Multiplicative and Fractional Strategies When Journals Are Assigned To Several Subfields

Neus Herranz
Department of Economics, University of Illinois at Urbana-Champaign, 214 David Kinley Hall, 1407 W. Gregory, Urbana, IL 61801. E-mail: neushg@gmail.com

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

In many data sets, articles are classified into subfields through the journals in which they have been published. The problem is that while many journals are assigned to a single subfield, many others are assigned to several. This article discusses a multiplicative and a fractional strategy to deal with this situation. The empirical part studies different aspects of citation distributions under the two strategies, namely: the number of articles, the mean citation rate, the broad shape of the distribution, their characterization in terms of size- and scale-invariant indicators of high and low impact, and the presence of extreme distributions, that is, distributions that behave very differently from the rest. We found that, despite large differences in the number of articles according to both strategies, the similarity of the citation characteristics of articles published in journals assigned to one or several subfields guarantees that choosing one of the two strategies may not lead to a radically different picture in practical applications. Nevertheless, the characterization of citation excellence through a high-impact indicator may considerably differ depending on that choice.

Introduction

Assume that we are given a hierarchical Map of Science that distinguishes between several aggregation levels, say, between scientific subfields, disciplines, and fields from the lowest to the highest aggregation level. Each category at any aggregate level is assumed to belong to only one item at the next level, so that each subfield belongs to a single discipline, and each discipline to a single field. Assume also that, as in the Thomson Scientific and Scopus databases, publications in the periodical literature are assigned to subfields via the journal in which they have been published. Many journals are assigned to a single subfield, but many others are assigned to two, three, or more subfields. This is an important problem. For example, in the data set used in this article, 42% of the 3.6 million articles published in 1998–2002 are assigned to two or more, up to a maximum of six, subfields, where subfields are identified with the 219 Web of Science (WoS hereafter) categories distinguished by Thomson Scientific.

This article investigates the practical implications arising from this situation. Two issues must be addressed: (a) the allocation of individual publications over the category set at each aggregate level and (b) the normalization procedure when closely related but heterogeneous subfields are brought together into some aggregate category.

We study two ways to solve the problem created when a journal is assigned to several subfields. The first follows a *fractional* strategy, according to which each publication is fractioned into as many equal pieces as necessary, with each piece assigned to a corresponding subfield. Because each subfield is assigned to a single discipline and the same rule applies at higher aggregate levels, the fractional assignment of individual papers to disciplines and fields poses no additional problem, and the total number of publications at each level coincides with the total number of publications in the original data set. (This is the approach often followed in the literature; see also Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011.) The second procedure follows a *multiplicative* strategy, according to which each paper is wholly counted as many times as necessary in the several subfields to which it is assigned. In this way, the space of articles is expanded as much as necessary beyond the initial size in what we call the *subfield extended count*. When this strategy is applied at higher aggregate levels, we end up with

different extended counts in which the total number of publications is always greater than the total number in the original data set. However, for reasons explained below, the size of the extended counts decreases as we move upward in the aggregation scheme.

Second, it is generally agreed that widely different citation practices at the subfield level require some normalization when considering aggregate categories comprising closely related but nevertheless heterogeneous subfields. Under the fractional strategy, the standard procedure is to use the subfield fractional mean citation rate (MCR hereafter) as the normalization factor (see also Waltman et al., 2011a, in the context of average-based indicators of citation impact). However, as will be seen below, under the multiplicative strategy the normalization procedure is not obvious at all. To the best of our knowledge, this article is the first to suggest a reasonable normalization procedure in this case.

The two strategies and their normalization procedures should be evaluated in terms of the properties they satisfy. However, quite apart from the a priori advantages that may make one strategy preferable to another, it is important to verify the order of magnitude of the empirical differences that the alternative methods may bring. In particular, this article studies the following three empirical issues.

- Using size-invariant and scale-invariant statistical techniques it is possible to focus solely on the shape of citation distributions independently of their size and MCR differences. Applying the characteristic scores and scales (CSS hereafter) approach that satisfies these properties, Albarrán, Crespo, Ortuño, and Ruiz-Castillo (2011a) find that the partition of un-normalized citation distributions in the multiplicative case over three broad classes is strikingly similar across 219 subfields, as well as across other aggregate categories built according to several aggregation schemes. Thus, an important issue is whether or not the above-mentioned regularities are maintained for the un-normalized distributions in the fractional case, as well as for the normalized distributions in both cases.
- Using limited evidence that, nevertheless, spans broad areas of science, Radicchi, Fortunato, and Castellano (2008) claim that normalization by subfield means leads to a universal distribution (see also Glänzel, 2010). However, for the multiplicative case, Albarrán et al. (2011a) present evidence against the universality claim across scientific subfields and other aggregate categories (see also Waltman, van Eck, & van Raan, 2012). In this article, we evaluate this issue in terms of the size-invariant and scale-invariant indicators of high impact and low impact introduced in Albarrán, Ortuño, and Ruiz-Castillo (2011b, 2011c, 2011d). The lack of universality will manifest itself through the presence of what we call *extreme distributions*, or citation distributions characterized by truly extreme indicator values.
- It turns out that the broad shape of citation distributions as well as the set of extreme distributions under both strategies are very similar indeed at all aggregate levels. These results seem to suggest that the choice between a multiplicative and a fractional strategy is of lesser importance. But this conclusion is not warranted. Even if citation distributions under both strategies may share a number of basic general characteristics, it is important for the user to isolate those categories at each aggregation level for which there are dramatic differences between the two strategies.

The rest of this article comprises four sections. The Two Strategies section introduces the multiplicative and the fractional strategies, as well as the normalization procedure in the multiplicative case. The Data section presents the data, and the empirical results about the similarities between the multiplicative and the fractional strategies, while the Differences Between the Multiplicative and the Fractional Strategies section is devoted to the differences between them. The Conclusions section offers some concluding comments and suggestions for future work.

The Two Strategies

Suppose we have an initial citation distribution $\boldsymbol{c} = \{c_l\}$ comprising N distinct articles, indexed by $l = 1,\ldots, N$, where c_l is the number of citations received by article l. The total number of citations is denoted by $\gamma = \sum_l c_l$. There are S subfields, indexed by $s = 1,\ldots, S$. Assume for the moment that there is only one other aggregation level comprising $D < S$ disciplines, indexed by $d = 1,\ldots, D$, as well as a rule that indicates the discipline to which each subfield belongs. As indicated in the Introduction section, the problem is that only about 58% of all the articles in our data set are assigned to a single subfield.

