Average-based indicators at the sub-field and other aggregation levels

Neus Herranz[a] and Javier Ruiz-Castillo[b]

[a] Department of Economics, University of Illinois at Urbana- Champaign,

neushg.correo@gmail.com

[b] Departamento de Economía, Universidad Carlos III, and Research Associate of the CEPR Project SCIFI-GLOW <jrc@eco.uc3m.es>

Abstract

This paper investigates the citation impact of three large geographical areas –the U.S., the European Union (EU), and the rest of the world (RW)– at different aggregation levels in two scenarios: (i) when all articles published in 1998-2002 are assigned to a single broad field, and (ii) when it is recognized that more than half of these articles are assigned to several sub-fields among a set of 219 Web of Science categories. The study focuses on the consequences of using the crown indicator, the Mean Normalized Citation Score (MNCS), and a third average-based indicator that does not correct for differences across sub-fields. It is found that: (1) Using either of the two normalized indicators, and following a multiplicative or fractional strategy at the sub-field level in scenario (ii) is of little consequence. (2) The U.S. outperforms the EU in 179 of the 219 sub-fields in scenario (ii), and at all aggregate levels in both scenarios. (3) The US/EU and RW/EU gaps for all sciences as a whole according to both normalized indicators are 25-28% and 23%, respectively. Without normalization, these gaps increase by 4.5 and 5.6 percentage points –a non-negligible difference.

INTRODUCTION

From the methodological point of view, this paper contributes to recent discussions on aggregate issues and average-based indicators. The problem, of course, is that for citation-based research performance evaluations of research units working in broad, aggregate scientific *fields*, it is crucial that one carefully controls for wide differences in the average number of citations per publication at the lowest level of aggregation in what we call *sub-fields*. As it is well known, there are two main normalization mechanisms in contention: the *crown indicator*, previously recommended by the Center for Science and Technological Studies (CWTS) at Leiden University (De Bruin *et al.*, 1993, and Moed *et al.*, 1995), and an alternative normalization mechanism sometimes referred to as the *item-oriented field-normalized citation score average* (Lunberg, 2007), or as the *mean normalized citation score*. Apparently, because of its better theoretical properties the CWTS is currently moving towards a new crown indicator that relies on the second mechanism (see Waltman *et al.*, 2011 for a clear rendition of the issues, as well as for relevant references about the polemic involving the two mechanisms).

In this scenario, this paper emphasizes the distinction between an *ideal world* in which all papers are assigned to a single sub-field, and the *real world* in which papers are assigned to sub-fields through the journals where they have been published, and journals are often assigned to two, three, or more sub-fields.

(i). In the ideal world, both average-based indicators coincide at the sub-field level. At the aggregate level, we suggest a third indicator that serves a useful purpose in spite of the fact that, contrary to the previous ones, does not correct for differences in mean citation rates across sub-fields.

(ii). Naturally, in the real world the issues are more complex. What is to be done with the publications assigned to two or more sub-fields? Section 6 in Waltman *et al.* (2011) solves the problem by adopting a fractional strategy according to which if a paper is assigned, for example, to three sub-fields, the paper is broken down into three pieces, one third of the citations originally received are assigned to each piece, and the result is assigned to each of the three sub-fields in question. We partially disagree. At the sub-field level we recommend a multiplicative strategy according to which if, as in the previous example, a paper is assigned to three sub-fields, it should be independently counted three times, once in each of the sub-fields in question, without altering the original number of citations in each case. At the aggregate level, if in the previous example the three sub-fields belong to the same field, then the paper should be counted once. This means that the number of papers at the field and sub-field levels will be different, so that the link between the two levels will be broken. In this situation, we can always compute our third indicator that only uses aggregate data. However, this is accomplished at a high cost: ignoring differences across sub-fields. Alternatively, to correct for these differences in the presence of multiple assignments we agree with Waltman *et al.* (2011) that a fractional strategy is appropriate. Interestingly enough, the third indicator computed from the sub-field level through a fractional strategy coincides with the one already referred to directly computed with aggregate data. Thus, in the real world we recommend a multiplicative strategy at the sub-field level and a fractional strategy at any aggregate level.

So far it has been implicitly assumed that each paper has been written by one or more authors belonging to the same research unit. Empirically, this will always be the case only when research units are journals. On the contrary, in an international context and many other situations it is likely that there is some cooperation between research units. Although this old problem admits different solutions (see *inter alia* Anderson *et al.*, 1988, for a discussion in an international context), we side with many other authors in recommending a multiplicative strategy at all aggregation levels that is analogous to the one already recommended at the sub-field level for the treatment of multiple assignment publications (see the influential contributions by May, 1997, and King, 2004, as well as the references in Section II in Albarrán *et al.*, 2010b).

