a standard deviation of 0.9 for an article in Biology and Biochemistry between, approximately, the 71st and the 99th percentile of its citation distribution, are equivalent to 10 citations for an article in that interval in the all-sciences case. We find it useful to divide fields into three groups according to the CV_f . Group I (colored in green in Table 1), consisting of 10 fields, has a CV_f smaller than or equal to 0.05. This means that the SD of the exchange rate, σ_f is less than or equal to five percent of the exchange rate itself. Hence, we consider ERs in this group as highly reliable. Group II (black), consisting of 10 fields, has a CV_f between 0.05 and 0.10. We

⁶ Crespo *et al.* (2012) discusses the choice of the inequality index I_f among the members of the GE family. To solve the problem that I_f is not defined for articles without citations (see expression 3), we have followed the convention $0 \log(0) = 0$. However, we have also experimented with the assignment of values 0.1 and 0.01 to these articles. Crespo *et al.* (2012) establishes the robustness of our results to this choice, as well as to the alternatives $\Pi = 10, 50, 100, 1000$.

⁷ It is important to emphasize that this is consistent with the stylized facts characterizing citation distributions documented in Albarrán and Ruiz-Castillo (2011), and Albarrán *et al.* (2012): although citation distributions are rather different in a long lower tail and at the very top of the upper tail, they behave very similarly over a partition into three broad classes.

consider *ERs* in this group as fairly reliable. Group III (red), consists of two fields: *Computer Science*, with a CV_f greater than 0.10, which is known from previous work to behave as an outlier (Herranz and Ruiz-Castillo, 2012), and the *Multidisciplinary* field with a CV_f greater than 0.15. The results for these two fields should be considered unreliable.

As is observed in column 4 in Table 1, on average the interval [706, 998] includes 72.1% of all citations (with a SD of 3.9). Although this is a large percentage, expanding the interval in either direction would bring a larger percentage of citations. It turns out that the *ERs* do not change much. However, they exhibit greater variability (for details, see Crespo *et al.*, 2012). Therefore, we find it useful to retain the interval [706, 998] in the sequel.

Table 1 around here

III. 2. Normalization Results

Figure 2 focuses on the product $v^\pi I(\pi)$ as a function of π . Of course, the term *IDCP* is equal to the integral of this expression (for clarity, quantiles $\pi < 600$, and $\pi > 994$, are omitted from Figure 2). As we saw in Section II.3, the impact of any normalization procedure can be evaluated by the reduction in this term before and after normalization. In the first place, we want to assess the normalization procedure based on *ERs* whereby the citations received by any article i in field f , c_{if} , are converted into normalized citations c_{if}^* as follows: $c_{if}^* = c_{if}/e_f$. Relative to the blue curve, the red curve illustrates the correction achieved by normalization: the size of the *IDCP* term is very much reduced. The numerical results before and after this normalization are in Panels A and B in Table 2.

Table 2 and Figure 2 around here

Note that, as expected, the term \mathcal{W} is small, while the term S is large. Both terms remain essentially constant after normalization. However, in absolute terms the *IDPC* term is reduced from 0.1221 to 0.0156, a 87.2% difference. Of course, total citation inequality after normalization is also reduced. On balance, the

IDPC term after normalization only represents 1.96% of total citation inequality –a dramatic reduction from the 13.9% with the raw data (for the robustness of this estimate to different choices of Π , as well as different members of the GE family, see Crespo *et al.*, 2012).

However, it should be recognized that in the last two quantiles and, above all, in the [1, 705] interval normalization results quickly deteriorate. It would appear that a convenient alternative consists of normalizing the lower tail of the original distributions by some appropriate *ERs* within the [1, 705] interval. The problem is that citation inequality due to different citation practices in that interval is both high and extremely variable for different quantiles. This explains why the *ERs* computed according to equation (6) for the entire [1, 705] interval lead to a worsening of the situation. Moreover, the improvement achieved with the restriction to the interval [356, 705] is, at most, very slight (see Crespo *et al.*, 2012).