The Subfield Level

Let X_l be the nonempty set of subfields to which article l is assigned, and denote by x_l the cardinal of this set, that is, $x_l = |X_l|$. Because, at most, an article is assigned to six subfields, $x_l \in [1, 6]$. In the first step in the multiplicative strategy, each article is wholly counted as many times as necessary in the several subfields to which it is assigned. Thus, if an article l is assigned to three subfields, so that $x_l = 3$, it should be independently counted three times, once in each of the subfields in question, without altering the original number of citations in each case. Consequently, as long as $x_l > 1$ for some article l, the total number of articles in what we call the *subfield extended count*, N_{SF}, is greater than N. Formally, let N_s be the number of distinct articles, indexed by $i = 1,\ldots, N_s$, which are assigned to subfield s. Then, $\boldsymbol{c_s} = \{c_{si}\}$ is the citation distribution in subfield s, where c_{si} is the number of citations received by article i, and $c_{si} = c_l$ for some article l in the original distribution. The subfield extended count, ***SF-count***, is the union of all subfield distributions, namely, $\boldsymbol{SF\text{-}count} = \cup_s \boldsymbol{c_s}$, where $N_{SF} = \sum_s N_s$. For later reference, the MCR in subfield s, M_s, is defined by

$$M_s = \left(\sum_i c_{si}\right)/N_s. \tag{1}$$

In the fractional strategy, subfield s's citation distribution can be described by $\boldsymbol{cf_s} = \{w_{si}\, c_{si}\}$, where $w_{si} = (1/x_l)$ for all

$s \in X_l$ and some article l in the initial distribution for which $c_{si} = c_l$. Therefore, $\sum_{s \in X_l} w_{si} = 1$. The fractional number of articles in subfield s is $n_s = \sum_i w_{si}$, the citations received by each fractional article are $w_{si}\, c_{si}$, and the fractional number of citations in subfield s is $\sum_i w_{si} c_{si}$. Subfield s's MCR, m_s, is defined by

$$m_s = (\textstyle\sum_i w_{si} c_{si}) / (\sum_i w_{si}). \tag{2}$$

By comparing Equations 1 and 2, it should be clear that the difference between the multiplicative and the fractional strategies amounts to a question of weighting. In the first strategy, the N_s distinct articles belonging to subfield s receive a weight equal to one, while in the second strategy, each of these articles is weighted by $w_{si} = (1/x_l)$ for some article l in the initial distribution. It should be noted that $\sum_s n_s = \sum_s \sum_i w_{si} = \sum_l \sum_{s \in X_l} w_{si} = N$ and $\sum_s c_s = \sum_s \sum_i w_{si} c_{si} = \gamma$ that is, in the fractional strategy the total number of articles and citations in the original data set are preserved at the subfield level.

The Discipline Level

Because each subfield belongs to a single discipline at the next aggregation level, there is no particular problem in associating the subfield fractional numbers of articles and citations to the corresponding discipline. As a matter of fact, the discipline distribution in the fractional strategy, $\boldsymbol{cf_d}$, is equal to the union of the corresponding subfield distributions, that is, $\boldsymbol{cf_d} = \cup_{s \in d}\, \boldsymbol{cf_s}$. Again, the number of articles and citations in a particular discipline, $n_d = \sum_{s \in d} \sum_i w_{si}$ and $\sum_{s \in d} \sum_i w_{si} c_{is}$, may typically be fractional. However, the sum of these numbers over all disciplines necessarily coincides with the original ones:

$$\textstyle\sum_d n_d = \sum_d \sum_{s \in d} \sum_i w_{si} = \sum_s \sum_i w_{si} = N,$$

and

$$\textstyle\sum_d \sum_{s \in d} \sum_i w_{si} c_{is} = \sum_s \sum_i w_{si} c_{is} = \gamma.$$

In other words, in the fractional strategy the total number of articles and citations in the original data set are preserved at the discipline level. Consequently, discipline d's MCR, $m_d = (\sum_{s \in d} \sum_i w_{si} c_{is}) / (\sum_{s \in d} \sum_i w_{si})$, is equal to the weighted sum of its subfields' MCRs, with weights equal to the proportion that the number of articles in each subfield represents in the total number of articles in the discipline, that is,

$$m_d = \textstyle\sum_{s \in d} \beta_s m_s, \tag{3}$$

where $\beta_s = (\sum_i w_{si}) / (\sum_{s \in d} \sum_i w_{si})$.

Instead, according to the multiplicative strategy, at the next aggregate level each article is wholly counted as many times as necessary given the several disciplines to which it belongs. Formally, for any article l, let Y_l be the nonempty set of disciplines to which article l is assigned, and let $y_l = |Y_l|$ be the cardinal of this set. At the discipline level, article l is counted y_l times with c_l citations each time. Of course, $y_l \leq x_l$ for all l. Let N_d be the number of distinct articles in discipline d, and denote by $\boldsymbol{c_d} = \{c_{dj}\}$ the citation distribution in discipline d, where c_{dj} is the number of citations received by article $j = 1, \ldots, N_d$. Thus, there must exist at least one subfield s belonging to d, some $i = 1, \ldots, N_s$, and some article l in the original distribution such that $c_{dj} = c_{si} = c_l$. The *discipline extended count*, ***D-count***, is the union of all discipline distributions, namely, $\boldsymbol{D\text{-}count} = \cup_d\, \boldsymbol{c_d}$, where $N_D = \sum_d N_d$ is the number of articles in the discipline extended count. Since $D < S$, as long as there exists some l and some d for which $y < x_l\ N_d < \sum_{s \in d} N_s$ and $N_D < N_{SF}$. The MCR of distribution $\boldsymbol{c_d}$, M_d, is defined by

$$M_d = \gamma_d / N_d, \tag{4}$$

where $\gamma_d = \sum_j c_{dj}$ is the total number of citations in $\boldsymbol{c_d}$. Because the link between the two levels is broken,

$$M_d \neq \textstyle\sum_{s \in d} \alpha_s M_s,$$

where $\alpha_s = N_s / N_d$, and the mean M_s and M_d are defined in Equations 1 and 4, respectively. However, there is an expression similar to Equation 3 for M_d. To show this, we need to introduce some more notations. For any $d \in Y_l$, let $X_{ld} \subseteq X_l$ be the nonempty set of subfields in X_l that belong to discipline d, and let $x_{ld} = |X_{ld}|$ be the number of subfields in X_{ld}. Finally, for any s, let $\boldsymbol{c'_s} = \{v_{si}\ c_{si}\}$ be a new subfield distribution where

$$v_{si} = 1/x_{ld} \text{ for all } s \in X_{ld},$$

so that $\sum_{s \in X_{ld}} v_{si} = 1$. It turns out that the number of articles and citations in the union of the new subfield distributions, $\cup_{s \in d}\, \boldsymbol{c'_s}$, coincides with N_d and γ_d, respectively. To see this, for any article l assigned to some subfield s that belongs to some discipline d, we must consider two possibilities depending on the cardinality of x_l.