Quite apart from the a priori advantages that may make a normalization procedure preferable to another one, it is important to empirically verify the order of magnitude of the differences that the alternative methods may bring. For that purpose, this paper uses a large data set acquired from Thomson Scientific consisting of more than 3,6 million articles published in 1998-2002, and the 28 million citations they have received during a five year citation window. Thomson Scientific assigns each article to a single broad field among a set of 20 natural sciences and two social sciences. This provides a unique opportunity to compare the three average-based indicators in a version of what we call the ideal world where the lowest aggregation level –the sub-fields– are identified with these 22 fields. The comparison is done at two aggregate levels: (i) four grand fields including the Life Sciences, the Physical Sciences, Other Natural Sciences, and the Social Sciences, and (ii) all 22 fields as a whole. On the other hand, Thomson Scientific assigns each article to one, two, three, and up to six Web of Science (WoS hereafter) categories. Therefore, in what we call the real world sub-fields are identified with 219 WoS categories. Here we compare the three indicators at three aggregate levels: 20 broad fields, the four grand fields, and all sciences as a whole.

From the empirical point of view, this paper complements previous contributions that study aggregation issues for different types of research units: individual scientists in Van Raan *et al.* (2010), as well as 158 research groups, 365 universities, 58 countries, and 8,423 journals for only seven WoS categories in Waltman *et al.* (2010). For the 22 and 219 sub-fields in our two situations, we partition the entire world into three large geographical areas: the U.S., the EU, namely, the 15 countries forming the European Union before the 2004 accession, and any other country of the rest of the world (RW hereafter).

THE AGGREGATION PROBLEM

Notation

Assume that there are S sub-fields, indexed by $s = 1,\ldots, S$, as well as K research units indexed by $k = 1,\ldots, K$. Assume that each article is written by one or more authors belonging to a single research unit, and that it is assigned to a single sub-field. Research unit k has a set $\boldsymbol{x}^k_s = \{c^k_{si}\}$ of n^k_s distinct articles in sub-field s, indexed by $i = 1,\ldots, n^k_s$, where c^k_{si} is the number of citations received by article i. The total number of articles for research unit k is $n^k = \Sigma_s\, n^k_s$. The mean citation rate (MCR hereafter) of unit k at the sub-field level s, m^k_s, is defined as

$$m^k_s = \Sigma_i\, c^k_{si}/n^k_s.$$

The corresponding MCR at the aggregate level, m^k, is defined as

$$m^k = \Sigma_s\, \Sigma_i\, c^k_{si}/n^k.$$

Note that this MCR can also be written as the weighted average of the m^k_s with weights α^k_s equal to the area's publication effort in each sub-field s:

$$m^k = \Sigma_s\, \alpha^k_s\, m^k_s,$$

where $\alpha^k_s = n^k_s/n^k$, and $\Sigma_s\, \alpha^k_s = 1$.

We are aware that most empirical work in bibliometrics is typically concerned with a relatively small subset of research units that do not constitute a partition of the world. This is the case, for example, of the empirical studies on aggregation issues mentioned in the Introduction. Quite independently of the fact that the empirical part of this paper does involve a world partition into three geographical areas, we proceed to develop the general case in order to clearly establish the inter-relationship between what happens in all members of a world partition and what happens in the world as a whole. Even if one is solely interested in investigating a certain subset of research units, normalization procedures should be defined – and understood– at the general level. Then, under the assumption that the K research units form a partition of the world, and that there is no article written in cooperation between research units, the total number of articles in sub-field s and at the aggregate world level are, respectively, $n_s = \Sigma_k\, n^k_s$ and $n = \Sigma_s\, n_s = \Sigma_k\, n^k$. What is the connection between the research units MCRs and the corresponding measures at the world level? In the first place, the world sub-field s MCR, m_s, is defined as

$$m_s = (\Sigma_k\, \Sigma_i\, c^k_{si})/n_s = \Sigma_k\, w^k_s\, m^k_s,$$

where $w^k_s = n^k_s/n_s$ is the publication share of research unit k in sub-field s, and $\Sigma_k\, w^k_s = 1$. In the second place, the world aggregate MCR, m, is defined as

$$m = (\Sigma_k\, \Sigma_s\, \Sigma_i\, c^k_{si})/n = \Sigma_s\, \alpha_s\, m_s = \Sigma_k\, W^k\, m^k = \Sigma_s\, \Sigma_k\, z^k_s\, m^k_s,$$

where $\alpha_s = n_s/n$, $W^k = n^k/n$, $z^k_s = n^k_s/n$, and $\Sigma_s\, \alpha_s = \Sigma_k\, W^k = \Sigma_s\, \Sigma_k\, z^k_s = 1$.

