

The measurement of low- and high-impact in citation distributions: Technical results



Pedro Albarrán, Ignacio Ortuño, Javier Ruiz-Castillo *

Departamento de Economía, Universidad Carlos III, Madrid 128, Getafe 28903, Spain

A B S T R A C T

This paper introduces a novel methodology for comparing the citation distributions of research units of a certain size working in the same homogeneous field. Given a critical citation level (CCL), we suggest using two real valued indicators to describe the shape of any distribution: a high-impact and a low-impact measure defined over the set of articles with citations above or below the CCL. The key to this methodology is the identification of a citation distribution with an income distribution. Once this step is taken, it is easy to realize that the measurement of low-impact coincides with the measurement of economic poverty. In turn, it is equally natural to identify the measurement of high-impact with the measurement of a certain notion of economic affluence. On the other hand, it is seen that the ranking of citation distributions according to a family of low-impact measures is essentially characterized by a number of desirable axioms. Appropriately redefined, these same axioms lead to the selection of an equally convenient class of decomposable high-impact measures. These two families are shown to satisfy other interesting properties that make them potentially useful in empirical applications, including the comparison of research units working in different fields.

Keywords:

Research performance
Citation distribution
Poverty measurement
Impact indicators

1. Introduction

From an operational point of view, a scientific field is a collection of papers published in a set of closely related professional journals. A field is said to be *homogeneous* if the number of citations received by its papers is comparable independently of the journal where each has been published. Consequently, if one paper has twice the number of citations as another in the same homogeneous field, not only it can be said that it has twice the international impact but also that it has twice as much merit as the other. Following a common practice in bibliometrics, the scientific performance of a research unit is identified with the impact achieved by the papers the unit publishes in the serial literature, where the impact is measured by the citations the papers receive. The main aim of this paper is to introduce a novel methodology for comparing the citation distributions of research units working in the same homogeneous field. Since citation distributions of individual scientists tend to be small and, therefore, noisy from a statistical inference point of view, it should be understood at the outset that this methodology is only applicable to research units of a certain minimal size (a university department, research institute, journal, region, country, or supra-national geographic area).

* Corresponding author. Tel.: +34 91 624 95 88; fax: +34 91 624 93 29.
E-mail address: jrc@eco.uc3m.es (J. Ruiz-Castillo).

It is well known that citation distributions are highly skewed, that is, many papers receive zero or few citations, while a few articles receive a disproportionate amount of all citations.¹ Correspondingly, the upper and the lower part of any citation distribution are typically very different. An important consequence is that a single statistic of centrality, such as the mean citation rate (MCR hereafter) or the median, may not adequately capture this feature (see Bornmann, Mutz, Nehaus, & Daniel (2008) and, in a different context, Glänzel (2002). Tijssen, Visser, & van Leeuwen (2002) and Aksnes and Sivertsen (2004) also stress that average or median scores will tend to be inadequate for describing general features of scientific excellence). This is why we suggest using two indicators to describe this key aspect of citation distributions. Assume we are given a criterion for selecting a *critical citation level* (CCL hereafter). Articles in a homogeneous field with citations above or below the CCL are referred to as high- and low-impact articles. Any citation distribution in that field will be described using two real valued functions, or indicators: a *high-impact* and a *low-impact measure* defined over the corresponding subsets of high- and low-impact articles.

Economists will readily recognize that the key to this methodology is the identification of a citation distribution with an income distribution. The measurement of low-impact, which starts with the identification of low-citation papers as the ones with citations below the CCL, coincides with the measurement of economic poverty that, as originally suggested in Sen's (1976) seminal contribution, starts with the identification of the poor as those individuals whose incomes are below a certain poverty line. On the other hand, once low-impact has been identified with economic poverty, it is equally natural to identify the measurement of high-impact with the measurement of a certain notion of economic affluence.

There are three measurement issues that deserve attention. Firstly, by borrowing results from the economic literature on poverty analysis in Foster and Shorrocks (1991) – FS hereafter – it will be seen that the ranking of citation distributions induced by a family of low-impact indicators has been essentially characterized in terms of a number of interesting properties. This is the FGT family of low-impact measures, originally suggested by Foster, Greener, and Thorbecke (1984) for the measurement of economic poverty. These same properties lead to the selection of an equally convenient class of FGT high-impact measures that is the counterpart of the family just mentioned. Secondly, beyond the ranking of citation distributions one is interested in cardinal comparisons between them. Although their cardinal aspects have not been equally characterized, the properties that the FGT indicators satisfy justify their use in empirical work for both ordinal and cardinal purposes. Finally, comparisons of citation distributions must be extended from the homogeneous to the heterogeneous case. Since our methodology describes the shape of citation distributions, independently of their size and their scale, it will be seen that it permits the comparison of research units working in different scientific fields, as well as the comparison of entire heterogeneous fields.

The only information required from research units in the approach advocated in this paper is the homogeneous field to which each unit's publications belong and the number of citations they receive.² In this scenario, the MCR is a good overall indicator of scientific performance. However, as indicated above, the MCR is silent about the distribution characteristics to either side of the mean—an undesirable feature when evaluating very skewed citation distributions. Scientometrics literature offers three main ways of dealing with this situation. Firstly, recent evaluations of research units focus on the upper tail of the citation distribution. We refer to measures such as the *h-index*, first suggested by Hirsch (2005) for the evaluation of individual scientists (see the recent survey by Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009). Secondly, there are methods for evaluating citation distributions by partitioning them into subgroups of articles with different number of citations. We refer to the technique of Characteristic Scaling, or Characteristic Scores and Scales (CSS hereafter) pioneered by Schubert, Glänzel, and Braun (1987) and Glänzel and Schubert (1988a, 1988b), where a set of characteristic scores defined independently of any preconceived rule or law are used to group papers in a given subject field into several categories of citation. Thirdly, among the battery of indicators used by the Leiden group for the monitoring of research groups (see, *inter alia* Moed, Burger, Frankfort, & van Raan, 1985; Moed, De Bruin, & van Leeuwen, 1995; Moed & van Raan, 1988; Van Leeuwen, Visser, Moed, Nederhof, & van Raan, 2003; Van Raan, 2004), one could select the most appropriate according to the limited information assumed to be available in this paper. One possibility, which will be referred to as the Leiden triad, would be to complete the unit's MCR in a given field with the percentage of uncited papers, and the percentage contribution to the set to the top 5% of highly cited papers (for similar proposals, see Aksnes & Sivertsen, 2004; Tijssen et al., 2002).

The problems with these alternatives are as follows. Firstly, one important feature of the *h-index* (shared by its many variants, extensions or alternatives) is that, in the terminology of Pinski and Narin (1976), it is size dependent: the greater the number of articles published by a group of researchers, the greater the value of the *h-index* will tend to be. This, of course, precludes the direct comparison of the *h-indices* of, say, Chemistry articles published in 2000 and cited during 2000–2005 by authors working in countries like the Netherlands and the U.S. with vastly different sizes. Secondly, CSS lacks the monotonicity property according to which a new citation is always desirable (see, *inter alia*, Marchant, 2009; Quesada, 2009, 2010; Woeginger, 2008a, 2008b). Thirdly, what we call the Leiden triad is an evaluation procedure insensitive to

¹ See *inter alia* Seglen (1992), Schubert, Glänzel, and Braun (1987) for evidence concerning scientific articles published in 1981–85 in 114 sub-fields, Glänzel (2007) for articles published in 1980 in 12 broad fields and 60 sub-fields, Albarrán and Ruiz-Castillo (2009) for articles published in 1998–2002 in 22 broad fields, and Albarrán, Crespo, Ortuño, and Ruiz-Castillo (2010a) for these same articles classified in 219 sub-fields and some aggregates.

² In particular, the knowledge of the journals where the papers are published, or the number of citations received by the papers citing an article in the original distribution are not required. However, the weighting of citations by their importance as a function, for example, of where the citing paper was published, can be easily incorporated into our framework.

distributional changes that leave the mean constant. Instead, among other desirable properties, the FGT high- and low-impact measures introduced in this paper are size independent, monotone, and sensitive to distributional considerations.

The rest of this paper is organized into five Sections and an [Appendix A](#). Section 2 introduces some notations and a number of properties high- or low-impact measures should satisfy. Section 3 presents a number of basic high- or low-impact indicators, as well as the FGT families that we recommend for empirical applications. A more detailed discussion of the technical results characterizing the FGT rankings is relegated to the [Appendix A](#). Section 4 is devoted to the evaluation of research units in the heterogeneous case. Section 5 reviews the properties satisfied by the comparable procedures available in the literature, and Section 6 offers some conclusions.

2. Properties of low- and high-impact indicators

2.1. Notation and definitions

Consider discrete citation distributions represented by vectors $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$, where $x_i \geq 0$ is the number of citations the i th article receives. Denote the total number of articles, the total number of citations, and the MCR by $n(\mathbf{x})$, $X = \sum_i x_i$, and $\mu(\mathbf{x}) = X/n(\mathbf{x})$, respectively. In the sequel, a CCL will always be a positive number, say $z > 0$. For any given CCL, z , and distribution \mathbf{x} , classify as low- or high-impact articles all papers with citation $x_i < z$, or $x_i \geq z$, respectively. Once the sets of low- and high-impact articles have been identified, the next step consists of combining this data to obtain a number that indicates the overall or aggregate low- and high-impact level of a citation distribution. A *low-impact index* is a real valued function L , whose typical value $L(\mathbf{x}; z)$ indicates the low-impact level associated with the distribution \mathbf{x} and the CCL z , while a *high-impact index* is a real valued function H , whose typical value $H(\mathbf{x}; z)$ indicates the high-impact level associated with the distribution \mathbf{x} and the CCL z .