(i) Assume that $x_l = 1$, so that $X_l = \{s\}$ is a singleton. Then, there exists some $i = 1, \ldots, N_s$ for which $c_{si} = c_l$. Because subfield s belongs to discipline d, we have: $Y_l = \{d\}$. Then there exists a single article $j = 1, \ldots, N_d$ with $c_{dj} = c_{si} = c_l$. On the other hand, $X_{ld} = X_l$, and $y_l = x_{ld} = x_l = 1$, so that $v_{si} = 1/x_{ld} = 1$, and $v_{si}\ c_{si} = c_l$. Therefore, article l is counted once in $\cup_{s \in d}\, \boldsymbol{c'_s}$ and receives c_l citations.

(ii) Assume that $x_l > 1$, so that X_l comprises several subfields. Note that for every $s \in X_l$, there exists some $i = 1, \ldots, N_s$ for which $c_{si} = c_l$. Next, we must consider three cases. (ii.a) If all subfields in X_l belong to a single discipline, then $Y_l = \{d\}$ with $y_l = 1$, and there exists a single $j = 1, \ldots, N_d$ such that $c_{dj} = c_{si} = c_l$ for every $s \in X_l$. On the other hand, $X_{ld} = X_l$ with $x_{ld} = x_l$, $\sum_{s \in X_{ld}} v_{si}$ is always equal to one, and $\sum_{s \in X_{ld}} v_{si} c_{si} = \sum_{s \in X_{ld}} (c_l / x_{ld}) = c_l$. Therefore, as before, article l is counted once in $\cup_{s \in d}\, \boldsymbol{c'_s}$ and receives c_l citations. (ii.b) If each subfield in X_l belongs to a different discipline, then $y_l = x_l$, and article l is counted y_l times at the discipline level with c_l citations each time. In particular, for each $d \in Y_l$, there exists some $j = 1, \ldots, N_d$ with $c_{dj} = c_l$. On the other hand, for each $d \in Y_l$, we have that X_{ld} is a singleton with $x_{ld} = 1$, so that

$\sum_{s\in X_l} v_{si} = \sum_{d\in Y_l}\sum_{s\in X_{ld}} v_{si} = x_l$, and $v_{si}\ c_{si} = c_l$ for each $s \in X_l$. Therefore, article l will be counted $y_l = x_l$ times in $\cup_{s\in d}\ \boldsymbol{c'_s}$, each time receiving c_l citations. (ii.c) If some subfields in X_l belong to a certain discipline and some others belong to one or several more disciplines, then $1 < y_l < x_l$ and article l is counted y_l times at the discipline level with c_l citations each time. On the other hand, $X_l = \bigcup_{d\in Y_l} X_{ld}$ with $x_l = \sum_{d\in Y_l} x_{ld}$. In this case, $\sum_{s\in X_{ld}} v_{si} = 1$ for each $d \in Y_l$, so that $\sum_{s\in X_l} v_{si} = y_l$. Therefore, article l is counted y_l times in $\cup_{s\in d}\ \boldsymbol{c'_s}$, each time receiving $\sum_{s\in X_{ld}} v_{si}c_{si} = c_l$ citations. Thus, in the previous example with $x_l = 3$ for some l, assume that the first two subfields belong to one discipline whereas the third belongs to another discipline, so that $y_l = 2$. In the multiplicative strategy, article l is counted three times at the subfield level but only twice at the discipline level.

As stated above, we conclude that γ_d is equal to the total number of citations in $\cup_{s\in d}\ \boldsymbol{c'_s}$, and N_d is equal to $\sum_{s\in d} N'_s$, where $N'_s = \sum_i v_{si}$ is the possibly fractional number of articles in the new subfield distribution $\boldsymbol{c'_s}$. Thus, we can obtain an expression analogous to Equation 3, namely:

$$\begin{aligned} M_d &= \left(\textstyle\sum_{s\in d}\sum_i v_{si}c_{si}\right)/\left(\sum_{s\in d}\sum_i v_{si}\right) \\ &= \left(\textstyle\sum_{s\in d}\left[N'_s\left[\sum_i v_{si}c_{si}/N'_s\right]\right]\right)/\left(\sum_{s\in d}\sum_i v_{si}\right) \\ &= \textstyle\sum_{s\in d}\left(N'_s/N_d\right)M'_s, \end{aligned}$$

where M'_s is the new subfield s's MCR defined by

$$M'_s = \left(\textstyle\sum_i v_{si}c_{si}\right)/\left(\sum_i v_{si}\right), \tag{5}$$

By comparing Equations 1 and 5, it should be clear that the difference between the multiplicative strategy at the subfield and the discipline level amounts to a question of weighting. In the first case, the N_s distinct articles belonging to subfield s receive a weight equal to one, while in the second case, an article l in the original distribution belonging to a new subfield s and discipline d is weighted by the inverse of the number of subfields belonging to discipline d, namely, is weighted by $v_{si} = (1/x_{ld})$, so that the MCR at the discipline level is seen to be equal to the weighted sum of its new subfields MCRs, with weights equal to the proportion that the number of articles in each new subfield represents in the total number of articles in the discipline.

Normalization Procedures

As indicated in the Introduction, whenever possible we must normalize aggregate distributions, say, at the discipline level, taking into account differences in citation practices across their subfields. In the fractional case, normalization is straightforward. The normalized distribution of subfield s, $\boldsymbol{zf_s}$, is simply equal to the original one, where each fractional article is divided by the fractional subfield mean m_s defined in Equation 2. Discipline d's normalized distribution, $\boldsymbol{zf_d}$, is simply equal to the union of the corresponding $\boldsymbol{zf_s}$ distributions. Thus, $\boldsymbol{zf_s} = \{\boldsymbol{zf_s}/m_s\} = \{(w_{si}\ c_{si})/m_s\}$ for all s belonging to d, and $\boldsymbol{zf_d} = \cup_{s\in d}\ \boldsymbol{zf_s}$. Of course, the MCRs of distributions $\boldsymbol{zf_s}$ and $\boldsymbol{zf_d}$ for all s and all d are equal to one.