Average-based Indicators at the Aggregate Level

Let e^k_{si}, $i = 1,\ldots, n^k_s$, denote the expected number of citations of article i published by research unit k in sub-field s. Consider the following two well known indicators. Firstly, the so-called *crown indicator* for research unit k, C^k, is the ratio CPP/FCSm where CPP and FCSm stand for, respectively, the research unit's MCR and the mean sub-field citation score. Therefore, C^k is defined as

$$C^k = \mathrm{CPP}^k/\mathrm{FCSm}^k = (\Sigma_s\, \Sigma_i\, c^k_{si}/n^k)/(\,\Sigma_s\, \Sigma_i\, e^k_{si}/n^k) = (\Sigma_s\, \Sigma_i\, c^k_{si})/(\,\Sigma_s\, \Sigma_i\, e^k_{si}),$$

which is equation (1) in Waltman *et al.* (2010). The rationale is that the articles of a research unit are seen as a single integrated *ouvre* rather than as a number of independent works. Since the distribution of citations over the individual articles is not considered important, normalization is performed at the level of the research unit's *ouvre* as a whole rather than at the level of the research unit's individual publications. Secondly, the *mean normalized citation score, $MNCS^k$*, is defined as

$$MNCS^k = (1/n^k)\,(\,\Sigma_s\, \Sigma_i\, c^k_{si}/e^k_{si}),$$

which is equation (2) in Waltman *et al.* (2010). The M^k indicator first performs a normalization at the level of individual articles, and then obtains the average of the normalized articles. The idea is that once the number of citations received by an article has been normalized for differences among sub-fields, all articles should be treated equally.

It is natural to take e^k_{si} equal to the sub-field's MCR, μ_s, for all i in s. In this case

$$C^k = (\Sigma_s\, \Sigma_i\, c^k_{si})/(\,\Sigma_s\, \Sigma_i\, m_s) = (\Sigma_s\, \Sigma_i\, c^k_{si})/\Sigma_s\, n^k_s\, m_s = (\Sigma_s\, \Sigma_i\, c^k_{si})/n^k)/(\Sigma_s\, (n^k_s/n^k)\, m_s = m^k/m^{k\#}, \quad (1)$$

where $m^{k\#} = \Sigma_s\, \alpha^k_s\, m_s$ is the field MCR that unit k would obtain if each of its publications in a given sub-field s were to receive m_s citations, that is, if $c^k_{si} = m_s$ for all $i = 1,\ldots, n^k_s$, and $s = 1,\ldots, S$. Similarly, we have

$$MNCS^k = (1/n^k)\,(\Sigma_s\, \Sigma_i\, (c^k_{is}/m_s)). \quad (2)$$

It is easy to observe that, for any s, the sub-field indicators C^k_s and M^k_s are equal:

$$C^k_s = MNCS^k_s = m^k_s/m_s. \quad (3)$$

The relationship between the indicators at the sub-field and field levels is the following:

$$MNCS^k = \Sigma_s\, \alpha^k_s\, M^k_s,\ \ \Sigma_s\, \alpha^k_s = 1.$$

$$C^k = \Sigma_s\, \beta^k_s\, M^k_s,\ \beta^k_s = (\alpha^k_s\, m_s)/\Sigma_s\, \alpha^k_s\, m_s,\ \ \Sigma_s\, \beta^k_s = 1. \quad (4)$$

A New Type of Indicator at the Aggregate Level

It might be argued that it is not obvious why we should evaluate a research unit's *ouvre* independently of the differences between its publication effort across sub-fields, α^k_s, and the world publication effort, W_s. This is exactly what is done in C^k, where the normalization process is tailored to the research unit publication effort, as well as in M^k, where all sub-fields count the same regardless of their relative importance at the world level. Alternatively, we can take $e^k_{si} = \mu$ for all $i = 1,\ldots, n^k_s$, and all $s = 1,\ldots, S$, normalize each article so that c^k_{si}/μ, and find the MCR over all articles published by unit k. In this case we have a new indicator, I^k, defined as

$$I^k = (1/n^k)\ (\Sigma_s\ \Sigma_i\ (c^k_{si}/m)) = m^k/m. \tag{5}$$

This indicator can also be seen as the result of a normalization at the level of the research unit's *ouvre* as a whole, where the expected number of citations of the *ouvre* is taken to be the aggregate MCR, m, in which case

$$I^k = (\Sigma_s\ \Sigma_i\ c^k_{si})/(\Sigma_s\ \Sigma_i\ m) = m^k/m.$$

Finally, note that if we were to take the entire field as a homogeneous field, then

$$I^k = MNCS^k = C^k = m^k/m.$$

This provides a third interpretation of indicator I^k: it is the natural indicator to take when the entire field itself is taken to be homogeneous. However, this shows that I^k does not correct for differences across sub-fields.