As indicated in the Introduction, the difficulties of combining heterogeneous citation distributions into broader aggregates have been traditionally confronted using mean citations as normalization factors (see Crespo *et al.*, 2011, for a review of this literature). In our dataset, the *IDCP* term after the traditional normalization procedure only represents 1.92% of total citation inequality (see Panel C in Table 2). The two solutions are so near that we refrain to illustrate the latter in Figure 2 because it will be indistinguishable with the red curve after normalization by our *ERs*. This confirms the results in Radicchi and Castellano (2012a) – RC hereafter–, where it is concluded that, despite not being strictly correct, this procedure is a very good approximation of the two-parameter transformation able to make citation counts independent of the scientific field. All of which justifies the use of mean citations as normalization factors, as traditional practiced in Scientometrics since the mid 1980s, and as recently suggested by Radicchi *et al.* (2008) and Radicchi and Castellano (2012b).

The question is, how can this similarity of results be accounted for? The explanation is as follows. As we have seen in Figure 1, field citation distributions differ approximately by a set of scale factors only in the

[706, 998] interval. Therefore, these scale factors should be well captured by any average-based measure of what takes place in that interval –such as our *ERs*. However, as documented in Albarrán *et al.* (2011), field mean citations μ_f are reached, on average, at the 69.7 percentile with a SD of 2.6, that is, at the lower bound of our [706, 998] interval. Thus, the *ERs* based on mean citations, $e_f(\mu) = \mu_f/\mu$, computed with data from *all* quantiles of citation distributions, are approximately equal to the average of our own $e_f(\pi)$ for the [706, 998] interval (see column 5 in Table 1). In other words, let $e_f(\mu') = \mu'_f/\mu'$ be the *ERs* based on the mean citations restricted to the [706, 998] interval (see column 6 in Table 1). The scale factors separating field citation distributions over that interval should be well captured by such new *ERs*. However, as can be seen in Table 1, these *ERs* are essentially equal to the old ones, that is, for each f , $e_f(\mu') \approx e_f(\mu) \approx e_f$.

IV. CONCLUSIONS

The lessons that can be drawn from this paper can be summarized in the following four points.

1. We have provided a simple model for the measurement of the effect of differences in citation practices across scientific fields. Using a member of the GE family of inequality indices, this effect is well captured by a between-group term –denoted by *IDCP*– in the double partition by field and quantile of the overall citation distribution in the all-fields case. This method is distribution free, that is, it imposes no *a priori* restrictions whatsoever on citation distributions. The success of any normalization procedure in eliminating as much as possible the impact of differences in citation practices can be evaluated by the reduction it induces in the *IDCP* term.

2. Using a large dataset of 4.4 million articles, the striking similarity of citation distributions allows the effect of idiosyncratic citation practices to be rather well estimated over a wide range of quantiles where citation distributions essentially differ by a scale factor. Consequently, a set of exchange rates is estimated for two purposes: the translation of citation counts of articles in different fields within that interval into the

citations in a reference situation, and the normalization of the raw citation data. The advantage of this procedure is that, contrary to the field citation means, the exchange rates are estimated with a reasonably low standard deviation for 20 out of 22 fields.

It should be stressed that, for uncited and poorly cited articles below the mean, and for articles in the very upper tail of citation distributions, no clear answer to the comparability of citation counts for articles in different fields can be provided. Since the citation process evolves at different velocity in different fields, using variable citation windows to ensure that the process has reached a similar stage in all fields should improve field comparability at the lower tail of citation distributions. Naturally, we may also worry about how to compare citation counts in the last two quantiles of citation distributions. Given the fact that in this key segment the citation impact appears to be very diverse across fields, perhaps this task should not be even attempted. Until we know more concerning how differential citation practices operate in these top quantiles, the most we can do within this paper's framework is to use exchange rates $e_j(\pi)$ for $\pi = 999, 1000$.