Discipline d's normalized distribution in the multiplicative case is $\boldsymbol{z_d} = \{z_{dj}\}$, where

$$z_{dj} = c_{dj}\textstyle\sum_{s\in X_{ld}}\left(v_{si}/M'_s\right) = \left(c_l/x_{ld}\right)\sum_{s\in X_{ld}}\left(1/M'_s\right),$$

and M'_s is defined in Equation 10. For each s belonging to d, let $\boldsymbol{z'_s} = \{c'_s/M'_s\} = \{(v_{si}\ c_{si})/M'_s\}$ be the new subfield normalized distribution. As before, the MCR of the normalized distribution $\boldsymbol{z_d}$ is seen to be equal to the MCR of the union $\cup_{s\in d}\ \boldsymbol{z'_s}$. Of course, the MCRs of distributions $\boldsymbol{z'_s}$ and $\boldsymbol{z_d}$ for all s and all d are equal to one. (Appendix I in the Working Paper version of this paper, Herranz and Ruiz-Castillo, 2011a [HR-C hereafter], contains a numerical example in which the two strategies and the corresponding normalization procedures are illustrated.)

To understand the procedure at higher aggregate levels, say, for F fields with $F < D$, indexed by $f = 1,\ldots, F$, it suffices to redefine Y_l as the nonempty set of fields to which article l is assigned, and X_{lf} as the nonempty set of subfields in X_l that belongs to field f in Y_l. Then, as before, if $x_{lf} = |X_{lf}|$ is the number of subfields in X_{lf}, then for any s, let $\boldsymbol{c''_s} = \{u_{si}\ \ c_{si}\}$ be a new subfield distribution where $u_{si} = 1/x_{lf}$ for all $s \in X_{lf}$, so that $\sum_{s\in Xlf} u_{si} = 1$. The new fractional number of articles in subfield s is equal to $N''_s = \sum_i u_{si}$, and the new MCR of distribution $\boldsymbol{c''_s}$ is denoted by M''_s. The number of distinct articles in the field distribution $\boldsymbol{c_f}$, N_f, is seen to be equal to $\sum_{s\in f} N''_s$, and the MCR of $\boldsymbol{c_f}$, M_f, is equal to the weighted sum of its new subfields MCRs, with weights equal to the proportion that the number of articles in each new subfield represents in the total number of articles in the field:

$$M_f = \textstyle\sum_{s\in f}\left(N''_s/N_f\right)M''_s.$$

The *field extended count*, ***F-count***, is the union of all discipline distributions, namely, $\boldsymbol{F\text{-}count} = \cup_f\ \boldsymbol{c_f}$, where $N_F = \sum_f N_f$ is the number of articles in the field extended count with $N < N_F < N_D < N_{SF}$. From this point, normalization proceeds as in the discipline case. Eventually, when we reach the maximum aggregation level the weighting system in the multiplicative strategy coincides with the one in the fractional strategy.

A Priori Evaluation of Both Procedures

The preservation of the total number of papers and citations at each aggregate level in the fractional case lends this strategy an aura of "normalcy." However, the fractional strategy is not beyond criticism. First, assume that there are two articles assigned to a certain subfield. The first article is assigned only to this subfield, while the second is also assigned to other subfields. Why should the weights associated to both articles in computing any statistic, such as the MCR, for example, be entirely different as implied by the fractional strategy? It can be argued that in the study of any subfield, all articles should count equally regardless of the role some of them may play on other

subfields.[1] Of course, as we have seen, at the lowest aggregation level this leads to an artificially large subfield extended count. However, this is not that worrisome in the sense that because this strategy does not create any interdependencies among the subfields involved, it is still possible to separately investigate every subfield in isolation, independently of what takes place in any other subfield.

Similarly, consider a situation in which two articles are assigned to the same discipline, but one is assigned only to a single subfield, and hence to only that discipline, and the other is assigned to several subfields and possibly to other disciplines. In the fractional strategy, the second article will be weighted by $1/x_l$, while in the new subfield according to the multiplicative strategy, it will be weighted only by $1/x_{ld}$ where $x_{ld} < x_l$. Consequently, in this discipline, the second article's citations in the multiplicative approach will be c_l, while in the fractional approach, they will be $\sum_{s \in X_{ld}} w_{si} c_{si} = \sum_{s \in X_{ld}} (1/x_l) c_l = (x_{ld}/x_l) c_l$. Why should the role of the second article be diminished as much as demanded by the fractional strategy, when in the study of any discipline, all articles should count equally regardless of the role some of them may play in other disciplines? This is the reason why, in their study of citation distributions, Albarrán et al. (2011a) follow a multiplicative strategy at all aggregate levels.

Second, assume without loss of generality that we want to evaluate the citation impact of different research units in a certain subfield (as before, a similar argument can be offered when the evaluation is performed at any other aggregate level). In the computation of any citation impact indicator, a fractional strategy reduces the role of articles published in journals assigned to several subfields. Therefore, this strategy would hurt relatively more those research units with highly cited articles of this type. It can be argued that, from a normative point of view, this implication distorts the evaluation of research units in a given subfield. This is the additional reason why in their comparison of citation impact performance for three geographical areas (the United States, the European Union, and the rest of the world), Herranz and Ruiz-Castillo (2011b, 2012b, in press) also follow a multiplicative strategy.

Admittedly, others will see the issue differently depending on, among other things, the particular view one has about the criteria used in the assignment of journals to subfields. The more credit you attach to such criteria, the more you might be in favor of a multiplicative strategy. However, we may all agree that knowing the empirical consequences of following the two strategies is worth investigating. This is the topic explored in the rest of the article.

Data and Similarities Between the Multiplicative and the Fractional Strategies

This section deals with four issues: the data, the characteristics of the shape of citation distributions, the high- and low-impact citation indicators, and what we call "extreme distributions."

The Data

Because we wish to address a homogeneous population, in this study only research articles such are studied. The data set comprises about 3.6 million articles published in 1998–2002, and the 28 million citations they receive after a common 5-year citation window for every year.