Remark. Observe that if we define $C^{k\#} = m^{k\#}/m$, then we have that

$$C^k\ C^{k\#} = I^k.$$

Therefore, I^k penalizes (rewards) a research unit when $C^{k\#} < 1$ ($C^{k\#} > 1$), that is, when the research unit publication share α^k_s is smaller than the world publication share W_s for sub-fields with high m_s. Another interpretation is that I^k can be broken down in a useful way in two components, C^k and $C^{k\#}$. For any pair k and v of research units we have

$$I^k/I^v = m^k/m^v = (C^k/C^v)(m^{k\#}/m^{v\#}).$$

Finally, we could define the following indicators:

$$I_s = (1/n_s)\ (\Sigma_k\ \Sigma_i\ (c^k_{si}/m)) = m_s/m,$$

$$I = \Sigma_s\ W_s\ I_s = \Sigma_k\ W^k\ I^k = (1/n)\ (\Sigma_k\ \Sigma_s\ \Sigma_i\ (c^k_{si}/m)) = m/m = 1.$$

II.4. A Comparison of the Alternatives In the Ideal World

Clearly, sub-fields with a high m_s have more weight in the calculation of C^k (see equation 4). In view of the fact that $I^k = C^k\ C^{k\#}$, this criticism can be equally raised for I^k. According to the critics of the original crown indicator, there is no reason to treat normalized articles from different sub-fields differently. After normalization, articles from different sub-fields should be treated equally, which is what $MNCS^k$ does. Waltman *et al.* (2010) are of the same opinion.

Moreover, they establish that $MNCS^k$ is the only indicator that has both the properties of homogeneous normalization and consistency. Consequently, as indicated in the Introduction the CWTS is currently moving towards a new crown indicator that relies on the second normalization mechanism.

On the other hand, disregarding differences in publication efforts provides incentives to specialize in sub-fields with a small m_S. Indeed, one way to rise $MNCS^k$ is to achieve a few high c^k_{si}/m_S ratios. The greater is the publication effort in that sub-field, α^k_s, the greater will be the impact of this strategy on $MNCS^k$. In so far as high c^k_{si}/m_S ratios might lead to a high m^k_s, with $m^k_s > m_S$, C^k will also be high. At the same time, $m^k_s > m_S$ implies that $C^{k\#} = m^{k\#}/m < 1$, so that I^k will be low and immune to this strategy.

Of course, if m_S is close to m for every s, then the three indicators C^k, $MNCS^k$, and I^k would be close to each other. Otherwise, any attempt to aggregate sub-fields with rather different m_S is bound to be problematic. In the absence of a clear advantage of one indicator relative to the other two, it seems reasonable to study the robustness of any evaluation exercise to the different indicators.

II.5. Average-based Indicators When There Is Cooperation Between Research Units

Let us take a step towards the real world introducing the possibility that some publications are written by authors from two research units. For example, consider our own case in which the world is partitioned into three geographical areas. The assignment of internationally co-authored papers among areas is problematic. From a U.S. geopolitical point of view, for example, we want to give as much weight to an article written in a U.S. research center as we give to another co-authored by researchers from a U.S. and a European university. Thus, as indicated in the Introduction, in this paper in every internationally co-authored article a whole count is credited to each contributing area. Only domestic articles, or articles exclusively authored by one or more scientists affiliated to research centers either in the U.S., the EU, or the RW alone, are counted once. In this way, the space of articles is expanded as much as necessary beyond the initial size in what we call the *geographical extended count*.

Let us denote by N^k_s, $N^k = \Sigma_s N^k_s$, $N_S = \Sigma_k N^k_s$, and $N = \Sigma_s N_S = \Sigma_k N^k$ the number of publications of unit k in sub-field s, the total number of publications of unit k, the total number of publications in sub-field s, and the total number of publications at the aggregate level, respectively, in the geographical extended count. Of course, $N^k_s \geq n^k_s$ for all $k = 1,\ldots, K$, and $s = 1,\ldots, S$, with some strict inequality for some k and s, so that $N_S \geq n_S$ with strict inequality for some s, and $N > n$. As long as internationally co-authored publications are typically found in the upper tail of citation distributions, we expect that the MCRs in the geographically extended count, M, M_S, M^k, and M^k_s, are greater than in the ideal world, that is, $M > m$; $M_S > m_S$ for some s; $M^k > m^k$ for some k, and $M^k_s > m^k_s$ for some k and s.