Given a citation distribution \mathbf{x} and a CCL z , we can think of the (citation) deprivation associated to any low-impact article. A convenient measure of such deprivation for any article with x_i citations is the *low-impact gap*:

$$g_i(\mathbf{x}; z) = \max \{z - x_i, 0\}. \quad (1)$$

Thus, $g_i(\mathbf{x}; z) > 0$ for low-impact articles, while $g_i(\mathbf{x}; z) = 0$ for high-impact articles. As seen below, given the value of the CCL z many low-impact indices may be defined as functions of the vector $\mathbf{g}(\mathbf{x}; z) = (g_1, \dots, g_i, \dots, g_n)$. But many of these indices depend on the CCL only through the ratio of each deprivation gap to the CCL. Therefore, it is convenient to define the *normalized deprivation gaps* as follows:

$$\Gamma_i(\mathbf{x}; z) = \frac{g_i(\mathbf{x}; z)}{z} = \max \left\{ \frac{z - x_i}{z}, 0 \right\}, \quad i = 1, \dots, n. \quad (2)$$

Similarly, we can think of the (citation) affluence associated to any high-impact article. A convenient measure for any x_i is the *high-impact gap*

$$g_i^*(\mathbf{x}; z) = \max\{(x_i - z), 0\}. \quad (3)$$

Thus, $g_i^*(\mathbf{x}; z) \geq 0$ for high-impact articles, while $g_i^*(\mathbf{x}; z) = 0$ for low-impact articles. Again, given the value of the CCL z many high-impact indices may be defined as functions of the vector $\mathbf{g}^*(\mathbf{x}; z) = (g_1^*, \dots, g_i^*, \dots, g_n^*)$. Since many of these indices depend on the CCL only through the ratio of each affluence gap to the CCL, it is convenient to define the *normalized affluence gap* by:

$$\Gamma_i^*(\mathbf{x}; z) = \frac{g_i^*(\mathbf{x}; z)}{z} = \max \left\{ \frac{x_i - z}{z}, 0 \right\}, \quad i = 1, \dots, n. \quad (4)$$

There are clearly many conceivable functions that do not correspond to anyone's notion of low- and high-citation impact, and so the first question we should ask is: what properties should these measures satisfy? For later reference, [Table 1](#) contains a list of the 12 properties – referred to as axioms – that are used in the paper. For expository purposes, it is useful to distinguish between basic and subsidiary properties. The following four sub-sections include an informal, intuitive discussion of all axioms, while the formal definitions of those necessary in a number of crucial results are relegated to [Appendix A](#). To facilitate the comparison, the presentation stays as close as possible to the economic poverty literature.

2.2. Basic properties

The first six axioms form the core of the low- and high-impact notions. The first one requires that if two citation distributions are identical except for the fact that one is a permutation of the other, then the low- and high-impact level of both distributions should be the same. For the record, we have:

A1. Symmetry: Low- and high-impact measures are invariant to permutations.

This property allows articles to be reordered without affecting the low- or high-impact values. In the sequel, any distribution \mathbf{x} will be taken to be ordered, so that $x_1 \leq x_2 \leq \dots \leq x_n$. The next axiom is a version of the size independent property discussed in [Pinski and Narin \(1976\)](#). The idea is that if a citation distribution is exactly replicated any number of times, say

Table 1
A list of axioms.

Basic axioms

- A1. *Symmetry*: Low- and high-impact measures are invariant to permutations.
 A2. *Replication invariance (size independence)*: Low- and high-impact measures are invariant to replications of the citation distribution.
 A3. *Continuity*: Low- and high-impact measures are continuous functions.
 A4. *Focus*: Low- and high-impact measures are invariant to changes in the set of high- and low-impact articles, respectively.
 A5. *Monotonicity*: A new citation of a low-impact (high-impact) article reduces (increases) the low- impact (high-impact) level of a citation distribution.
 A6. *Normalization*: A low- or a high-impact measure takes the zero value when the set of low or high-impact articles is empty, respectively.

Subsidiary axioms

- A7. *Subgroup consistency*: Consider a partition of a citation distribution into two subgroups, and assume that the low-impact (high-impact) level in one of the subgroups increases while it remains constant in the other subgroup. Then in the new situation the overall low-impact (high-impact) level increases.
 A8. *Decomposability*: For any partition of a distribution, a low- or a high-impact index is decomposable if the overall low- or high-impact level is the weighted average of the low- or high-impact levels of the subgroups, with weights equal to the subgroups' publication shares.
 A9. *Transfer axiom*: If a citation transfer takes place between two high- or low-impact articles, then the high- or low-impact level increases.
 A10. *Invariance to equal proportional variations (scale invariance)*: Low- and high-impact indicators are invariant to equal proportional changes in the citations received by all articles and the CCL.
 A11. *Invariance to equal absolute variations (translation invariance)*: Low- and high-impact indicators are invariant to equal absolute changes in the citations received by all articles and the CCL.
 A12. *Increasing CCL axiom*: If the CCL increases, then the low-impact (high-impact) level increases (decreases).
-

twice, the low- and high-impact level of the resulting distribution should be equal to that of the original one. Consider the distributions $\mathbf{x} = (0, 0, 2, 5, 8)$ and $\mathbf{y} = (0, 0, 0, 2, 2, 5, 5, 8, 8)$. If, for example, the CCL is equal to 4, then the set of low-impact articles for the two distributions are $(0, 0, 2)$ and $(0, 0, 0, 2, 2)$, while the sets of high-impact articles are $(5, 8)$ and $(5, 5, 8, 8)$ in the two cases. The citation per article has not changed, the only difference being the number of times a certain citation sequence appears in each case. Since clearly the *shape* of the distribution has remained constant, there are reasons to say that the structure of the low- and high-impact phenomena has not changed. Hence:

A2. Replication Invariance (Size Independence): Low- and high-impact measures are invariant to replications of the citation distribution.

Of course, like any axiom, this one is debatable. Some observers may say that the two distributions of the example should not be equally ranked. Insofar as the research unit with twice the number of articles can be said to be more productive, either the low-impact aspect of distribution \mathbf{y} can be said to be smaller than that of \mathbf{x} , and/or the high-impact aspect of \mathbf{y} can be said to be greater. In any case, this property is crucial, since it allows us to compare the low- and high-impact aspects of two citation distributions of different sizes. Consider two distributions \mathbf{x} and \mathbf{y} with n and m articles, respectively. They are not directly comparable, but we can always replicate the first one m times and the second one n times. Let us denote the result by \mathbf{x}' and \mathbf{y}' . Thanks to A2, the low- and high-impact levels of distribution \mathbf{x}' are equal to those of distribution \mathbf{x} , and similarly with \mathbf{y}' and \mathbf{y} . But the two new distributions now have the same size: n times m . Thus, by comparing \mathbf{x}' and \mathbf{y}' the original problem is solved.³

It is technically convenient to ensure that small changes in citation distributions generate small changes in low- or high-impact levels. This is the content of the third axiom:

A3. Continuity: Low- and high-impact measures are continuous functions.

The next two axioms are very important because they serve to differentiate low- from high-impact indicators. The first one captures the idea that low- and high-impact measurement should be completely independent of what takes place in high- and low-impact articles, respectively. Hence:

A4. Focus: Low- and high-impact measures are invariant to changes in the set of high- and low-impact articles, respectively.

Next, consider a situation in which a low-impact article receives a new citation. What should we require from a well-behaved low-impact measure? Surely that the low-impact level is reduced in the new situation. Similarly, if a high-impact article receives a new citation we expect the high-impact level to increase. Hence:

A5. Monotonicity: A new citation of a low-impact (high-impact) article reduces (increases) the low-impact (high-impact) level of a citation distribution.

As will be seen in Section 3, perhaps the most important implication of the monotonicity axiom is that it rules out measures based upon a simple counting of the low- or high-impact articles. The sixth axiom is elementary and innocuous and introduces a normalization rule according to which, when the set of low- or high-impact articles is empty, the low- or the high-impact measure takes the value zero.

³ Dasgupta, Sen, and Starrett (1973) first discussed this axiom in their seminal contribution to income inequality. Chakravarty (1983) and Thon (1983) introduced it into poverty measurement.

A6. Normalization: A low- or a high-impact measure takes the zero value when the set of low or high-impact articles is empty, respectively.

2.3. Subgroup consistency and decomposability

There are many L and H indices satisfying properties 1–6. As a matter of fact, the sheer quantity of possible indicators makes it difficult to select those to be used in practice. Therefore, it would be useful to count on other interesting axioms that help restrict the class of admissible indicators, that is, the functional form that L and H indices can take. The next axiom is a case in point.

Consider any partition of the set of articles under evaluation into a number of non-overlapping subgroups.⁴ The subgroup consistency property for high-impact measures requires the overall high-impact level to fall if a subgroup experiences a high-impact reduction, while the high-impact levels in the rest of the population subgroups remain unchanged. The subgroup consistency axiom for low-impact measures is analogously defined. This property –which has not yet been discussed in scientometrics literature– is desirable for a number of reasons. From a practical point of view, consider a situation where the object of study is the citation distribution of articles in a certain homogeneous field published by research units in a certain country. Subgroup consistency is needed to coordinate the efforts of the country’s decentralization strategy towards, say, a high-impact increase in the field in question. Such a strategy may typically involve a set of policy measures targeted at specific research units. If the high-impact indicator is not subgroup consistent we may be faced with a situation in which each targeted research unit achieves the objective of increasing its own high-impact level, and yet the country’s high-impact level decreases. Subgroup consistency may therefore be viewed as an essential counterpart to a coherent low- or high-impact policy program. This requirement may also be regarded as a natural analogue of the monotonicity condition 5, which requires that the aggregate high- (or low-) impact level increases (falls) if one *article* receives a new citation, *ceteris paribus*, while subgroup consistency demands that the aggregate high- (or low-) impact level increases (falls) if one *subgroup* sees its high-impact increased (or its low-impact reduced), *ceteris paribus*; in this sense, [Zheng \(1997\)](#) suggests that subgroup consistency can be also termed subgroup monotonicity.

A7. Subgroup consistency: Consider a partition of a citation distribution into two subgroups, and assume that the low-impact (high-impact) level in one of the subgroups increases while it remains constant in the other subgroup. Then in the new situation the overall low-impact (high-impact) level increases.