3. It has been established that both the normalization procedure using our exchange factors, as well as the traditional method of taking the field citation means as normalization factors reduces the importance of the *IDCP* term relative to overall citation inequality from, approximately, 14% to 2%. The paper provides an empirical explanation of why the two methods are equally successful. Other normalization proposals –such as the one in RC, or those based on citing side procedures quoted in the Introduction, might be analogously evaluated. In turn, it would be interesting to evaluate the normalization procedure based on the exchange rates in terms of the reduction of the bias in the RC model. Given how near our *ERs* are from those based on the fields' mean citation rates, the conjecture is that our procedure would perform as well as the approximation provided by these means in the RC paper.

4. Policy makers and other interested parties should be very cautious when comparing citation performance in different scientific fields. More research is still needed. In particular, we need to study the robustness of our strategy to other datasets, as well as to extend it to lower aggregation levels. However,

together with the important contribution by RC, the results of this paper indicate that the combination of interesting assumptions with the empirical similarity of citation distributions paves the way for meaningful comparisons of citation counts across heterogeneous fields.

REFERENCES

- 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), “The Skewness of Science In 219 Sub-fields and a Number of Aggregates”, *Scientometrics*, **88**: 385-397.
- Bornmann, L, and H-D. Daniel (2008), “What Do Citation Counts Measure”, *Journal of Documentation*, **64**: 45-80.
- Bourguignon, F. (1979), “Decomposable income inequality measures”, *Econometrica*, **47**: 901-920.
- Braun, T., W. Glänzel, and A. Schubert (1985), *Scientometrics Indicators. A 32 Country Comparison of Publication Productivity and Citation Impact*, World Scientific Publishing Co. Pte. Ltd., Singapore, Philadelphia.
- Crespo, J. A., Li, Yunrong, and Ruiz-Castillo, J. (2012), “The Comparison of Citation Impact Across Scientific Fields”, Working Paper 12-06, Universidad Carlos III (<http://hdl.handle.net/10016/14771>).
- Herranz, N. & 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* (DOI:10.1002/asi.22629).
- Garfield, E. (1979), *Citation Indexing: Its Theory and Applications in Science, Technology, and Humanities*, New York: Wiley.
- Leydesdorff, L., and Opthof, T. (2010), “Normalization at the Field level: Fractional Counting of Citations”, *Journal of Informetrics*, **4**: 644-646.
- Moed H. F. (2010), “Measuring contextual citation impact of scientific journals”, *Journal of Informetrics*, **4**: 265–77.
- Moed, H. F., W. J. Burger, J. G. Frankfort, and A. F. J. van Raan (1985), “The Use of Bibliometric Data for the Measurement of University Research Performance”, *Research Policy*, **14**: 131-149.
- Moed, H. F., and A. F. J. van Raan (1988), “Indicators of Research Performance”, in A. F. J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland: 177-192.