As indicated in the Introduction, subfields are identified with the 219 WoS categories distinguished by Thomson Scientific. To facilitate the reading of results, it will be useful to classify these subfields into other aggregate categories. The difficulty, of course, is how to construct a Map of Science—a question that is known to have no easy answer. In this article, we use a scheme comprising 80 intermediate categories, or disciplines, and 20 fields (for details, see HR-C, 2011a).[2]

As explained in the previous section, in the multiplicative strategy, the number of articles in the different extended counts is always greater than the number of articles in the original data set, and decreases as we move upwards in the aggregation scheme: The subfield extended count has more than 5.7 million articles, or 57.1% more than the number of articles in the original data set, while disciplines and fields lead to extended counts about 47%, and 34% larger than the original data set.

Characteristics of the Shape of Citation Distributions

We know that the broad shapes of un-normalized citation distributions in the multiplicative case are highly skewed and strikingly similar at all aggregation levels (see also Albarrán & Ruiz-Castillo, 2011; Albarrán et al., 2011a; Schubert, Glänzel, & Braun, 1987; Seglen, 1992). Therefore, it is very important to verify whether this is also the case for the original distributions in the fractional strategy at the subfield level, and for the normalized distributions according to both strategies at all aggregate levels.

Size-independent and scale-independent descriptive tools permit us to focus on the shape of distributions. In particular, the CSS approach, pioneered by Schubert et al. (1987) in

[1]We would like to take this opportunity to correct the idea that "*fractionally assigned articles have a much smaller chance of occupying the upper tail of citation distributions than articles assigned to a single WoS category*" (Albarrán et al., 2011a, p. 389). Fractionally assigned articles would play a smaller role than articles assigned to a single subfield, but they would have the same chance of occupying the upper tail of citation distributions.

[2]We should make clear that it is not claimed that this aggregation scheme provides an accurate representation of the structure of science. It is rather a convenient simplification or a realistic tool for the discussion of the aggregation issue in this paper.

TABLE 1. Characteristic scores and scales (means and standard deviations).

	Percentage of articles in categories		Percentage of citations in categories	
	1 + 2	4 + 5	2	4 + 5
A. Un-normalized subfields				
Multiplicative strategy[a]	68.6	10.0	21.1	44.9
	(3.7)	(1.7)	(5.0)	(4.6)
Fractional strategy	68.3	10.2	21.5	44.7
	(3.4)	(1.6)	(4.2)	(3.9)
b. Normalized disciplines				
Multiplicative strategy	68.4	10.0	22.3	43.9
	(2.6)	(1.3)	(3.2)	(2.9)
Fractional strategy	68.4	10.0	21.8	44.5
	(2.8)	(1.3)	(3.3)	(3.0)
c. Normalized fields				
Multiplicative strategy	68.7	9.7	21.6	44.6
	(1.8)	(1.0)	(3.4)	(3.3)
Fractional strategy	68.7	9.7	21.1	45.1
	(2.0)	(1.1)	(3.5)	(3.3)

[a]The information in this row is taken from Table 6 in the Working Paper version of Albarrán et al. (20011a).

citation analysis, permits the partition of any distribution of articles into five convenient classes according to the citations they receive. Denote by s_1 the MCR; by s_2 the mean citation of articles above s_1, and by s_3 the mean citation of articles above s_2. The first category includes articles without citations. As for the remaining four, articles are said to be *poorly cited* if their citations are below s_1, *fairly well cited* if they are between s_1 and s_2, *remarkably cited* if they are between s_2 and s_3, and *outstandingly cited* if they are above s_3. For the partition of citation distributions at the subfield level into three broad classes—comprising categories 1+2, 3, and 4+5—, the relevant information at different aggregate levels according to both strategies is in Table 1. (For the individual information for the un-normalized and the normalized distributions in both strategies at all aggregate levels, see HR-C, 2011a.)

According to Albarrán et al. (2011a), approximately 69% of all articles in the multiplicative case at the subfield level receive citations below the mean and account for about 21% of all citations, while articles with a remarkable or outstanding number of citations represent about 10% of the total, and account for approximately 45% of all citations. This is exactly what we find for the un-normalized distributions in the fractional case at the subfield level, as well as for the normalized distributions according to both strategies at the discipline and field levels. In brief, the partition into three broad citation categories is, approximately, 69/21/10 of all articles, accounting for 21/34/45 of all citations.

However, when we move inside the union of categories 1 and 2 and categories 4 and 5, differences across categories at all aggregation levels become much larger (see HR-C, 2011a, for details). Thus, dispersion statistics formally reveal that the universality of citation distributions breaks down at both the lower and the upper tails at all aggregation levels. This conclusion contrasts with the more optimistic view offered by Radicchi et al. (2008) with a methodology that does not explain whether a multiplicative or a fractional strategy has been used, omits articles without citations, examines distributions at a limited set of points and, above all, covers only 14 of the 219 subfields. In addition, Albarrán et al. (2011a) find considerable differences in the power law characteristics of 140 un-normalized subfield distributions and a variety of un-normalized aggregate distributions in the multiplicative case. Thus, the lack of universality is particularly apparent at one key segment of citation distributions: the tip of the upper tail, or the place where citation excellence resides. The estimation of power laws is beyond the scope of this article. However, in the remainder of this section, we pursue the study of the lack of universality by detecting the presence of extreme distributions, or citation distributions characterized by extreme values of certain indicators.

High-Impact and Low-Impact Citation Indicators

As we have seen, citation distributions are highly skewed in the sense that a large proportion of articles receive no or few citations, while a small percentage accounts for a disproportionate amount of all citations. An important consequence is that average-based indicators may not adequately summarize these distributions for which the upper and lower parts are typically very different. This leads to the idea of using two indicators to describe any citation distribution: a *high-impact* and a *low-impact measure* defined over the set of articles with citations above or below a *critical citation line* (CCL hereafter). In the first empirical application of this methodology, Albarrán et al. (2011c) use a family of high-impact and low-impact indicators that satisfies a number of desirable properties. In this article, we use one high-impact and one low-impact indicator, denoted by H and L, which are members of these families (for a brief presentation of these indicators and their main properties, see Appendix III in HR-C, 2011a). The reason for using these indicators is twofold.

First, while average-based measures are silent about the distributive characteristics on either side of the mean, H and L are sensitive to the citation inequality in the sense that an increase in the coefficient of variation increases both of them. Second, it is well known that wide differences in publication and citation practices give rise to wide differences in size and MCR across subfields. However, in this article, we are interested in studying distributions that are very different from the rest abstracting from differences in those two characteristics. Fortunately, H and L allow us to pursue this aim because they are size-invariant and scale-invariant, namely, the value they take is invariant under replication and scalar multiplication of citation distributions.