Thus, subgroup consistency merely ensures that the aggregate, or overall low- (or high-) impact value does not respond perversely to changes in the level of low- (or high-) impact within one subgroup while the level of the other stays constant. Note that a principal restriction of the conditions under which subgroup consistency applies is that subgroup sizes are fixed to rule out any changes in subgroup low- (or high-) impact due entirely to inter-groups shifts of articles. On the other hand, the constraints on the number of subgroups and the precise way the subgroups’ low- (or high-) impact levels alter are less restrictive than they appear at first glance. The number of fixed-sized groups may be extended to any number greater than two, on the condition that the overall low- (or high-) impact level increases if the low- (or high-) impact does not fall in any subgroup and increases in at least one. Finally, the subgroup consistency property is closely allied to the stronger condition of decomposability, originally defined by [Foster et al. \(1984\)](#) as follows:

A8. Decomposability: Consider any partition of a distribution \mathbf{x} into $K \geq 2$ subgroups so that $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^K)$, where \mathbf{x}^k is the citation distribution of subgroup k , and $\omega_k = n(\mathbf{x}^k)/n(\mathbf{x})$ is the proportion that the articles in subgroup k represent in the total volume of articles in the original distribution. A low- or a high-impact index is decomposable if the overall low- or high-impact level is the weighted average of the low- or high-impact levels of the subgroups, with weights equal to the subgroups’ publication shares ω_k ; that is, a low- or a high-impact index, L or H , is decomposable if the overall low-impact level can be written as

$$L(\mathbf{x}^1, \dots, \mathbf{x}^K; z) = \sum_k \omega_k L(\mathbf{x}^k; z),$$

and the high-impact level as

$$H(\mathbf{x}^1, \dots, \mathbf{x}^K; z) = \sum_k \omega_k H(\mathbf{x}^k; z).$$

Of course, decomposable measures are also subgroup consistent, but not vice-versa. Given a partition, if we are merely interested in comparing the subgroup low- or high-impact levels with one another, the decomposability requirement is quite unnecessary. On the other hand, if the analysis involves comparisons between subgroup and overall levels, then

⁴ Consider the world distribution of articles in Chemistry and its partition into the Chemistry papers attributed to the different countries in the world. Or consider a country’s citation distribution in Mathematics and its partition into the Mathematics articles attributed to the different university departments and research institutes active in that country.

decomposability can be very useful indeed. To appreciate this, consider a situation in which the citation distribution of a given scientific discipline, \mathbf{x} , is partitioned into the citation distributions of K countries, so that $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^k, \dots, \mathbf{x}^K)$. If the L and H indicators are decomposable, then given a certain CCL z , each country's observed relative contribution to the low- or high-impact measure at the worldwide level is equal to $\omega_k L(\mathbf{x}^k; z)/L(\mathbf{x}; z)$ or $\omega_k H(\mathbf{x}^k; z)/H(\mathbf{x}; z)$, respectively. Each country's expected contribution can be taken to be its publication share ω_k . Therefore, the ratios $[\omega_k L(\mathbf{x}^k; z)/L(\mathbf{x}; z)]/\omega_k = L(\mathbf{x}^k; z)/L(\mathbf{x}; z)$ and $[\omega_k H(\mathbf{x}^k; z)/H(\mathbf{x}; z)]/\omega_k = H(\mathbf{x}^k; z)/H(\mathbf{x}; z)$ will be greater than, equal to or smaller than one whenever country k 's contribution to the overall low- or high-impact level is greater than, equal to, or smaller than its publication share.

Decomposability can also be useful to express, say, the high-impact differences between two distributions as the sum of two terms involving differences in publication shares and differences in subgroups' high-impact levels. Consider the world citation distributions \mathbf{x} and \mathbf{y} in two periods of time, equally partitioned into K countries: $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^K)$, and $\mathbf{y} = (\mathbf{y}^1, \dots, \mathbf{y}^K)$. Let z_x and z_y be the possibly different CCLs in the two periods, and denote by ΔH the difference between the corresponding overall high-impact levels, that is, let

$$\Delta H = H(\mathbf{y}; z_y) - H(\mathbf{x}; z_x).$$

Denote by $\omega_k(\mathbf{x})$ and $\omega_k(\mathbf{y})$ the publication shares of country k in the two periods. Using A8, the quantity ΔH can be expressed as

$$\Delta H = \sum_k [\omega_k(\mathbf{y}) - \omega_k(\mathbf{x})] H(\mathbf{y}; z_y) + \sum_k [H(\mathbf{y}; z_y) - H(\mathbf{x}; z_x)] \omega_k(x).$$

The first term in the above expression captures the changes in publishing shares, while the second term captures the change in the countries high-impact levels.

From this discussion it can be concluded that conditions A7 and A8 constitute interesting properties for our measures to have.⁵ On the other hand, as will be seen in the Appendix A, FS show that under some reasonable conditions subgroup consistent and decomposable measures order equally all citation distributions, and that subgroup consistency is essential to characterize the class of indicators singled out in this paper for its use in empirical applications.

2.4. Sensitivity to distributional considerations

The next two axioms address the role that distributional considerations might play in the evaluation of citation distributions (the only two papers we know that stress the interest of distributional considerations in citation analysis are Allison, 1980, and Bornmann et al., 2008). Consider two research units whose citation distributions have the same size, the same MCR, and the same percentage of high-impact articles. Assume also that the total number of citations achieved by the high-impact articles is the same in both cases. Assume, however, that the high-impact articles of the first research unit receive a similar number of citations, while the number of citations among the high-impact articles in the second unit is very unequally distributed in the sense that most articles receive citations barely above the CCL while there are a few of Nobel prize quality because of the large number of citations they receive. It seems reasonable to conclude that the second research unit has a greater high-impact level than the first one.

For a numerical example, consider the citation distribution $\mathbf{x} = (0, 0, 2, 4, 5, 9)$, and assume that the CCL is 3. Thus, the set of high-impact articles is $(4, 5, 9)$. Assume that a transfer of one citation takes place between the second article in this set and the third, so that we reach a new distribution $\mathbf{y} = (0, 0, 2, 4, 4, 10)$. If distributional considerations matter, then we would say that distribution \mathbf{y} has greater high-impact level than distribution \mathbf{x} . To explain this idea, let us begin by defining a citation transfer as follows: given a distribution $\mathbf{x} = (x_1, \dots, x_n)$, a *citation transfer* between two articles with $0 < x_i < x_j$ is the transfer of one (or more) citations from article i to article j . Intuitively, a citation transfer increases the citation inequality of the original distribution.⁶ Hence:

A9.1. Transfer axiom for high-impact measures: If a citation transfer takes place between two high-impact articles, then the high-impact level increases.

It should be noted that distributional considerations do not play the same role in the evaluation of citation or income distributions. In an economic context, income inequality, even among the rich, has negative normative connotations. Therefore, the idea that greater income inequality should mean greater economic affluence would not be acceptable to economists. On the contrary, in an economic context it is natural to require that an increase in income inequality among the poor should also increase the economic poverty level. In our context, although this requirement is not so obvious, we believe that it is still acceptable. If the citation inequality among low-impact articles increases because a transfer takes place between a less cited to a more cited article, or in other words, from a more to a less deprived article, then in a straightforward sense the overall relative (citation) deprivation is increased and it seems acceptable to expect that the low-impact measure should

⁵ It should be noted that subgroup consistency and decomposability require strong doses of independence or autonomy among subgroups in all conceivable partitions. See Sen (1992, p. 106) for criticisms of this notion in an economic context.

⁶ A citation transfer might also be defined in terms of one (or more) citations from article j to article i without altering the ranking of the two articles, in which case a citation transfer intuitively decreases the citation inequality of the original distribution. 6

also increase. Although we see no major objection to using measures that satisfy it, it should be understood that the transfer axiom is not a fundamental property that all low-impact indices are expected to possess. Correspondingly, we list it separately here:

A9.2. Transfer axiom for low-impact measures: If a citation transfer takes place between two low-impact articles, then the low-impact level increases.

For later reference, note that the distributional consequences of a citation transfer are the same for the two notions of inequality usually distinguished by economists. To see this, consider the following conceptual experiment. Given a citation distribution $\mathbf{x}=(x_1, \dots, x_n)$, assume that there is a change in scale where the number of articles remain fixed but the total number of citations increases, say from X to X' where $X'=\lambda X$ for some $\lambda > 1$. How should the excess number of citations, $X'-X=(\lambda-1)X$, be allocated among the n articles so that the citation inequality originally associated with distribution \mathbf{x} remains constant? Economists usually provide two answers. The first is that citation inequality remains constant if relative positions remain unchanged, that is, if in the new distribution $\mathbf{x}'=(x'_1, \dots, x'_i, \dots, x'_n)$ we have that $x'_i=\lambda x_i$ for all $i=1, \dots, n$, so that $X'=\sum_i x'_i=\lambda X$. The second answer is that, instead of allocating the citation surplus in proportion to the originally number of citations, the total amount available should be allocated equally among all articles so that the absolute positions remain unchanged, that is, $x''_i=x_i+[(\lambda-1)X]/n$ for all $i=1, \dots, n$, with $X''=\sum_i x''_i=\lambda X=X'$. The two answers are said to capture a *relative* and an *absolute* notion of citation inequality. But note that if a citation transfer takes place, then citation inequality in both senses would increase.

2.5. Properties when the CCL varies

So far, implicitly the CCL has remained fixed. The last set of axioms deal with situations in which the CCL changes. In the first place, it is useful to introduce new properties to ensure the coherence of low- and high-impact values at different CCLs. This is typically accomplished by requiring the low- or high-impact value to be invariant to certain kinds of simultaneous changes in the citations to all articles and the CCL. As in the economic poverty literature, we consider here two types of changes that will simultaneously apply to both types of low- and high-impact indicators. Firstly, given a citation distribution $\mathbf{x}=(x_1, \dots, x_n)$ and a CCL z , consider a proportional, or relative change to a new situation in which $x'_i=\alpha x_i$ for all $i=1, \dots, n$, and $z'=\alpha z$ for some $\alpha > 0$. There is, of course, a change in scale, or in the total number of citations, $X'=\sum_i x'_i=\alpha X$, and hence in the distribution mean, $\mu(\mathbf{x}')=X'/n=\alpha X/n=\alpha \mu(\mathbf{x})$. However, the normalized low- and high-impact gaps defined in (2) and (4), as well as the relative citation inequality of the citation distribution, remain constant. It could be said that in a clear, “relative” sense the shape of the distribution remains constant, so that the low- and high-impact levels have not changed either, which is the idea expressed in the following axiom (that defines what the natural sciences call an *intensive* property):

A10. Invariance to equal proportional variations (scale invariance): Low- and high-impact indicators are invariant to equal proportional changes in all citations and the CCL.