- Moed, H. F., R. E. De Bruin, and Th. N. van Leeuwen (1995), “New Bibliometrics Tools for the Assessment of national Research Performance: Database Description, Overview of Indicators, and First Applications”, *Scientometrics*, **33**: 381-422.
- Murugesan, P., and Moravcsik, M.J. (1978), “Variation of the Nature of Citation Measures with Journal and Scientific Specialties”, *Journal of the American Society for Information Science and Technology*, **105**: 17268-17272.
- Pinski, G. and Narin, F. (1976), “Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Applications to the Literature of Physics”, *Information Processing and Management*, **12**: 297-312.
- Radicchi, F., Fortunato, S., and Castellano, C. (2008), “Universality of Citation Distributions: Toward An Objective Measure of Scientific Impact”, *PNAS*, **105**: 17268-17272.
- 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.
- Ravallion, M., and Wagstaff, A. (2011), "On measuring scholarly influence by citations", *Scientometrics*, **88**: 321-337.
- Roemer, J. (1998), *Equality of Opportunities*, Harvard University Press.
- Ruiz-Castillo, J. (2003), "The Measurement of Inequality of Opportunities", in J. Bishop y Y. Amiel (eds.), *Research in Economic Inequality*, **9**: 1-34.
- Shorrocks, A. F. (1980), "The Class of Additively Decomposable Inequality Measures," *Econometrica*, **48**: 613-625.
- Shorrocks, A. F. (1984), "Inequality Decomposition by Population Subgroups," *Econometrica*, **52**: 1369-1388.
- Schubert, A., and T. Braun (1986), "Relative Indicators and Relational Charts for Comparative Assessment of Publication Output and Citation Impact", *Scientometrics*, **9**: 281-291.
- Schubert, A., and T. Braun (1996), "Cross-field Normalization of Scientometric Indicators", *Scientometrics*, **36**: 311-324.
- Schubert, A., W. Glänzel, and T. Braun (1983), "Relative Citation Rate: A New Indicator for Measuring the Impact of Publications", in D. Tomov and L. Dimitrova (eds.), *Proceedings of the First National Conference with International Participation in Scientometrics and Linguistics of Scientific Text*, Varna.
- Schubert, A., W. Glänzel and T. Braun (1987), "A New Methodology for Ranking Scientific Institutions", *Scientometrics*, **12**: 267-292.
- Schubert, A., W. Glänzel, and T. Braun (1988), "Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools", in A. F. J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*: 137-176.
- Small, and Sweeney (1985), "Clustering the Science Citation Index Using Co-citations. I. A Comparison of Methods", *Scientometrics*, **7**: 393-404.
- Vinkler, P. (1986), "Evaluation of Some Methods For the Relative Assessment of Scientific Publications", *Scientometrics*, **10**: 157-177.
- Vinkler, P. (2003), "Relations of Relative Scientometric Indicators", *Scientometrics*, **58**: 687-694.
- Zitt M, and Bassecoulard, E. (1994), "Development of a Method for Detection and Trend Analysis of Research Fronts Built by Lexical or Cocitation Analysis", *Scientometrics*, **30**: 333-351.
- Zitt M, and Small H. (2008), "Modifying the journal impact factor by fractional citation weighting: The audience factor", *Journal of the American Society for Information Science and Technology*, **59**: 1856-1860.