Our indicators at the sub-field level are defined as before:

$$C^k_s(\text{GEC}) = MNCS^k_s(\text{GEC}) = M^k_s/M_S. \tag{6}$$

On the other hand, denote the publication effort by $a^k_s = N^k_s/N^k$, and the publication share $b^k_s = (a^k_s M_S)/\Sigma_s a^k_s M_S$ with $\Sigma_s a^k_s = \Sigma_s b^k_s = 1$. At the aggregate level we have:

$$MNCS^{k}(\text{GEC}) = (1/N^{k}) (\Sigma_{s} \Sigma_{i} (c^{k}_{is}/M_{s})) = \Sigma_{s} a^{k}_{s} MNCS^{k}_{s}(\text{GEC}), \quad (7)$$

$$C^{k}(\text{GEC}) = (\Sigma_{s} \Sigma_{i} c^{k}_{si})/(\Sigma_{s} \Sigma_{i} M_{s}) = \Sigma_{s} b^{k}_{s} C^{k}_{s} = M^{k}/M^{k\#}, M^{k\#} = \Sigma_{s} a^{k}_{s} M_{s}, \quad (8)$$

$$I^{k}(\text{GEC}) = M^{k}/M, \quad (9)$$

Finally, consider at an intermediate level what we call *fields*, indexed by $f = 1,\ldots, F$. If we let $N^{k}_{f} = \Sigma_{s\in f} N^{k}_{s}$ be the total number of articles in field f, we have:

$$MNCS^{k}_{f}(\text{GEC}) = (1/N^{k}_{f}) (\Sigma_{s\in f} \Sigma_{i} (c^{k}_{is}/M_{s})) = \Sigma_{s\in f} (N^{k}_{s}/N^{k}_{f}) MNCS^{k}_{s}(\text{GEC}), \quad (10)$$

$$C^{k}_{f}(\text{GEC}) = (\Sigma_{s\in f} \Sigma_{i} c^{k}_{si})/(\Sigma_{s\in f} \Sigma_{i} M_{s}) = M^{k}_{f}/M^{k\#}_{f}, \quad (11)$$

where $M^{k\#}_{f} = \Sigma_{s\in f} (N^{k}_{s}/N^{k}_{f}) M_{s}$, and

$$I^{k}_{f}(\text{GEC}) = M^{k}_{f}/M_{f}. \quad (12)$$

II.6. Average-based Indicators In the Real World: Multiplicative Strategy

Next, let us recognize that in the real world some publications are assigned to several sub-fields. At the sub-field level we recommend to take a multiplicative strategy for two reasons. Firstly, as indicated in Albarrán *et al.* (2010b) where sub-fields are identified with WoS categories, "*A crucial requirement is that all articles within a sub-field should count the same. Otherwise, if an article assigned to several WoS categories were fractionally assigned to them, then its place in the various citation distributions would be dramatically affected. In particular, fractionally assigned articles would have a much smaller chance of occupying the upper tail of citation distributions than articles assigned to a single WoS category.*" Secondly, a fractional assignment would affect research units' citation distributions in a normatively undesirable manner, hurting relatively more those units with overlapping articles that receive a larger number of citations. Therefore, we opt for classifying multiple assignment articles into as many WoS categories as necessary. An article assigned to three WoS categories, for instance, is classified into the three corresponding sub-fields; this means that this article –and the citations it originally received– would be counted three times. In this way, as before, the space of articles is expanded as much as necessary beyond the initial size. This artificially large number in what we call the *double extended count* is not that worrisome in the sense that, since the multiplicative strategy does not create any interdependencies among the sub-fields involved, it is still possible to separately investigate every sub-field in isolation, independently of what takes place in any other sub-field.

Let us denote by P^{k}_{s}, $P^{k} = \Sigma_{s} P^{k}_{s}$, and $P_{s} = \Sigma_{k} P^{k}_{s}$, the number of publications of unit k in sub-field s, the total number of publications of unit k, the total number of publications in sub-field s, respectively, in the double extended count at the sub-field level. Of course, the total number of articles in this case, $P = \Sigma_{s} P_{s} = \Sigma_{k} P^{k}$, is strictly greater than the total in the geographical extended count, N. Let us denote the MCRs in the double extended count by μ^k_s, μ_s, μ^k, and μ. Our indicators at the sub-field level are defined as before:

$$C^k_s(\text{DEC}) = MNCS^k_s(\text{DEC}) = \mu^k_s/\mu_s, \tag{13}$$

A key question is that as soon as we attempt to aggregate from the sub-field to the field and further aggregate levels the multiplicative strategy runs into difficulties. At the field level we do not want to count twice the papers that were assigned to two sub-fields when both sub-fields belong to the same field. If we were to do this, we would arbitrarily favor research units whose multiple assigned papers are highly cited. One alternative is to count only once such papers at the field level, maintaining a multiplicative strategy whenever a paper is assigned to two or more sub-fields and these sub-fields belong to two (or more fields). Denote by Q^k_f the number of papers of unit k in field f, $f = 1,\ldots, F$, and let $Q^k = \Sigma_f\, Q^k_f$, and $Q_f = \Sigma_k\, Q^k_f$ be the total number of papers in unit k and field f, respectively. Clearly, the total number of papers in this new double extended count, $Q = \Sigma_k\, Q^k = \Sigma_f\, Q_f$, will be smaller than P, but still greater than N. Denote the MCRs in this extended count by μ_f^{k*}, μ^{k*}, μ_f^{*}, and μ^{*}. The connection between the field and sub-field levels is broken in the sense that