Secondly, given a citation distribution $\mathbf{x}=(x_1, \dots, x_n)$ and a CCL z , consider an equal absolute change to a new situation in which $x''_i=x_i+\delta$ for all $i=1, \dots, n$, and $z''=z+\delta$ for some $\delta > 0$. Again, there is a change in scale, or in the total number of citations, $X''=\sum_i x''_i=X+n\delta$, and hence in the distribution mean, $\mu(\mathbf{x}'')=X''/n=X/n+\delta=\mu(\mathbf{x})+\delta$. However, the un-normalized low- and high-impact gaps defined in (1) and (3), as well as the absolute citation inequality of the citation distribution, remain constant. It could be said that in a clear, “absolute” sense the shape of the distribution remains constant, so that the low- and high-impact levels have not changed either, which is the idea expressed in the following axiom:

A11. Invariance to equal absolute variations (translation invariance): Low- and high-impact indicators are invariant to equal absolute changes in all citations and the CCL.

Low- and high-impact measures that satisfy axioms A10 or A11 are said to be scale or translation invariant. In the homogeneous case, MCRs from different research units are directly comparable. But this is not so in the heterogeneous case in which citations are not directly comparable across fields with different publishing and citation practices, and hence with different MCRs. However, both scale and translation invariance have very important practical consequences in this case because, as we will see in Section 4, they allow us to compare the low- and high-impact aspects of two citation distributions of the same size but different means.

Finally, it is also interesting to consider changes solely in the CCL. If the CCL, say increases, it is natural to expect the low-impact level to increase and the high-impact level to decrease. This scarcely debatable observation is the last axiom in our list:

A12. Increasing CCL Axiom: If the CCL increases, then the low-impact (high-impact) level increases (decreases).⁷

⁷ For a more detailed discussion of properties in the economic poverty literature that, after an appropriate reinterpretation, can be directly taken as properties of low- and high-impact citation distributions, see Foster (1984) and Zheng (1997).

3. A review of some specific L and H indicators and their alternatives

3.1. Two sets of L and H indicators

Given a citation distribution \mathbf{x} and a certain CCL z , let $l(\mathbf{x}; z)$ be the number of low-impact articles, and let $h(\mathbf{x}; z) = n(\mathbf{x}) - l(\mathbf{x}; z)$ be the number of high-impact articles. A first example of a specific low-impact indicator is the *low-impact ratio*, or the percentage of low-impact articles:

$$L_R(\mathbf{x}; z) = \frac{l(\mathbf{x}; z)}{n(\mathbf{x})}. \quad (5)$$

Similarly, we have the *high-impact ratio*, or the percentage of high-impact articles:

$$H_R(\mathbf{x}; z) = \frac{h(\mathbf{x}; z)}{n(\mathbf{x})}. \quad (6)$$

Of course,

$$L_R(\mathbf{x}; z) + H_R(\mathbf{x}; z) = 1,$$

so that if $L_R(\mathbf{x}; z)$ changes, then $H_R(\mathbf{x}; z)$ must change in the opposite direction. Clearly, both indicators satisfy the basic axioms A1–A4 and A6, as well as the convenient axioms A12, A8, and hence A7. Interestingly enough, both simultaneously satisfy the invariance axioms A10 and A11, a fact we shall return to below. However, both violate two crucial conditions: the monotonicity A5, and the transfers axioms A9.1 and A9.2 relating to the sensitivity of an indicator to distributional considerations.

There is a second type of natural, low-impact index that satisfies monotonicity, namely, the *low-impact gap ratio* defined as

$$L_I(\mathbf{x}; z) = \left(\frac{1}{z} \right) \left[\sum_{i=1}^{l(\mathbf{x}; z)} \frac{g_i(\mathbf{x}; z)}{l(\mathbf{x}; z)} \right] = \frac{z - \mu_L(\mathbf{x})}{z},$$

where $g_i(\mathbf{x}; z)$ is the low-impact gap defined in (1), and $\mu_L(\mathbf{x})$ is the mean citation rate of low-impact articles. Similarly, define the *high-impact gap ratio* as

$$H_I(\mathbf{x}; z) = \left(\frac{1}{z} \right) \left[\sum_{i=l(\mathbf{x}; z)+1}^{n(\mathbf{x})} \frac{g_i^*(\mathbf{x}; z)}{h(\mathbf{x}; z)} \right] = \frac{\mu_H(\mathbf{x}) - z}{z}$$

where $g_i^*(\mathbf{x}; z)$ is the high-impact gap defined in (3), and $\mu_H(\mathbf{x})$ is the mean citation rate of high-impact articles. These indices satisfy all basic axioms A1–A6; they are scale invariant but, due to the way they are normalized with $1/z l(\mathbf{x}; z)$ and $1/z h(\mathbf{x}; z)$ as a normalization factor, they are not subgroup consistent, a circumstance that might constitute a serious drawback in practice. Consider, however, the *per-article low-impact gap ratio* in which the normalization factor is $1/n(\mathbf{x})$:

$$L_G(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \left[\sum_{i=1}^{l(\mathbf{x}; z)} \frac{g_i(\mathbf{x}; z)}{z} \right] = \left[\frac{l(\mathbf{x}; z)}{n(\mathbf{x})} \right] \left\{ \left(\frac{1}{z} \right) \left[\sum_{i=1}^{l(\mathbf{x}; z)} \frac{g_i(\mathbf{x}; z)}{l(\mathbf{x}; z)} \right] \right\} = L_R(\mathbf{x}; z) L_I(\mathbf{x}; z). \quad (7)$$

This index represents the minimum number of citations required to bring all low-impact articles to the CCL, and is equal to the product of L_R and L_I . This convenient low-impact indicator satisfies the same properties as L_I but it is decomposable, and hence subgroup consistent. Define also the *per-article high-impact gap ratio*:

$$H_G(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \left[\sum_{i=l(\mathbf{x}; z)+1}^{n(\mathbf{x})} \frac{g_i^*(\mathbf{x}; z)}{z} \right] = H_R(\mathbf{x}; z) H_I(\mathbf{x}; z). \quad (8)$$

This high-impact indicator, which represents the surplus of citations actually received by high-impact articles above the CCL and is equal to the product of H_R and H_I , satisfies also axioms A1–A8 and A10.

It can be said that L_R and H_R only capture the *incidence* of the low- and high-impact aspects of any citation distribution, while L_G and H_G – the products of L_R and L_I and H_R and H_I , respectively – capture both the incidence and the *intensity* of the low- and high-impact aspects of a citation distribution. The problem, of course, is that none of the four indices are sensitive to distributional considerations. To observe this, consider two research units whose citation distributions have the same size, the same MCR, and the same percentage of high-impact articles, that is to say, the same H_R measure. Assume also that the total number of citations achieved by the high-impact articles is the same in both cases, so that both distributions have the same H_G . If this were all, then the two research units should be equally ranked from the high-impact point of view. Assume, however, that the citation inequality among high-impact articles is greater in the second research unit. It seems reasonable to conclude that this unit has a greater high-impact level than the first one.

3.2. The class of subgroup consistent low- and high-impact indicators

There are convenient decomposable, and hence subgroup consistent low- and high-impact indicators.⁸ Some of them, such as $L_G(\mathbf{x}; z)$ and $H_G(\mathbf{x}; z)$ defined in Eqs. (7) and (8), are also monotonic and satisfy all the basic properties A1–A5. The question is: are there many others simultaneously satisfying all these properties?

FS provide a definite answer to this question for poverty indices defined over income distributions which are equivalent to our low-impact measures defined over citation distributions. As seen in more detail in Appendix A, symmetric, replication invariance, continuous, focused, monotonic, and subgroup consistent low-impact measures are increasing transformations of decomposable measures, a result that justifies the use of the latter. Furthermore, the ranking of distributions obtained from a class of subgroup consistent low-impact indicators which satisfy axioms A1–A5 and other acceptable property introduced in Appendix A coincides with the ranking induced by the FGT class of low-impact indices defined by

$$L_\beta(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=1}^{l(\mathbf{x}; z)} \left[\frac{g_i(\mathbf{x}; z)}{z} \right]^\beta, \quad 0 \leq \beta. \quad (9)$$

The extension of these results to high-impact indicators is immediate, and the corresponding FGT class is defined by

$$H_\beta(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=l(\mathbf{x}; z)+1}^{n(\mathbf{x})} \left[\frac{g_i^*(\mathbf{x}; z)}{z} \right]^\beta, \quad 0 \leq \beta. \quad (10)$$

Together with the properties already mentioned, all members of these two families are scale invariant and satisfy axioms 6 and 12. Interestingly enough, the members of these families corresponding to parameter values $\beta=0$, 1 coincide with the indices already presented in this Section in Eqs. (5)–(8): $L_0(\mathbf{x}; z) = L_R(\mathbf{x}; z)$, $L_1(\mathbf{x}; z) = L_G(\mathbf{x}; z)$, $H_0(\mathbf{x}; z) = H_R(\mathbf{x}; z)$, and $H_1(\mathbf{x}; z) = H_G(\mathbf{x}; z)$. What are novel are the cases in which β is greater than 1. When $\beta=2$, the only case considered in this paper, we have:

$$L_2(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=1}^{l(\mathbf{x}; z)} \left[\frac{g_i(\mathbf{x}; z)}{z} \right]^2, \quad (11)$$

and

$$H_2(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=l(\mathbf{x}; z)+1}^{n(\mathbf{x})} \left[\frac{g_i^*(\mathbf{x}; z)}{z} \right]^2. \quad (12)$$