Extreme Distributions

In this article, the CCL is always fixed at the 80th percentile of all citation distributions. (For individual information about high-impact and low-impact levels according to the *H* and *L* indicators in the multiplicative and the fractional case at the subfield level, see Table B in Appendix II in HR-C, 2011a.)

Starting with the low-impact phenomenon, it is observed that the mean and the median of the 219 values that *L* takes in the multiplicative case at the subfield level practically coincide, and the standard deviation is very small. Only 59 of 219 subfields are slightly above or below the mean plus one standard deviation, and only five distributions can be considered as mildly extreme. The correlation coefficient of *L* values according to the two strategies is 0.96, and the analysis in the fractional case leads to exactly the same five mildly extreme distributions isolated in the multiplicative case. At the discipline and the field levels (individual information available on request), only the Multidisciplinary category deserves to be mentioned as a potential extreme distribution under both strategies. The conclusion is that for truly different behavior we must turn to what we call the structure of excellence at the upper tail of citation distributions.

Turning towards the high-impact phenomenon, we begin by noting that the distributions of *H* values at the subfield level for the two strategies are highly correlated (correlation coefficient equal to 0.96), and present similar general characteristics. In the multiplicative case, for example, the standard deviation and the coefficient of variation take very large values, and the mean is very much greater than the median, all of which indicates that the distribution of *H* values is highly skewed to the right and it is likely to present some important extreme cases. Panel A in Table 2 includes the 17 subfields with the highest *H* values in the multiplicative case, as well as five subfields with high *H* values in the fractional case that are not included in the previous set (subfields 18–22).

The following three points should be emphasized:

- The set of extreme distributions, comprising 8 or 22 distributions, depending on the critical *H* values we choose, is very similar indeed according to both strategies.
- There is no systematic tendency for *H* values to be greater according to one of the two strategies. Surely the most notable case is Statistics and Probability, where the *H* value in the multiplicative case is almost 100% greater than in the fractional case.
- Within the set of extreme distributions, the following comments are in order. First, two subfields—Crystallography and Medicine, Research and Experimental—were already characterized as "residual subfields" in Albarrán et al. (2011a). Second, six of eight subfields in computer science are considered extreme. The conclusion is inescapable: This field's structure of excellence is entirely different from the rest. Third, two important subfields within Physics are classified as extreme: Physics, Particle and Fields, and Physics, Multidisciplinary. Fourth, perhaps not surprisingly, the Multidisciplinary category behaves as a mildly extreme distribution at the subfield level. Fifth, only two social sciences can be considered as mildly extreme subfields: International Relations and Ethnic Studies.

TABLE 2A. Extreme un-normalized subfield distributions according to the multiplicative and the fractional approach.

High-impact values	Multiplicative (1)	Fractional (2)	(3) = (1) – (2) in %
1. Medicine, General and Internal	20.7	22.3	–7.2
2. Crystallography	17.7	17.2	2.7
3. Mathematical and Computational Biology	15.5	11.8	32.0
4. Statistics and Probability	14.8	7.6	93.1
5. Computer Science, Interdisciplinary Applications	12.9	9.9	29.5
6. Biochemical Research Methods	5.2	3.7	40.8
7. Physics, Particle and Field	3.7	4.0	–6.6
8. Medicine, Research and Experimental	3.0	3.5	–15.2
9. Engineering, Petroleum	1.1	4.7	–76.7
10. Physics, Multidisciplinary	3.1	3.3	–7.7
11. Computer Science, Information Systems	3.3	2.8	20.1
12. Computer Science, Hardware and Architecture	2.8	2.3	25.6
13. Computer Science, Theory and Methods	2.8	1.9	42.2
14. Multidisciplinary Sciences	2.1	2.2	–0.7
15. Computer Science, Artificial Intelligence	2.1	1.8	15.8
16. Biotechnology and Applied Microbiology	2.1	2.1	–2.7
17. Telecommunications	2.0	1.7	13.0
18. International Relations	1.9	2.3	–16.1
19. Materials Science, Characterization and Testing	1.8	1.8	–3.6
20. Psychology, Multidisciplinary	1.4	2.0	–31.1
21. Mining and Mineral Processing	1.3	2.0	–36.2
22. Ethnic Studies	1.1	2.3	–51.2
Mean subfield value	1.1	1.1	
Standard deviation	2.4	2.2	

At higher aggregate levels, together with the original distributions, we should take into account the normalized distributions according to both strategies. Panel B in Table 2 lists the disciplines and fields with the highest *H* values in both scenarios. (Individual information in the multiplicative and the fractional case is available in Table C in Appendix II in HR-C, 2011a.)

As expected, extreme *H* values decrease with normalization. The ranking of the first two disciplines remains unchanged after normalization, but as soon as differences in subfield MCRs are taken into account, Applied Mathematics and Particle and Nuclear Physics, which appear as third and

TABLE 2B. Extreme discipline and field distributions in the un-normalized and the normalized case.

Un-normalized discipline distributions	Multiplicative (1)	Fractional (2)	(3) = (1) – (2) In %	Normalized discipline distributions	Multiplicative (1)	Fractional (2)	(3) = (1) – (2) In %
1. Crystallography	17.7	17.2	2.7	1. Crystallography	17.7	17.2	2.7
2. General & Int. Med.	8.4	8.3	1.0	2. General & Int. Med.	4.6	5.1	–9.0
3. Applied Mathematics	5.9	2.5	136.3	3. Comp. Sci. & Inf. Tech.	3.6	2.8	29.5
4. Comp. Sci. & Inf. Tech.	5.4	5.5	–2.4	4. Applied Mathematics	3.5	2.5	36.3
5. Part. & Nucl. Physics	3.2	3.5	–8.1	5. Medicine, Res. & Exp.	3.0	3.5	–15.2
6. Medicine, Res. & Exp.	3.0	3.5	–15.2	6. Multidisciplinary Physics	2.2	2.4	–7.2
7. Multidisciplinary Physics	2.9	2.8	4 2	7. Part. & Nucl. Physics	2.2	2.7	–20.2
8. Multidisciplinary	2.1	2.2	–0.7	8. Multidisciplinary	2.1	2.2	–0.7
Mean values	1.3	1.2		Mean values	1.1	1.1	
Standard deviation	2.2	2.1		Standard deviation	2.0	2.0	

Un-normalized field distributions	Multiplicative (1)	Fractional (2)	(3) = (1) – (2) in %	Normalized field distributions	Multiplicative (1)	Fractional (2)	(3) = (1) – (2) in %
Mathematics	6.3	2.2	180.8	Computer Science	3.6	2.8	29.5
Computer Science	5.4	5.5	–2.4	Resid. subfields	3.0	3.7	–17.3
Resid. subfields	4.1	4.8	–15.1	Mathematics	2.4	1.6	53.3
Multidisciplinary	2.1	2.2	–0.7	Multidisciplinary	2.1	2.2	–0.7
Mean values	1.6	1.5		Mean values	1.2	1.2	
Standard deviation	1.7	1.4		Standard deviation	0.9	0.8	

fifth disciplines among the original distributions, now occupy rank four and seven among normalized distributions. A similar phenomenon takes place among fields: Because of the extreme behavior displayed by the Statistics and Probability subfield, Mathematics appears as the first extreme distribution among un-normalized fields. However, as soon as the low MCRs of other mathematical subfields are taken into account in the normalization process, Mathematics goes down to occupy rank three among normalized field distributions.