$I(\pi)$

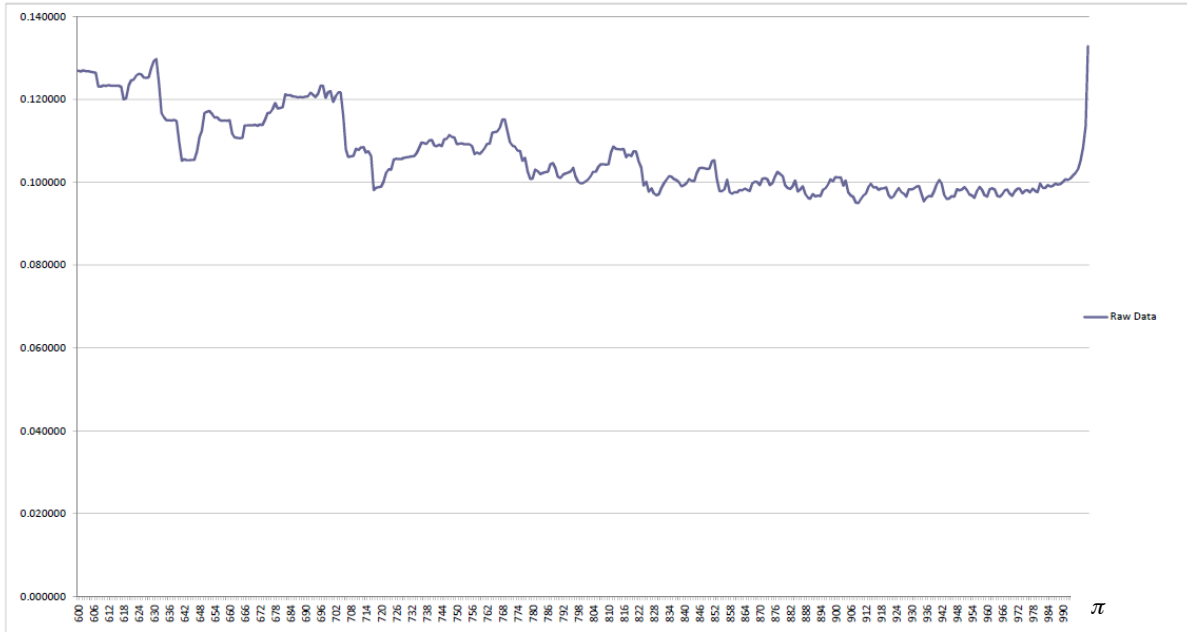


Figure 1. Citation Inequality Due to Differences in Citation Practices, $I(\pi)$ versus π . Raw Data

Table 1. Exchange Rates, Standard Deviations, and Coefficient of variation for the [706, 998] Interval, and Exchange Rates Based on Mean Citations

	Exchange Rates	Standard Deviation	Coefficient of Variation	% of Citations	ERs Based on Mean Citations	ERs Based on Mean Cits. In the [706, 998] Interval
1. Biology & Biochemistry	15.8	0.9	0.054	68.0	16.0	15.3
2. Clinical Medicine	12.1	0.6	0.049	71.8	12.4	12.5
3. Immunology	19.5	0.9	0.048	66.3	20.4	19.0
4. Microbiology	14.4	1.3	0.092	65.8	14.6	13.5
5. Molecular Biology & Genetics	25.7	0.6	0.022	71.1	25.9	25.9
6. Neuroscience & Behav. Science	17.1	0.8	0.050	67.2	17.5	16.5
7. Pharmacology & Toxicology	10.1	0.6	0.056	68.4	10.2	9.8
8. Psychiatry & Psychology	9.1	0.2	0.025	72.4	9.0	9.1
9. Chemistry	9.9	0.4	0.037	70.9	9.7	9.7
10. Computer Science	3.7	0.5	0.124	76.3	3.8	4.0
11. Mathematics	3.3	0.2	0.059	75.4	3.1	3.3
12. Physics	8.8	0.5	0.061	74.2	8.7	9.1
13. Space Science	14.2	0.3	0.019	71.9	14.0	14.2
14. Agricultural Sciences	6.5	0.4	0.056	72.5	6.2	6.3
15. Engineering	4.4	0.2	0.054	75.9	4.1	4.4
16. Environment & Ecology	9.1	0.7	0.073	68.3	9.1	8.7
17. Geoscience	8.9	0.6	0.069	70.1	8.6	8.5
18. Materials Science	5.9	0.3	0.048	75.0	5.8	6.1
19. Multidisciplinary	4.3	0.7	0.158	81.6	4.1	4.7
20. Plant & Animal Science	6.7	0.3	0.045	71.3	6.5	6.5
21. Economics & Business	5.2	0.4	0.068	75.6	5.0	5.3
22. Social Sciences, General	4.5	0.2	0.045	75.1	4.2	4.5
Mean				72.1		
	(1)	(2)	(3)	(4)	(5)	(6)