The measures H_2 and L_2 are the only ones in this paper which satisfy axioms A9.1 and A9.2. To better understand their sensitivity to distributional considerations, it is useful to realize that

$$L_2(\mathbf{x}; z) = L_R(\mathbf{x}; z) \{ [L_G(\mathbf{x}; z)]^2 + [1 - L_G(\mathbf{x}; z)]^2 (C_L)^2 \}, \quad (13)$$

$$H_2(\mathbf{x}; z) = H_R(\mathbf{x}; z) \{ [H_G(\mathbf{x}; z)]^2 + [1 - H_G(\mathbf{x}; z)]^2 (C_H)^2 \} \quad (14)$$

where $(C_L)^2$ and $(C_H)^2$ are the squared coefficient of variation (i.e. the ratio of the standard deviation over the mean). This implies that over distributions with the same L_R and L_G (or the same H_R and H_G), L_2 and $(C_L)^2$, or H_2 and $(C_H)^2$ give precisely the same ranking. The coefficient of variation is an inequality index that exhibits a *transfer neutrality* property based on citation differences: a citation transfer of a given amount between two low- or high-impact articles a certain distance apart will always have the same effect on low- or high-impact levels irrespective of the absolute number of citations of the articles involved.⁹ Finally, expressions (13) and (14) clearly indicate that the measures L_2 and H_2 capture simultaneously the incidence, the intensity, and the *citation inequality* aspects which have been discussed in Section 2.

4. The heterogeneous case

Assume we want to compare the performance of molecular biologists in MIT in the U.S. with mathematicians in the University of Lisbon in Portugal. Molecular Biology and Mathematics are clearly heterogeneous fields. Because of large differences in publication and citation practices, the number of citations received by articles in these two fields are not directly comparable. In the study of economic poverty, the analogous problem would be the comparison of

⁸ For examples of low-impact indices that are subgroup consistent but not decomposable, as well as indices that are not subgroup consistent, see FS (p. 692–693). By taking into account how properties A4 and A5 work in both cases, it is not difficult to convert all these low-impact indices into high-impact ones.

⁹ This property is emphasized by Allison (1980) when he recommends the coefficient of variation as an appropriate index of scientific productivity inequality.

poverty in the state of Massachusetts in the U.S., say, with poverty in the Lisbon region in Portugal. The heterogeneity arises from the differences in the standard of living between the two regions. The solution is to measure poverty in these two geographical areas relative to the standard of living in the two countries, the U.S. and Portugal. The usual way of proceeding in the poverty literature is to choose the poverty line of each country equal to some fraction of its mean (or its median) income, say 50% of the mean (or 60% of the median) in Portugal and the U.S. The poor in Massachusetts and the Lisbon region will be those individuals with incomes below the corresponding, relative poverty lines.

In our approach, all that is required in the heterogeneous case is to work with scale (or translation) invariant indicators and to fix appropriately the CCLs of the two fields in contention. Given the skewness of citation distributions and the special interest of high-impact articles, it may be appropriate to fix the CCL for any field at some common multiple $b > 1$ of the corresponding mean (or median) citation rates. Let \mathbf{x} and \mathbf{y} be the citation distributions of molecular biologists in MIT and mathematicians in the University of Lisbon, let μ_{MB} and μ_M be the MCRs in the world citation distribution of Molecular Biology and Mathematics, and let $z_{MB} = b\mu_{MB}$ and $z_M = b\mu_M$ be the CCLs in Molecular Biology and Mathematics fixed as a common multiple b of the corresponding MCRs. In principle, high-impact levels in the original situation, $H(\mathbf{x}; z_{MB})$ and $H(\mathbf{y}; z_M)$, appear not to be comparable. Let $\lambda = \mu_{MB}/\mu_M$, which in practice may be approximately equal to 10. If the high-impact measure is scale invariant, we can multiply by λ the citations of all articles in \mathbf{y} , as well as the CCL in Mathematics, z_M . Denote the new distribution and CCL by \mathbf{y}' and z'_M . According to A10, we now have $H(\mathbf{y}; z_M) = H(\mathbf{y}'; z'_M)$. But $z'_M = \lambda z_M = \lambda b\mu_M = b\mu_{MB} = z_{MB}$, so that $H(\mathbf{y}'; z'_M) = H(\mathbf{y}'; z_{MB})$. Therefore, the original comparison is equivalent to the comparison of $H(\mathbf{x}; z_{MB})$ with $H(\mathbf{y}'; z_{MB})$. Hence, as long as scale invariant indicators are used, in our approach the low- and high-impact levels of citation distributions representing the performance of research units in different fields can be easily compared. Naturally, the same applies if the comparison is between the low- and high-impact levels of entire heterogeneous fields, such as Molecular Biology and Mathematics at the world level—or the analogous comparison in the economic area between poverty levels in the U.S. and Portugal as a whole.

In brief, since size independent and scale invariant low- or high-impact indicators – such as the members of the FGT families – capture aspects of the shape of a citation distribution independently of its size and its scale, the performance of research units across heterogeneous fields, or the fields' citation distributions themselves, are directly comparable in our approach. For brevity, the possibility of heterogeneous comparisons using translation invariant indicators is left for the reader to work out.

5. Properties of alternative procedures

As indicated in Section 1, there are three alternative procedures in the literature for the evaluation of research units in a homogeneous field: indices of excellence, such as the h -index, the CSS technique, and what we call the Leiden triad. Naturally, the focus and normalization axioms A4 and A6, as well as A12 are not applicable in this case, while the monotonicity A5, the subgroup consistency and decomposability properties A7 and A8, as well as the transfer and invariance axioms A9 and A10 can be easily rephrased to apply to these alternative procedures by eliminating any reference to any CCL or to the partition into low- and high-impact articles. Table 2 summarizes which axioms are satisfied by these three procedures, in contrast to those satisfied by the three members of the FGT family of low- and high-impact indicators corresponding to $\beta = 0, 1, 2$.

Table 2
Axioms satisfied by the different procedures' indicators.

	(A) h -index	(B) CSS	(C) Leiden triad indicators			FGT indicators		
			(1) % uncited articles, (2) Share of top 5%	3) MCR	$\beta=0$	$\beta=1$	$\beta=2$	
A1. Symmetry	Y	Y	Y	Y	Y	Y	Y	
A2. Replication invariance	*	Y	Y	Y	Y	Y	Y	
A3. Continuity		Y	Y	Y	Y	Y	Y	
A4. Focus	NA	NA	NA	NA	Y	Y		
A5. Monotonicity	*	*	*	Y	Y	Y		
A6. Normalization	NA	NA	NA	NA	Y	Y		
A7. Subgroup consistency	*	Y	Y	Y	Y	Y		
A8. Decomposability		*	Y	Y	Y	Y	Y	
A9. Transfer axiom	*	*	*	*	*	Y		
A10. Scale invariance	*	Y	Y	Y	Y	Y		
A11. Translation invariance	*	*	E	*	*	*		
A12. Increasing CCL axiom	NA	NA	NA	NA	Y	Y	Y	

CSS=Characteristic Scores and Scales; MCR=Mean Citation Rate; Y=Yes, the axiom is satisfied; *=No, the axiom is not satisfied; NA=Not applicable. E(exception)=An asterisk * applies for procedure (1), while Y is the case for procedure (2).

Starting with the h -index, it clearly violates the replication invariance axiom A2.¹⁰ Moreover, it is not monotonic, subgroup consistent, sensitive to distributional considerations, and neither scale nor translation invariant. However, it has been characterized in terms of a number of interesting properties (Quesada, 2009, 2010; Woeginger, 2008a, 2008b). Moreover, contrary to any of the high-impact indices introduced in this paper, it is robust to the presence of extreme observations, a fact that may prove to be very important in empirical applications. Thus, the h -index and the replication invariant procedures discussed in this paper serve very different purposes, and therefore constitute essentially complementary approaches to the same evaluation problem in the homogeneous case. On the other hand, in the heterogeneous case the h -indices for research units working in different fields are nor directly comparable.

Next, consider the CSS technique. In the original applications the following four *characteristic scores* are determined: $s_0 = 0$; $s_1 = \text{MCR}$; $s_2 = \text{mean citation rate of articles with citations above the MCR}$; $s_3 = \text{mean citation rate of articles with citations above } s_2$. These scores are used to partition the set of articles into five categories of citedness:

Category 1 $x_i = s_0$	Articles that receive no citations;
Category 2 $x_i \in (s_0, s_1)$	Articles that receive <i>few</i> citations, namely, citations lower than the MCR;
Category 3 $x_i \in [s_1, s_2)$	Articles that receive a <i>fair</i> amount of citations, namely, at least the MCR but below s_2
Category 4 $x_i \in [s_2, s_3)$	Articles that receive a remarkable amount of citations, namely, no lower than s_2 but below s_3 ;
Category 5 $x_i \geq s_3$	Articles that receive an <i>outstanding</i> amount of citations, namely, greater or equal than s_3

In our view, the way to proceed with this approach in the homogeneous case is as follows; just as a fixed, common CCL for all research units has been assumed in a given field, here we have four characteristic scores s_0 to s_3 and five categories characterizing the world citation distribution. Therefore, one way to evaluate a number of research units is to compute their percentage distributions over the five categories, and compare them among themselves and with respect to the one for the entire field in question.¹¹

The classification of any ordered distribution over the five categories satisfies axioms A1–A3, which means that it is replication invariant. This technique is also additively decomposable, and hence subgroup consistent. On the other hand, CSS is scale invariant because the multiplication by a common scalar of the citations received by all articles in a citation distribution does not alter category 1, multiplies s_1 , s_2 , and s_3 by the same factor, and therefore leaves unchanged the percentage distribution over the citedness categories. However, after a common number of citations is added up to all articles no articles remain uncited and category 1 is altered, so that the CSS procedure is not translation invariant. More importantly, as the following examples show the procedure is neither monotonic nor sensitive to distributional considerations. Consider distribution $\mathbf{x} = (0, 0, 0, 3, 5, 6, 8, 10)$, for which $s_1 = 4$, $s_2 = 7.25$, and $s_3 = 9$. Thus category 1 represents 37.5%; category 2, 12.5%; category 3, 25%, and categories 4 and 5, 12.5% each. In distribution $\mathbf{y} = (0, 0, 0, 4, 5, 6, 8, 10)$, where the fourth article receives one more citation, $s'_1 = 4.25$ rather than 4 as before, but the remaining characteristic scores are unchanged. Clearly, the percentage distribution over categories is also unchanged, which shows that this procedure violates monotonicity. Finally, consider distribution $\mathbf{v} = (0, 0, 0, 3, 4, 6, 9, 10)$, where the seventh article gets one more citation taken from the fifth one in \mathbf{x} . The MCR is the same, so that s'_1 is still equal to 4, but the remaining characteristic scores change: $s_2 = 8.33$, and $s_3 = 9.5$. However, the percentage distribution over the five categories remains unchanged, which shows that this procedure is not sensitive to distributional considerations.¹²

It should be noted that the percentage distribution of the citations received by research unit k in field i over the citedness categories is not directly comparable with the corresponding distribution of the citations received by research unit l in field j . However, after appropriate normalization by the citedness categories of the two fields in question, the CSS approach is able to deal with the heterogeneous case.