Interestingly enough, there is now complete agreement between the multiplicative and the fractional strategies about extreme sets. The main difference is the ranking of Applied Mathematics and Mathematics at the discipline and the field levels, respectively, which is always higher in the multiplicative case. The reason, of course, is the large difference already noted about Statistics and Probability at the subfield level.

Not surprisingly, disciplines comprising single extreme subfields remain extreme at the discipline level. Not surprisingly either in view of results at the subfield level, Computer Science is a clear extreme distribution among both disciplines and fields.

Differences Between the Multiplicative and the Fractional Strategies

This section presents results on two issues: the number of articles according to the two strategies and other distribution characteristics that include the MCR and the low- and high-impact values.

The Number of Articles According to the Two Strategies

By construction, differences between the multiplicative and the fractional strategies start with the number of articles. (The individual information is in Table D in Appendix II in HR-C, 2011a.) The following three points should be emphasized.

- In our data set, there is no information about the distribution of subfields, disciplines, or fields by size, measured by the number of people working in them, but the numbers must be very different indeed. Moreover, publication practices vary very much across categories at every aggregate level. In some cases authors publishing one article per year would be among the most productive, while in other instances, authors—either alone or as members of a research team—are expected to publish several papers per year. Consequently, distribution sizes measured by the number of articles are expected to differ at all aggregation levels. In particular, judging by the large dispersion measures, subfield sizes according to both strategies are very different indeed.
- Interestingly enough, the correlation coefficient between subfield sizes according to the multiplicative and the fractional strategies is 0.98. The question the potential user needs to know is whether or not the differences are uniform across categories at each aggregate level. Focusing on the important subfield case, the median of the distribution of the differences between the number of articles according to both strategies is

about 64%, or seven points above the mean. Correspondingly, there are 58 of 219 subfields in which the number of articles in the multiplicative case is at least 100% greater than in the fractional case, while there are only 17 subfields in which this difference is below 20%.

- On the other hand, differences between the two strategies tend to diminish as we proceed towards higher aggregate levels. Thus, there are 3 of 80 disciplines (and 2 of 20 fields) in which the number of articles in the multiplicative case is at least 100% (or 60%) greater than in the fractional case, while only in the Multidisciplinary subfield, which appears as a single discipline and a single field, this difference is below 10%.
- A final interesting question is whether size differences increase with size. A correlation coefficient of –0.19 between these two variables in the subfield case indicates that this is not the case.

Other Characteristics: MCR, L, and H

The final question that needs to be investigated is the differences between the two strategies on dimensions other than size. In particular, we study differences in MCR, and the *L* and *H* indicators that are size-invariant and scale-invariant. The evidence (see Table E in Appendix II in HR-C, 2011a) deserves the following three comments.

- In a majority of cases, 136 subfields, the MCR is greater according to the multiplicative strategy. However, the opposite is the case in a nonnegligible number of cases: 82 subfields.
- In spite of very large differences in the number of articles according to both strategies, differences in MCRs are rather small: They amount to less than 5% in 114 subfields, and between 5% and 10% in another 59 cases. On the other hand, the correlation coefficient between MCRs according to both strategies is very high: 0.98.
- The correlation coefficient between differences in MCRs in absolute terms and differences in size is -0.01, an indication that to have a large number of articles in journals assigned to multiple subfields is not a sufficient condition for large MCR differences between the multiplicative and the fractional strategies.

Turning now to the low-impact phenomenon, it is observed that choosing either of the two strategies has truly minor consequences. However, differences in *H* values are rather significant. As can be seen in Table 3 (that summarizes the individual information in Tables B and C in Appendix II in HR-C, 2011a), in 120 of 219 subfields, 17 of 80 disciplines, and 4 of 20 fields, differences in *H* values between the two strategies are greater than 10%. Moreover, in 30 subfields and one discipline these differences exceed 30%. Thus, when we measure citation impact excellence with the *H* indicator with a CCL fixed at the 80th percentile of world distributions, the quantitative picture drawn through the multiplicative and the fractional strategies is quite different indeed. Nevertheless, the correlation coefficient of this indicator for the two strategies is 0.85 and 0.99 at the subfield and discipline levels, while, as we saw in the Extreme Distributions section, the set of high-impact extreme distributions for the two strategies is very similar indeed.

TABLE 3. Differences in high-impact values between the multiplicative and the fractional strategies at different aggregation levels.

A. Subfields

	0–10%	10–20%	20–30%	30–50%	>50%
Multiplicative > fractional	40	30	17	12	3
Multiplicative < fractional	59	26	16	12	3
Total	99	56	33	24	6

B. Disciplines

	0–10%	10–20%	20–30%	>30%
Multiplicative > fractional	16	9	7	1
Multiplicative < fractional	29	10	4	4
Total	45	19	11	5

C. Fields

	0–10%	10–20%	>20%
Multiplicative > fractional	3	2	2
Multiplicative < fractional	8	5	–
Total	11	7	2

Conclusions

The assignment of a number of journals to multiple subfields poses serious practical problems in many data sets. In this article, we have compared two alternative strategies to cope with this situation: a multiplicative strategy, according to which articles should be wholly counted as many times as necessary when the journal in which they have been published is assigned to several subfields, and a fractional strategy, in which articles should be weighted by the inverse of the number of subfields to which the publishing journal is assigned. Moreover, we have introduced a novel normalization procedure that in the construction of aggregate categories in the multiplicative case takes into account differences in MCRs across subfields at the lowest aggregation level.

Quite independently from the fact that we prefer the first solution on a priori grounds, the main empirical conclusions can be summarized in the following three points.