$$\frac{\pi}{v} I(\pi)$$

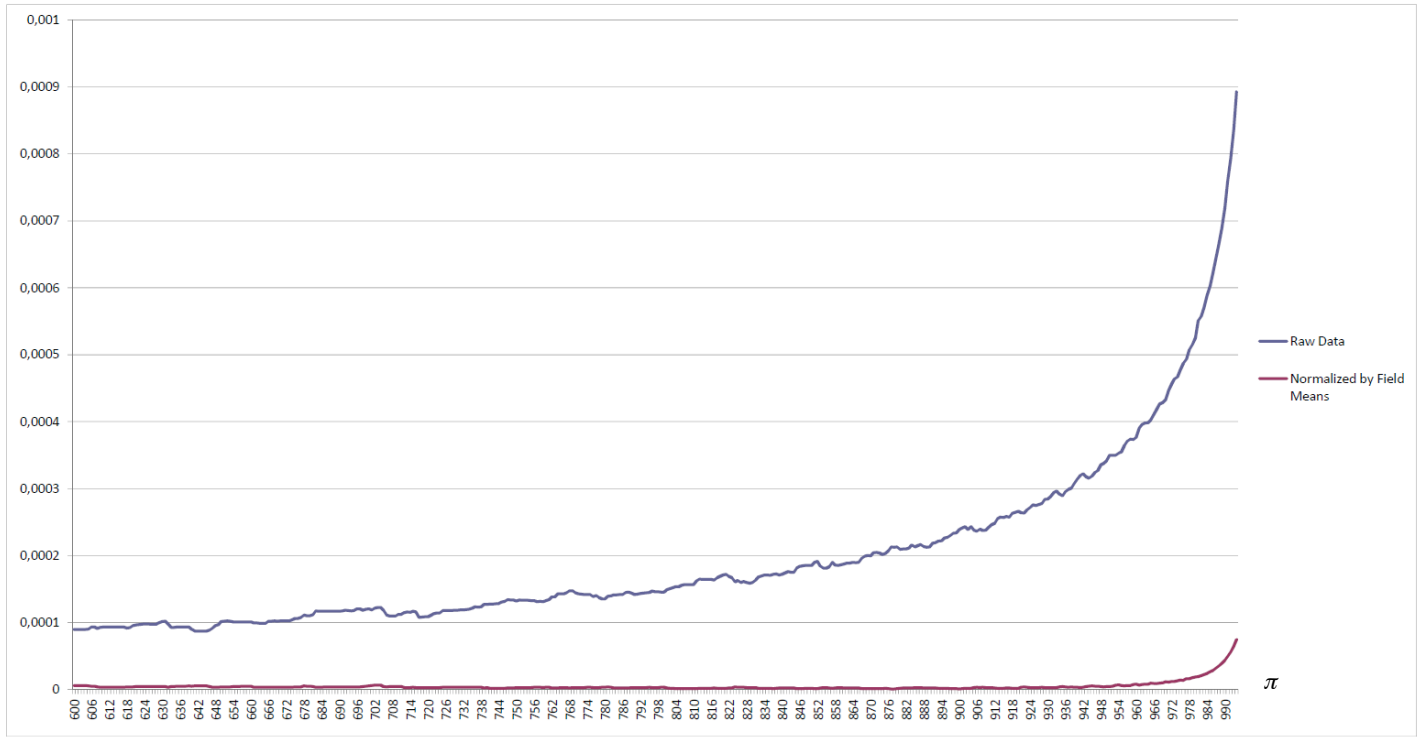


Figure 2. Weighted Citation Inequality Due to Differences in Citation Practices, $\frac{\pi}{v} I(\pi)$ vs. π . Raw vs. Normalized Data

Table 2. Total Citation Inequality Decomposition Before and After Normalization: *ICP* Interval Detail

	Within-group Term, \mathcal{W}	Skew. of Sc. Term, \mathcal{S}	<i>IDCP</i> Term	Total Citation Ineq., $I_I(C)$	Percentages In %:		
					(1)/(4)	(2)/(4)	(3)/(4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. RAW DATA							
All Quantiles	0.0046	0.7488	0.1221	0.8755	0.52	85.53	13.95
[1, 705]			0.0449				5.13
[706, 998]			0.0717				8.18
[999, 1000]			0.0056				0.64
B. EXCHANGE FACTOR NORMALIZATION							
All Quantiles	0.0050	0.7756	0.0156	0.7962	0.6	97.4	1.96
[1, 705]			0.0117				1.47
[706, 998]			0.0018				0.23
[999, 1000]			0.0022				0.27
C. MEAN NORMALIZATION							
All Quantiles	0.0050	0.7761	0.0153	0.7964	0.6	97.4	1.92
[1, 705]			0.0114				1.43
[706, 998]			0.0020				0.25
[999, 1000]			0.0020				0.25