Finally, relative impact indicators have been used and recommended by many authors, notably, those from the Leiden group (Moed et al., 1985, 1995; Van Raan, 2004), authors from the Budapest group (Braun et al., 1985; Glänzel et al., 2002; Schubert & Braun, 1986; Schubert, Glänzel, & Braun, 1983, 1988), and Vinkler (1986, 2003). For a research unit working in a number of sub-fields, a relative impact indicator is the ratio between the unit's observed performance, measured by the citations received by their published articles, and the expected citation rate of a relevant reference standard.¹³ As

¹⁰ For interesting attempts to correct for this feature in search of size independent indicators see Katz (1999, 2000) or Molinari and Molinari (2008a, 2008b) and Kinney (2007), as well as the discussion in Alonso et al. (2009).

¹¹ Alternatively, Schubert et al. (1987) would compute the percentage distribution over the five categories for the world citation distribution in a given (homogeneous) field, and would compare a set of research units by placing their the MCRs into the category they belong to determine if on average the units receive a few, a fair, a remarkable or an outstanding number of citations. They actually do this for a number of journals that belong to one of the 114 subfields to which they applied the procedure.

¹² On the other hand, note that if we choose a $\text{CCL}z = s_2$, then the low-impact ratio L_R defined in Eq. (5) coincides with the sum of the first three categories of citedness in these authors' approach, while the high-impact ratio H_R defined in (6) coincides with the sum of the last two citedness categories. This also serves to show that the procedure is not monotonic and not sensitive to distributional considerations.

¹³ It should be emphasized that, contrary to what is assumed in this paper, these authors define their measures counting on information about the journal where each paper is published. In the case of the Leiden group, this information allows them to compare the observed behavior of relatively small research

pointed out in Section 1, under the restricted information assumed to be available in this paper, judging research units in a homogeneous field by their MCRs is in line with the literature on relative impact indicators. However, taking into account the typical skewness of citation distributions, the authors of the Leiden group have completed their indicators based on average values in two directions. Thus, what we have called the Leiden triad consists of the MCR, the percentage of uncited articles, and the percentage contribution to the top 5% of the most highly cited articles.

The last two indicators satisfy several interesting properties (A1–A3, A7, A8 and A10), but are neither monotonic nor sensitive to distributional considerations.¹⁴ As far as the MCR, it should be emphasized that it satisfies most applicable axioms: it is symmetric and continuous, as well as replication invariant, decomposable, and scale invariant. Since it is also monotonic, from the point of view of the properties discussed in Section 2 the only shortcoming of the MCR as an indicator of scientific performance is that it is not sensitive to distributional considerations. To illustrate this feature, consider the possibility of accepting the existence of a reasonable CCL and evaluating citation distributions by two different means: the inverse of the MCR among low-impact articles, $1/\mu_L$, and the MCR among high-impact ones, μ_H . In the paper's conceptual framework, $1/\mu_L$ and μ_H constitute a low-impact and a high-impact indicator, respectively. Consider a citation distribution $\mathbf{x} = (0, 0, 0, 2, 3, 5, 7, 9)$ and a CCL equal to 4. Clearly, $\mu_L(\mathbf{x}; z) = 1$, and $\mu_G(\mathbf{x}; z) = 7$. Consider now the distribution $\mathbf{y} = (1, 1, 1, 1, 1, 7, 7, 7)$ for which μ_L and μ_G remain constant. Of course, the evaluation according to $1/\mu_L$ and μ_H would rank equally \mathbf{x} and \mathbf{y} . The same would be the case using the overall MCR. However, a high-impact index H satisfying A9.1 would indicate that $H(\mathbf{x}; z) > H(\mathbf{y}; z)$, while a low-impact index L satisfying A9.2 would indicate that $L(\mathbf{x}; z) < L(\mathbf{y}; z)$.¹⁵

Because of large differences in publication and citation practices, the MCRs of research units of two different homogeneous fields are not directly comparable. But, of course, this is the problem originally addressed by relative indicators. After appropriate normalization, the MCRs of research units in heterogeneous fields become comparable. However, this is not the case if we are interested in the comparison of two entire heterogeneous fields: we lack an appropriate relative reference for, say, the MCR of Molecular Biology and Mathematics.

6. Conclusions

This paper has addressed the evaluation of the scientific performance of research units of a certain size when the only available information is the citation distribution of the papers published by the research unit in the serial literature in a homogenous field, namely, when the number of citations received by each pair of articles directly reflects the different merit, or international impact of each of the two articles. Given that citation distributions are typically highly skewed, we introduce an evaluation procedure in which each (ordered) citation distribution is characterized in terms of two indicators, a low- and a high-impact index defined to the left and to the right of an appropriate CCL. These indicators must be understood as 'partial indicators' in the sense of Martin and Irvine (1983), or as potentially useful elements in the scoreboard approach advocated in Tijssen (2003).

Which indicators should be used in practice? This is a question that must be answered after the purpose of the evaluation exercise has been determined. However, we believe that in many contexts the FGT family of indicators, closely associated with the family of economic poverty indices originally suggested by Foster et al. (1984), may prove to be very useful. The reasons suggested in the paper can be summarized as follows. Firstly, the properties that characterize the ranking induced by this family of indicators are known. Among them, the following three should be emphasized. Replication invariance ensures that an indicator is independent of the size of the set of papers under evaluation. Monotonicity requires that a new citation should lower the low-impact index, and should increase the high-impact indicator. Subgroup consistency, a property with useful practical implications, ensures that the aggregate, or overall low- (or high-) impact value does not respond perversely to changes in the level of low- (or high-) impact within the subgroups of any partition. Secondly, the first member of each family coincides with the low- or the high-impact ratio and captures the incidence aspect of the two phenomena under investigation; the second member coincides with the per-article low-impact gap ratio, and captures both the incidence and the intensity aspect of each of the two phenomena, while the third member – in addition to these two aspects – is sensitive to distributional considerations, in the sense that an increase in the citation inequality according to the coefficient of variation, increases both the low- and the high-impact measures. Thirdly, all members of the two families of indicators satisfy other convenient properties. In particular, they are all scale invariant in the sense that multiplying all elements of a citation distribution and the CCL by a common scalar factor leaves the low- and high-impact measures unchanged. This opens the door to the comparison of citation distributions with different means (medians, or other first order moments of the citation distribution) in the heterogeneous case.

units, namely research groups, with the expected behavior of the set of journals where the research group is known to publish. The ratio of such expected behavior to the behavior of the journals in the entire field constitutes another interesting indicator in this case. Finally, the possibility of ordering the set of journals in a field in terms of their relative impact allows the authors in the Budapest group to graphically represent relative impact indicators in two-dimensional relational charts (Glänzel et al., 2002; Schubert & Braun, 1986).

¹⁴ Note that the same applies to the median, a centrality statistic that is robust to extreme observations, a common feature of citation distributions.

¹⁵ Together with J. Crespo, in Albarrán, Crespo, Ortuño, and Ruiz-Castillo (2010b) we compare the U.S. and the European Union using their MCRs as well as their percentage shares at many percentiles of the world citation distribution in 22 homogeneous scientific fields. Of course, this variant of the Leiden triad is neither monotonic nor sensitive to distributional considerations.

How do these indicators compare with some of the alternatives available in the literature? In the first place, it has been shown that CSS for partitioning any citation distribution into a number of citation classes is neither monotonic nor sensitive to distributional considerations. In the second place, within the Leiden triad the percentages of uncited articles or share of the top 5% most highly cited articles have the same shortcomings as the CSS procedure. As far as the MCR is concerned, or the possibility of computing two means among the low- and the high-impact articles, the only shortcoming is that average-based indicators are not sensitive to distributional considerations. In the third place, note that if the CCL is fixed sufficiently high, the high-impact indicator may be considered an index of excellence comparable with the h -index and its variants but possessing very different properties. Thus, the replication invariant high-impact indicators presented in this paper and the size dependent h -index constitute essentially complementary approaches to the same evaluation problem.

Of course, whether the extra properties enjoyed by the indicators we have introduced are of any interest is not merely a formal issue. The value added by these properties, if any, can only be revealed by their use in practice (see Albarrán et al., 2010c) for the first application of this methodology to articles published by the U.S., the European Union and the rest of the world in 1998–2002, with a five year citation window in 22 scientific fields).

Acknowledgements

The authors acknowledge financial support by Santander Universities Global Division of *Banco Santander*. Albarrán and Ruiz-Castillo acknowledge financial support from the Spanish MEC, the former through grants SEJ2007-63098 and SEJ2006-05710, and the latter 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. Ruiz-Castillo acknowledges conversations with Coral del Río ten years ago about the notion of economic affluence in income space. Comments and suggestions by Juan A. Crespo are also gratefully acknowledged.