- By construction, the number of articles according to the multiplicative strategy is always greater than the number of articles in the fractional strategy. At a maximum—at the lowest aggregation level—this difference is 57%. More importantly, differences between the two strategies are far from uniform across categories at different aggregation levels.

- It turns out that—in certain respects—the citation characteristics of articles coming from journals assigned to multiple subfields do not differ much from the rest. Thus, in spite of the wide differences in the mix between the two types of articles, the two strategies lead to un-normalized and normalized citation distributions that have many important features in common. First, MCRs for individual subfields according to the two strategies are not very different from each other. Furthermore, the MCR distributions according to the two strategies are highly correlated. Second, normalized and un-normalized citation distributions according to either the multiplicative or the fractional strategies share the same skewed shape. The proportion of articles that receive (1) none or few citations, (2) are fairly cited, and (3) are remarkably or outstandingly cited is, approximately, 69/21/10. These three classes of articles account for the proportions 21/34/45 of all citations. Third, the measures of low-impact according to both strategies are very close to each other.
- There is no question that the most important part of citation distributions is the upper tail. By fixing the CCL at the 80th percentile, this article focuses attention on the 20% of most highly cited articles. The main conclusion is that excellence is not equally structured in all citation distributions. It turns out that this structure is differently captured by our high-impact indicator under the two strategies in contention: In 63 of 219 subfields, 16 of 80 disciplines, and 2 of 20 fields, differences in H values between the two strategies are greater than 20%. On the other hand, there is a set of extreme citation distributions that behave very differently from the rest in the sense that they are characterized by a very high H value. An important finding in this article is that this set essentially coincides under the multiplicative and the fractional strategies.

In brief, although the similarity of citation characteristics of articles published in journals assigned to one or several subfields guarantees that choosing one of the two strategies may not lead to a radically different picture in practical applications, the list of categories with high-impact values at any aggregate level may considerably differ depending on that choice.

Four possible extensions might be mentioned. First, it is worthwhile to explore whether the main conclusions of the article are robust in relation to the CCL choice. Second, as indicated in the Characteristics of the Shape of Citation Distributions section, Albarrán et al. (2001a) investigated the existence of a power law representing the very top of the upper tail of un-normalized citation distributions in the multiplicative case. It would be certainly interesting to extend this work to the fractional case, as well as to normalized distributions under both strategies. Third, it should be noted that our high-impact indicator is not robust to the presence of a handful of articles with a truly phenomenal number of citations. Therefore, it would be interesting to explore the issue of extreme distributions using indicators of citation excellence robust to extreme observations. Fourth, an important research question is to explain why excellence is not equally structured in all citation distributions, and why in some of them it behaves so differently from the rest.

We should not end this article without pointing out how convenient it would be to have a classification system available in which each article could be assigned to a single subfield. Thomson Scientific does that for the data set used in this article, but only for a notion of "subfield" that leads to a set of only 22 broad categories. (This is the classification system used in Albarrán & Ruiz-Castillo, 2011 and Albarrán et al., 2011b, 2011c.) In this context, we should welcome the recent work by Archambault, Beauchesne, and Caruso (2011), in which individual journals are assigned to single, mutually exclusive categories using a hybrid approach that combines algorithmic methods and expert judgment. Nevertheless, in our view, it would be important to verify whether citation distributions at every aggregation level in the new classification system satisfy the broad features that in both Albarrán et al. (2011a) and this article have been seen to characterize distributions under the multiplicative and the fractional strategies.

Acknowledgments

We acknowledge financial support by Santander Universities Global Division of Banco Santander. Ruiz-Castillo acknowledges financial support from the Spanish MEC through grant SEJ2007-67436. This article is produced as part of the CEPR project "SCience, Innovation, FIrms and markets in a GLObalized World (SCI-FI GLOW)" funded by the European Commission under its Seventh Framework Programme for Research (Collaborative Project) Contract no. 217436. Conversations with Pedro Albarrán, Félix de Moya, Vicente Guerrero, Nees Jan van Eck and, above all, Ludo Waltman are deeply appreciated. Comments from two referees helped to improve the original version of the article. All remaining shortcomings are the authors' sole responsibility.

References

Albarrán, P., Crespo, J., Ortuño, I., & Ruiz-Castillo, J. (2011a). The skewness of science in 219 subfields and a number of aggregates. Scientometrics, 88(2), 385–397.

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011b). The measurement of low- and high-impact in citation distributions: Technical results. Journal of Informetrics, 5(1), 48–63.

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011c). High- and low-impact citation measures: Empirical applications. Journal of Informetrics, 5, 122–145.

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011d). Average-based versus high- and low-impact indicators for the evaluation of citation distributions. Research Evaluation, 20, 325–339.

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. Journal of the American Society for Information Science and Technology, 62(1), 40–49.

Archambault, É., Beauchesne, O., & Caruso, J. (2011). Towards a multilingual, comprehensive, and open scientific journal ontology. Paper presented at the 13th International Conference on Scientometrics and Informetrics, Durban, Republic of South-Africa.

Glänzel, W. (2010). The application of characteristics scores and scales to the evaluation and ranking of scientific journals. Journal of Information Science, 37(1), 40–48.

Herranz, N., & Ruiz-Castillo, J. (2011a). Multiplicative and fractional strategies when journals are assigned to several subfields (Working Paper 11–20). Madrid, Spain: Universidad Carlos III.

Herranz, N., & Ruiz-Castillo, J. (2011b). The end of the European paradox. (Working Paper 11–27). Madrid, Spain: Universidad Carlos III. Availableat: http://hdl.handle.net/10016/12195

Herranz, N., & Ruiz-Castillo, J. (2012). Subfield normalization procedures in the multiplicative case: Average-based citation indicators. Journal of Informetrics, 6, 543–556.

Herranz, N., & Ruiz-Castillo, J. (in press). Subfield normalization procedures in the multiplicative case: High- and low-impact citation indicators. Research Evaluation. doi: 10.1093/reserval/rv5006.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. PNAS, 105, 17268–17272.

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

Seglen, P. (1992). The skewness of science. Journal of the American Society for Information Science, 43(9), 628–638.

Waltman, L, van Eck, N.J., van Leeuwen, T.N. Visser, M.S., & van Raan, A.F.J. (2011). Towards a new crown indicator: Some theoretical considerations. Journal of Informetrics, 5, 37–47.

Waltman, L, van Eck, N.J., & van Raan, A.F.J. (2012). Universality of citation distributions revisited. Journal of the American Society for Information Science and Technology, 63, 72–73.