$$\mu^{k*} \neq \mu^k = \Sigma_s\, (P^k_s/P^k)\, \mu^k_s,$$

and

$$\mu^{*} \neq \mu = \Sigma_s\, (P_s/P)\, \mu_s.$$

However, for each k we can always compute the field indicators

$$I_f^k = \mu_f^{k*}/\mu_f^{*}, f = 1,\ldots F. \tag{14}$$

At the maximum aggregate level all papers would count only once, regardless of whether some of them were originally assigned to two or more sub-fields. Consequently, the MCRs at this level for all units and for the world as a whole correspond to those at the geographical extended count. Thus, we would have available again the indicator defined in equation 9, $I^k(\text{GEC}) = M^k/M$. Note, however, that using the means μ^k, and μ computed in the first double extended count, as well as μ^{k*} and μ^{*} computed in the second one, two systems of aggregate indicators of this type can be defined for each k:

$$I^k(\text{DEC}) = \mu^k/\mu, \tag{15}$$

$$I^k(\text{DEC2}) = \mu^{k*}/\mu^{*}. \tag{16}$$

II.7. Average-based Indicators In the Real World: Fractional Strategy

The problem with the previous strategy, of course, is that it amounts to taking fields and all sciences as a whole at the maximum aggregate level as if they were homogeneous, something we know it is not the case. Moreover, moving away from the double extended count at the sub-field level means that we are no longer able to compute MCRs at that level. This implies renouncing to compute the sub-field indicators $C^k{}_s$ and $MNCS^k{}_s$ and, hence, aggregate indicators $C^k{}_f$, C^k, as well as $MNCS^k{}_f$ and $MNCS^k$. An alternative is to pursue the following fractional strategy.

Consider a paper i, written in research unit k, receiving $c^k{}_i$ citations, which had been assigned to, say x sub-fields indexed by $s = j + 1, \ldots, j + x$. This paper is made equivalent to $(1/x)$-th of a paper with $c^{k'}{}_{si} = (c^k{}_i/x)$ citations in each $s = j + 1, \ldots, j + x$. Let $N^{k'}{}_s$ be the number of papers of unit k in sub-field s according to the fractional strategy, and let $N^{k'} = \Sigma_s N^{k'}{}_s$. For each unit, counting only once a paper assigned to x sub-fields is equivalent to the sum of $(1/x)$ fractional papers x times. Consequently, $N^{k'} = N^k$ for all k, so that the total number of papers in the fractional strategy $N' = \Sigma_k N^{k'}$ coincides with N, the total number of papers in the geographical extended count. Let us denote by μ', μ'_s, $\mu^{k'}$, and $\mu^{k'}{}_s$ the MCRs computed following the fractional strategy, where

$$\mu^{k'}{}_s = \Sigma_i c^{k'}{}_{si}/N^{k'}{}_s, \tag{17}$$

$$\mu^{k'} = \Sigma_s a^{k'}{}_s \mu^{k'}{}_s,$$

$a^{k'}{}_s = N^{k'}{}_s/N^{k'}$, and $\Sigma_s a^{k'}{}_s = 1$,

$$\mu'_s = (\Sigma_k \Sigma_i c^{k'}{}_{si})/N'_s = \Sigma_k w^{k'}{}_s \mu^{k'}{}_s,$$

$N'_s = \Sigma_k N^{k'}{}_s$, $w^{k'}{}_s = N^{k'}{}_s/N'_s$, and $\Sigma_k w^k{}_s = 1$, and

$$\mu' = (\Sigma_k \Sigma_s \Sigma_i c^{k'}{}_{si})/N',$$

where $N^{k'} = N^k$ for all k, and $N' = N$.

At the sub-field level we can now compute the indicators

$$C^{k'}{}_s = MNCS^{k'}{}_s = \mu^{k'}{}_s/\mu'_s. \tag{18}$$

At the intermediate level we have:

$$MNCS^{k'}{}_f = (1/N^{k'}{}_f)\,(\Sigma_{s\in f} \Sigma_i\,(c^{k'}{}_{is}/\mu'_s)) = \Sigma_{s\in f}(N^{k'}{}_s/N^{k'}{}_f)\, MNCS^{k'}{}_s, \tag{19}$$

$$C^{k'}{}_f = (\Sigma_{s\in f} \Sigma_i c^{k'}{}_{si})/(\Sigma_{s\in f} N^{k'}{}_s \mu'_s) = \mu^{k'}{}_f/\mu^{k\#'}{}_f, \tag{20}$$