Appendix A. The characterization of subgroup consistent low- and high-impact indicators (this Appendix borrows heavily from FS)

In order to cover the more general case, consider discrete citation distributions represented by vectors drawn from the set $\Psi = \cup_{n=1}^{\infty} \Omega^n = \Omega^n$, where $\Omega = (a, b)$ is some nondegenerate real interval, and Ω^n is the set of all n -tuples of elements from Ω . For any given CCL $z \in \Omega$ and distribution $\mathbf{x} \in \Psi$, the low-impact domain is $D_L(z) = \{t \in \Omega | t < z\}$ and the high-impact domain is $D_H(z) = \{t \in \Omega | t \geq z\}$. As in the text, the number of articles in \mathbf{x} , and the number of low- and high-impact articles is denoted by $n(\mathbf{x})$, $l(\mathbf{x}; z)$ and $h(\mathbf{x}; z)$.

A low-impact index is a function $L: \Psi \times \Omega \rightarrow R$ whose typical value $L(\mathbf{x}; z)$ indicates the low-impact level associated with the distribution \mathbf{x} and the CCL z , while a high-impact index is a function $H: \Psi \times \Omega \rightarrow R$ whose typical value $H(\mathbf{x}; z)$ indicates the high-impact level associated with the distribution \mathbf{x} and the CCL z . In order to formally state some of the basic properties introduced in the text, we say that $\mathbf{x} \in \Psi$ is obtained from $\mathbf{y} \in \Psi$ by a permutation if $\mathbf{x} = \mathbf{y}\Pi$ for some permutation matrix Π ; by a $(k-)$ replication if $n(\mathbf{x}) = kn(\mathbf{y})$ and $\mathbf{x} = (\mathbf{y}, \mathbf{y}, \dots, \mathbf{y})$ for some positive integer k ; and by an increment to a low- (resp. high-) impact article if $x_i = y_i$ for all $i \neq j$ and $x_j > y_j$ for some $y_j < z$ (resp. $y_j \geq z$). It is assumed throughout that the L and H indices satisfy the following five basic properties for any given CCL $z \in \Omega$.

- A1. *Symmetry*: $L(\mathbf{x}; z) = L(\mathbf{y}; z)$ and $H(\mathbf{x}; z) = H(\mathbf{y}; z)$ whenever $\mathbf{x} \in \Psi$ is obtained from $\mathbf{y} \in \Psi$ by a permutation.
- A2. *Replication invariance*: $L(\mathbf{x}; z) = L(\mathbf{y}; z)$ and $H(\mathbf{x}; z) = H(\mathbf{y}; z)$ whenever $\mathbf{x} \in \Psi$ is obtained from $\mathbf{y} \in \Psi$ by a replication.
- A3. *Continuity*: $L(\mathbf{x}; z)$ and $H(\mathbf{x}; z)$ are continuous as functions of \mathbf{x} on Ψ .¹⁶
- A4. *Focus*: $L(\mathbf{x}; z) = L(\mathbf{y}; z)$ and $H(\mathbf{v}; z) = H(\mathbf{w}; z)$ whenever $\mathbf{x} \in \Psi$ is obtained from $\mathbf{y} \in \Psi$ by an increment to a high-impact article, and $\mathbf{v} \in \Psi$ is obtained from $\mathbf{w} \in \Psi$ by an increment to a low-impact article.
- A5. *Monotonicity*: $L(\mathbf{x}; z) \leq L(\mathbf{y}; z)$ and $H(\mathbf{v}; z) \geq H(\mathbf{w}; z)$ whenever $\mathbf{x} \in \Psi$ is obtained from $\mathbf{y} \in \Psi$ by an increment to a low-impact article and $\mathbf{v} \in \Psi$ is obtained from $\mathbf{w} \in \Psi$ by an increment to a high-impact article (This weak form of the monotonicity condition is closely related to the 'Citations Do Not Harm' axiom in Marchant, 2009).

The characterization argument in FS can be summarized in two steps. First, take the CCL z as fixed and omit it as an argument in all expressions to simplify the notation. FS start by proving that any continuous subgroup consistent low-impact index must be a continuous and increasing transformation of what will be called a *canonical low-impact index*, L^ϕ , defined as

$$L^\phi(\mathbf{x}) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=1}^{n(\mathbf{x})} \phi(x_i) \quad \text{for all } \mathbf{x} \in \Psi, \quad (15)$$

¹⁶ This strong version could be weakened to a notion of *restricted continuity* of the L and H indicators on the $D_L(z)$ and $D_H(z)$ domains, respectively—a technical point to which we will return below.

where $\phi: \Omega \rightarrow R$ is a continuous and non-increasing function such that $\phi(t) = 0$ for all $t \geq z$. The function ϕ may be regarded as a measure of (citation) deprivation—a non-increasing function that attains its minimum level of 0 at z , and maintains this value for all high-impact articles. Consequently, the index L^ϕ takes overall low-impact to be the average deprivation for the distribution \mathbf{x} as a whole. Formally, in FS we have:

Proposition 1. *$L: \Omega \rightarrow R$ is a continuous, subgroup consistent low-impact index if, and only if, there exist $\phi: \Omega \rightarrow R$ and $F: \phi(\Omega) \rightarrow R$ such that*

$$L(\mathbf{x}) = F[L^\phi(\mathbf{x})] \quad \text{for all } \mathbf{x} \in \psi$$

where F is continuous and increasing; $L^\phi(\mathbf{x}) = [1/n(\mathbf{x})] \sum_{i=1}^{n(\mathbf{x})} \phi(x_i)$; ϕ is continuous and non-increasing, and $\phi(t) = 0$ for all $t \geq z$.

The next observation is that canonical indices essentially form the subclass of continuous, decomposable low-impact indices. For if L is continuous and decomposable it is also subgroup consistent and canonical, and applying Proposition 1 the following result is easy to prove:

Corollary 1. *L is a continuous decomposable low-impact index if, and only if, $L = L^\phi + c$ for some canonical index L^ϕ and some constant c .*

Therefore, combining Proposition 1 and Corollary 1 we obtain

Corollary 2. *L is a continuous, subgroup consistent low-impact index if and only if L is a continuous, increasing transformation of a continuous, decomposable low-impact index.*

Therefore, there is a direct relationship between subgroup consistent and decomposable low-impact measures. All decomposable measures are subgroup consistent and all subgroup consistent, under some reasonable conditions, are increasing transformations of decomposable low-impact measures. For those that regard decomposability as putting too detailed a restriction on the functional form of a low-impact index, the FS results justify, from an ordinal point of view, the use of low-impact measures satisfying this property. In FS's own words, "subgroup consistency thus provides a means of justifying the use of decomposable low-impact measures. For, corresponding to each continuous subgroup consistent index, there is a continuous decomposable index¹⁷ which ranks distributions in precisely the same way" (p. 696–697).

These results can be immediately extended to the high-impact case by introducing the concept of a *canonical high-impact index*, H^ϕ , defined as

$$H^\phi(\mathbf{x}) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=1}^{n(\mathbf{x})} \phi(x_i) \quad \text{for all } \mathbf{x} \in \Psi, \quad (16)$$

where $\Phi: \Omega \rightarrow R$ is a continuous and nondecreasing function such that $\phi(t) = 0$ for all $t < z$. The function Φ may be regarded as a measure of (citation) affluence—a nondecreasing function that attains its minimum level of 0 at z , and maintains this value for all low-impact articles. Consequently, the index H^ϕ takes overall high-impact to be the average affluence for the distribution \mathbf{x} as a whole. The FS procedures applied to high-impact monotonic indices would lead to the following result:

Corollary 3. *H is a continuous, subgroup consistent high-impact index if, and only if, H is a continuous, increasing transformation of a continuous, decomposable high-impact index.*

In the second place, when variations in the CCL z are allowed we have introduced axioms A10 and A11 in Section 2.5 to ensure the coherence of low- and high-impact values at different CCLs. An obvious question to ask is whether a subgroup consistent low-impact index can be both a scale invariant and a translation invariant low-impact index. The answer by FS is that the class of subgroup consistent, scale invariant, and translation invariant low-impact indices is not very interesting, since it only includes monotonic transformations of the low-impact ratio $L_R(\mathbf{x}; z)$ (see Proposition 6 in FS).¹⁸ However, a natural generalization comes immediately to the forefront: a scale invariant index and a translation invariant index are said to be *compatible* if, at any fixed CCL, they give the same ranking of distributions, although not necessarily the same values. FS then ask if compatible pairs of indices exist that are subgroup consistent. It is shown that if a pair of subgroup consistent low-impact indices is compatible, then the scale invariant index must be an increasing transformation of a member of the FGT class defined by

$$L_\beta(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=1}^{l(\mathbf{x}; z)} \left[\frac{g_i(\mathbf{x}; z)}{z} \right]^\beta, \quad 0 \leq \beta. \quad (17)$$

¹⁷ From a technical point of view, it is important to point out that by relaxing the continuity assumption as indicated in note 16, we gain only transformations of the low-impact ratio, or of some combination of this ratio and a canonical index (see Propositions 2 and 3, and Corollaries 2 and 3 in FS).

¹⁸ For simplicity, FS prove this and the following results for distributions consisting of strictly positive elements; in our context, this implies the restriction to articles with a positive number of citations.