where $N^{k'}{}_f = \Sigma_{s\in f} N^{k'}{}_s$, and $\mu^{k\#'}{}_f = \Sigma_{s\in f}(N^{k'}{}_s/N^{k'}{}_f)\,\mu'_s$. Finally, at the aggregate level we have

$$MNCS^{k'} = (1/N^k)\,(\Sigma_s \Sigma_i\,(c^{k'}{}_{is}/\mu'_s)) = \Sigma_s a^{k'}{}_s MNCS^{k'}{}_s, \tag{21}$$

$$C^{k'} = (\Sigma_s \Sigma_i c^{k'}{}_{si})/(\Sigma_s \Sigma_i \mu'_s) = \Sigma_s b^{k'}{}_s C^{k'}{}_s = \mu^{k'}/\mu^{k\#'}, \tag{22}$$

where $\mu^{k\#'} = \Sigma_s a^{k'}{}_s \mu'_s$, $b^{k'}{}_s = (a^{k'}{}_s \mu'_s)/\Sigma_s a^{k'}{}_s \mu'_s$, $\Sigma_s b^{k'}{}_s = 1$.

SUMMARY OF EMPIRICAL RESULTS

(To save space, all numerical information is available in the Working Paper version of this paper)

The Data

In this paper, only research articles or, simply, articles are studied. We begin with a large sample acquired from Thomson Scientific, consisting of about 3.6 million articles published in 1998-2002, as well as more than 28 million citations these fields receive using a five-year citation window for each one. Articles are assigned to geographical areas according to the institutional affiliation of their authors as recorded in the Thomson Scientific database on the basis of what had been indicated in the by-line of the publications. As indicated in the Introduction, in this paper in every internationally co-authored article a whole count is credited to each contributing area. In this way we arrive to the geographical extended count of 4.1 million articles, 13.5% larger than in the ideal world. However, it can be shown that working with the geographical extended count does not distort much the features that characterize the ideal world.

As indicated in the Introduction, in our version of the real world sub-fields are identified with the 219 WoS categories distinguished by Thomson Scientific in our original dataset. In the original dataset the multiple assigned articles represent 58,8% of the total. Following a multiplicative strategy, we arrive to a double extended count at this level of more than 6,5 million articles and 38 million citations. These amounts represent, respectively, 57.7% and 53% more than the number of articles and citations in the geographical extended count used in the previous Sub-section.

There is no generally agreed upon Map of Science or aggregation scheme that allows us to climb from the sub-field up to other aggregate levels. Among the many alternatives, borrowing from the schemes recommended by Tijssen and van Leeuwen (2003) and Glänzel and Schubert (2003), Albarrán *et al.* (2010b) constructed a third scheme consisting of 80 intermediate categories, called disciplines, and 19 fields with the aim of maximizing the possibility that a power law represents the upper tail of each of the corresponding citation distributions. For our purposes, we separate Computer Sciences from Engineering to work with a total of 20 fields. Finally, we also distinguish the four grand fields previously considered.

Conclusions

From a methodological point of view, the main conclusions are the following four.

1. Using the crown or the *MNCS* indicator gives rise in both scenarios to some differences in the measurement of citation impact of all areas, as well as in the measurement of the US/EU and the RW/EU gaps. However, these differences are of a small order of magnitude generally below a few percentage points.

2. Since the publication shares of both the U.S. and the EU are rather close to the world ones, there is not much of a difference between the crown and the *I* indicator in spite of the fact that the latter does not account for differences in MCRs across sub-fields. In both scenarios there are many cases in which, for each geographical area, the crown indicator is larger. But the opposite is also the case in many other cases at all aggregation levels. However, these differences have small consequences for the US/EU and RW/EU gaps.

3. The results about the impact of normalization on the size of the two gaps are interesting. In the second with 219 sub-fields, the US/EU gap at the field and grand field levels is greater in half of the cases when the *MNCS* indicator is used, while the RW/EU gap is most of the time smaller when this normalized indicator is used. For all sciences as a whole, when we normalize the US/EU and the RW/EU gap decrease by 4.5 and 5.6 percentage points –a non-negligible difference.

4. At the sub-field level both normalized indicators coincide. The issue is about the consequences of solving the multiple assignments of articles to sub-fields by means of a multiplicative or a fractional strategy in the second scenario. We recommend the first, in which case the dominance of the U.S. over the EU reaches about 80% of the 219 cases. When a fractional strategy is followed, the US/EU gap increases in 137 cases, or 63% of the total. However, this gap increases by more than 10% in only 20 cases (of which 17 represent a worsening of the European position).

From a substantive point of view, the main conclusions are the following two.

5. As soon as we climb from the sub-filed to higher aggregate levels in the second scenario, all indicators agree that the few cases in which the EU maintains some dominance over the U.S. yield to the ranking that also obtains at all levels in the first scenario and previous research: the U.S., the EU, and the RW.

6. From a quantitative point of view at the field level in the second scenario, the two normalized indicators agree that the US/EU gap is less than 20% in six fields, and between 20% and 40% for the remaining 14, while the RW/EU gap is between 10% and 20% in seven cases, and between 20% and 50% in the remaining 13. For all sciences as a whole, the normalized US/EU and RW/EU gaps are 25-28% and 23%, respectively.

Extensions

As it is well known, references made by articles in any sub-field give rise to a highly skewed distribution of citations received, in which a large proportion of articles gets none or few citations while a small percentage of them account 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 the lower part are typically very different. This leads to the idea of using two indicators to describe any citation distribution: a pair of a high- and a low-impact measure defined over the set of articles with citations below or above a critical citation level (see Albarrán *et al.*, 2011c, for a discussion of technical properties). While average-based measures are silent about the distributive characteristics on either side of the mean, the high- and low-impact measures used for the evaluation of the U.S, the EU, and the RW in Albarrán *et al.* (2011a) are sensitive to the citation inequality in the sense that an increase in the coefficient of variation increases both of them.

Previous results –Albarrán *et al.* (2011a, b) – are restricted to the scenario in which articles are assigned to only one of the 22 broad fields distinguished by Thomson Scientific. It remains to investigate how to apply this approach in what we have called the real world. Naturally, this extension should include aggregation procedures capable of correcting for differences in citation practices across sub-fields. This is the subject matter of ongoing research.

Acknowledgements

The authors acknowledge financial support from the Santander Universities Global Division of *Banco Santander*. Ruiz-Castillo also acknowledges financial help from the Spanish MEC through grant SEJ2007-67436. This paper is part of the SCIFI-GLOW Collaborative Project supported by the European Commission's Seventh Research Framework Programme, Contract number SSH7-CT-2008-217436. Conversations with Pedro Albarrán, Félix de Moya, Vicente Guerrero, Ludo Waltman, and Nees Jan van Eck are deeply appreciated. All shortcomings are the authors' sole responsibility.

REFERENCES

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

Albarrán, P., J. Crespo, I. Ortuño, and J. Ruiz-Castillo (2010b), "The Skewness of Science In 219 Sub-fields and A Number of Aggregates", Working Paper 10-15, Universidad Carlos III.

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

Albarrán, P., I. Ortuño, and J. Ruiz-Castillo (2011b), "Average-based *versus* High- and Low-impact Indicators For The Evaluation of Citation Distributions With", forthcoming in *Research Evaluation*.

Albarrán, P., I. Ortuño, and J. Ruiz-Castillo (2010b). "The Measurement of Low- and High-impact In Citation Distributions: Technical Results", *Journal of Informetrics*, **5**: 48-63.

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

De Bruin, R.E., Kint, A., Luwel, M., and Moed, H. F. (1993), "A Study of Research Evaluation and Planning: The University of Ghent", *Research Evaluation*, **3**: 25-41.

EC (1994), *First European Report on Science and Technology Indicators*, Directorate-General XII, Science, Research, and Development. Luxembourg: Office for Official Publications of the European Community.

Glänzel, W. (2000), "Science in Scandinavia: A Bibliometric Approach", *Scientometrics*, **48**: 121-150.

Glänzel, W. (2001), "National Characteristics in International Scientific Co-authorship Relations", *Scientometrics*, **51**: 69-115.

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

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

Lunberg, J. (2007), "Lifting the Crown –Citation *z*-score", *Journal of Informetrics*, **1**: 145-154.

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

Moed, H. F., R. E. De Bruin, and T. van Leeuwen (1995), "New Bibliometrics Tools for the Assessment of National Research Performance: Database Description, Overview of Indicators, and First Applications", *Scientometrics*, **133**: 381-422.

Tijssen, R., and T. van Leeuwen (2003), "*Bibliomeric Analysis of World Science*", Extended Technical Annex to Chapter V of *Third European Report on Science and Technology Indicators*, Directorate-General for Research. Luxembourg: Office for Official Publications of the European Community, http://www.cordis.lu/rtd2002/indicators/home.htlm.

Van Raan, A. F. J., Van Leeuwen, T. N., Visser, M. S., Van Eck, N. J., and Waltman, L, (2010), "Rivals for the Crown: Repply to Opthof and Leydesdorff", *Journal of Informetrics*, **4**: 431-435.

Waltman, L, N. J. van Eck, T. N. van Leeuwen, M. S. Visser, and A. F. J. van Raan (2010), "Towards a New Crown Indicator: An Empirical Analysis", arXiv:1004.1632v2 [cs.DL].

Waltman, L, N. J. van Eck, T. N. van Leeuwen, M. S. Visser, and van Raan (2011), "Towards a New Crown Indicator: Some Theoretical Considerations", *Journal of Informetrics*, **5**: 37-47.