The equivalent result for high-impact indices can be stated as follows: if a pair of subgroup consistent high-impact indices is compatible, then the scale invariant index must be an increasing transformation of a member of the FGT class of high-impact measures defined by

$$H_{\beta}(\mathbf{x}; z) = \left[\frac{1}{n(\mathbf{x})} \right] \sum_{i=(\mathbf{x};z)+1}^{n(\mathbf{x})} \left[\frac{g_i^*(\mathbf{x}; z)}{z} \right]^{\beta}, \quad 0 \leq \beta. \quad (18)$$

Three comments are in order. First, the FGT class defined in (17) results from making $\phi(t) = [(z-t)/z]^{\beta}$ for $t < z$ in the definition of a canonical low-impact index in Eq. (15), that is, from measuring the (citation) deprivation of any low-impact article by means of its low-impact gap defined in Eq. (2) to the β power. Similarly, the FGT class defined in (18) results from making $\phi(t) = [(t-z)/z]^{\beta}$ for $t \geq z$ in the definition of a canonical low-impact index in Eq. (16), that is, from measuring the (citation) affluence of any high-impact article by means of its high-impact gap defined in Eq. (4) to the β power. Second, the above results may be viewed as a complete axiomatic characterization of two important classes of rankings in the following sense. If we choose to adopt a nontrivial scale invariant measure of low- or high-impact which is not an increasing function of one of the members of the L_{β} and H_{β} families, respectively, then the chosen index must either violate continuity or subgroup consistency, or else have no translation invariant low- or high-impact index counterpart which ranks citation distributions in the same way for any given CCL. Third, these characterization results mean that we know exactly which axioms or value judgments are invoked when we use the FGT families to obtain a ranking of citation distributions. However, in practice we use members of these families, for example those corresponding to the choice $\beta = 2$, in order to establish cardinal comparisons of the sort: distribution \mathbf{x} has twice the low- or high-impact level as distribution \mathbf{y} according to L_2 and H_2 . But different indices that induce the same ranking will typically lead to different cardinal statements. Consider the square of the above indices: $L'_2 = (L_2)^2$, and $H'_2 = (H_2)^2$. They will rank distributions \mathbf{x} and \mathbf{y} exactly as L_2 and H_2 , but the cardinal statement would be changed as follows: distribution \mathbf{x} has four times the low- or high-impact level as distribution \mathbf{y} according to L'_2 and H'_2 . The characterization of the FGT indices is not yet known. In the meantime, a defense of its use for cardinal purposes only rests on the interest in specific evaluation contexts of the properties they are shown to satisfy in Section 3 and Table 2.

Finally, it should be observed that many common indices widely used in the income poverty area, which in our context can be taken as low-impact indicators, are functions of the normalized low- and high-impact gaps defined in Eqs. (2) and (4).¹⁹ As indicated in footnote 8, by taking into account how properties A4 and A5 work in both cases, it is not difficult to convert all those low-impact indices into high-impact ones.

References

- Aksnes, D., & Sivertsen, G. (2004). The effect of highly cited papers on national citation indicators. *Scientometrics*, 59, 213–224.
- Albarrán, P., Crespo, J., Ortuño, I., & Ruiz-Castillo, J. (2010a). The skewness of science in 219 sub-fields and a number of aggregates. *Mimeo*.
- Albarrán, P., Crespo, J., Ortuño, I., & Ruiz-Castillo, J. (2010b). A comparison of the scientific performance of the U. S. and Europe at the turn of the 21st century. *Scientometrics*, 82, 329–344.
- Albarrán, P., & Ruiz-Castillo, J. (2009). References made and citations received by scientific articles. *Working Paper 09-81, Economics Series 45*, Universidad Carlos III, accepted for publication in *Journal of the American Society for Information Science and Technology*.
- Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2010c). High- and low-impact citation measures: empirical applications. *Working Paper 10-09*, Universidad Carlos III.
- Allison, P. (1980). Inequality and scientific productivity. *Social Studies of Science*, 10, 163–179.
- Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3, 273–289.
- Bornmann, L., Mutz, R., Nehaus, C., & Daniel, H.-D. (2008). Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, 8, 93–102.
- Braun, T., Glänzel, W., & Schubert, A. (1985). *Scientometrics indicators. A 32 country comparison of publication productivity and citation impact*. Singapore, Philadelphia: World Scientific Publishing Co. Pte. Ltd.
- Chakravarty, S. R. (1983). A new index of poverty. *Mathematical Social Science*, 6, 307–313.
- Clark, S., Hemming, R., & Ulph, D. (1981). On indices for the measurement of poverty. *Economic Journal*, 91, 515–526.
- Dasgupta, P., Sen, A., & Starrett, D. (1973). Notes on the measurement of inequality. *Journal of Economic Theory*, 6, 180–187.
- Foster, J. E. (1984). *On economic poverty: A survey of aggregate measures*. *Advances in econometrics* JAI Press Inc., pp. 15–251.
- Foster, J. E., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52, 761–766.
- Foster, J. E., & Shorrocks, A. (1991). Subgroup consistent poverty indices. *Econometrica*, 59, 687–709.
- Glänzel, W. (2002). Co-authorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, 50, 461–473.
- Glänzel, W. (2007). Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, 1, 92–102.
- Glänzel, W., & Schubert, A. (1988a). Characteristic scores and scales in assessing citation impact. *Journal of Information Science*, 14, 123–127.
- Glänzel, W., & Schubert, A. (1988b). Theoretical and empirical studies of the tail of scientometric distributions. In L. Egghe, & R. Rousseau (Eds.), *Informetrics 87–88: International Conference Proceedings* (pp. 75–83). Elsevier Science Publisher.
- Glänzel, W., Schubert, A., & Braun, T. (2002). A relational charting approach to the world of basic research in twelve science fields at the end of the second millennium. *Scientometrics*, 55, 335–348.
- Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the national Academy of Sciences of the United States of America*, 102, 16569–16572.

¹⁹ According to Jenkins and Lambert (1997, Table 1, p. 318), these are the indices first introduced in Chakravarty (1983); Clark et al. (1981) Type 2; Johnson (1988); Pyatt (1987); Shorrocks (1995) modified-Sen; Thon (1979), but only for the limiting form as $n \rightarrow \infty$; and Watts (1968).

- Jenkins, S., & Lambert, P. (1997). Three 'I's of poverty curves, with an analysis of U.K. poverty trends. *Oxford Economic Papers*, 49, 317–327.
- Johnson, D. T. (1988). The measurement of poverty in Australia: 1981–82 and 1985–86. *Australian Economic Review*, 3rd quarter, 13–24.
- Katz, S. (1999). The self-similar science system. *Research Policy*, 28, 501–517.
- Katz, S. (2000). Scale-independent indicators and research evaluation. *Science and Public Policy*, 27, 23–36.
- Kinney, A. L. (2007). National scientific facilities and their science impact on non-biomedical research. *Proceedings of the National Academy of Sciences*, 104, 17943–17947.
- Marchant, T. (2009). An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors. *Scientometrics*, 80, 325–342.
- Martin, B., & Irvine, J. (1983). Assessing basic research. Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12, 61–90.
- Moed, H. F., Burger, W. J., Frankfort, J. G., & van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14, 131–149.
- Moed, H. F., & van Raan, A. F. J. (1988). Indicators of research performance. In A. F. J. van Raan (Ed.), *Handbook of quantitative studies of science and technology* (pp. 177–192). North Holland.
- Moed, H. F., De Bruin, R. E., & van Leeuwen, Th. N. (1995). New bibliometrics tools for the assessment of national research performance: Database description, overview of indicators, and first applications. *Scientometrics*, 33, 381–422.
- Molinari, J. F., & Molinari, A. (2008). A new methodology for ranking scientific institutions. *Scientometrics*, 75, 163–174.
- Molinari, J. F., & Molinari, A. (2008). Mathematical aspects of a new criterion for ranking scientific institutions. *Scientometrics*, 75, 339–356.
- Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing and Management*, 12, 297–312.
- Pyatt, G. (1987). Measuring welfare, poverty, and inequality. *Economic Journal*, 97, 459–467.
- Quesada, A. (2009). Monotonicity and the Hirsch Index. *Journal of Informetrics*, 3, 158–160.
- Quesada, A. (2010). More axiomatics for the Hirsch Index. *Scientometrics*, 82, 413–418.
- Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9, 281–291.
- Schubert, A., Glänzel, W., & Braun, T. (1983). Relative citation rate: A new indicator for measuring the impact of publications. In D. Tomov, & L. Dimitrova (Eds.), *Proceedings of the first national conference with international participation in scientometrics and linguistics of scientific text*, Varna.
- Schubert, A., Glänzel, W., & Braun, T. (1987). A new methodology for ranking scientific institutions. *Scientometrics*, 12, 267–292.
- Schubert, A., Glänzel, W., & Braun, T. (1988). Against absolute methods: Relative scientometric indicators and relational charts as evaluation tools. In A. F. J. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology* (pp. 137–176).
- Seglen, P. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43, 628–638.
- Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, 44, 219–230.
- Sen, A. (1992). *Inequality reexamined*. Cambridge: Harvard University Press.
- Shorrocks, A. (1995). Revising the Sen's poverty index. *Econometrica*, 63, 1225–1230.
- Tijssen, R. M., Visser, M., & van Leeuwen, T. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference. *Scientometrics*, 54, 381–397.
- Tijssen, R. M. (2003). Scoreboards of research excellence. *Research Evaluation*, 12, 91–103.
- Thon, D. (1979). On measuring poverty. *Review of Income and Wealth*, 25, 114–126.
- Thon, D. (1983). A note on a troublesome axiom for poverty indices. *Economic Journal*, 93, 199–200.
- Van Leeuwen, T., Visser, M., Moed, H., Nederhof, T., & van Raan, A. (2003). The holy grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, 57, 257–280.
- Van Raan, A. F. J. (2004). Measuring science. In H. F. Moed, H. F. Moed, et al. (Eds.), *Handbook of quantitative science and technology research* (pp. 19–50).
- Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics*, 10, 157–177.
- Vinkler, P. (2003). Relations of relative scientometric indicators. *Scientometrics*, 58, 687–694.
- Watts, H., & Moynihan, D. P. (1968). An economic definition of poverty. In *Understanding poverty: Perspectives from the social sciences*. New York: Basic Books.
- Woeginger, G. (2008a). An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, 56, 224–232.
- Woeginger, G. (2008b). A symmetry axiom for scientific impact indices. *Journal of Informetrics*, 2, 298–303.
- Zheng, B. (1997). Aggregate poverty measures. *Journal of Economic Surveys*, 11, 123–162.