Journal of Informetrics 11 (2017) 689–703

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Measuring the citation impact of journals with generalized Lorenz curves

Tommaso Lando[a,b], Lucio Bertoli-Barsotti[a,*]

[a] Department of Management, Economics and Quantitative Methods, University of Bergamo, via dei Caniana 2, 24127, Bergamo, Italy
[b] Department of Finance, VŠB-TU Ostrava, Sokolskà Trida 33, Ostrava, Czech Republic

ARTICLE INFO

Article history:
Received 7 September 2016
Received in revised form 23 February 2017
Accepted 16 May 2017
Available online 20 June 2017

Keywords:
Journal impact factor
Stochastic dominance
Quantile function
Lorenz curve

ABSTRACT

To improve comparisons of journals, which are typically based on single-value indicators, such as the journal impact factor (JIF), this paper proposes a *functional* approach. We discuss interpretatively three progressively finer dominance relations. The first one corresponds to a comparison between the *quantile functions* of the citation distributions. The second one consists in comparing the integrals of the quantile functions—namely, the *generalized Lorenz curves* (GLCs). The third one consists in comparing the integrals of the GLCs, where the integration is designed to emphasize the role of the "central body" of the articles of the journal. Although dominance relations are generally not complete orders, we demonstrate with an empirical analysis that it is possible to increase significantly the proportion of pairs of journals that are comparable by moving from the first to the second criterion, and then from the second to the third.

Because, in practical applications, it may be convenient to reduce such a functional comparison to a scalar comparison between indicators, we follow an axiomatic approach to identify classes of indicators that are *isotonic* with the criteria introduced. We demonstrate that the established JIF may be usefully improved if it is corrected simply by multiplying it by one minus the Gini coefficient. The resulting index, defined as *stabilized-JIF*, has many attractive features and it is isotonic with all the dominance relations introduced.

1. Introduction

Research evaluation is a topic whose theoretical importance is widely recognized by the scientific community, and it has several practical implications of great interest for the policy and orientation of scientific research. In this present context, the ranking of scientific journals is a major issue. Authors and institutions are interested in quantifying the "impact" of a journal on the scientific community, and the most widely used impact measures are based on citation data. On one hand, authors generally aim to identify the journals that may provide the largest audience and, hence, (possibly) the highest number of citations of their papers. On the other hand, researchers or research institutions may be directly rewarded for publishing in highly ranked journals.

The indicator most widely used to evaluate the impact of journals is based on the average number of citations per paper. This simple indicator, generally referred to as the *journal impact factor* (JIF), is ascribable to Garfield and Sher (1963) (see also Garfield, 1972; Garfield, 2006). The JIF, basically the mean citedness, is not uniquely defined insofar as it may vary according

* Corresponding author.
E-mail address: lucio.bertoli-barsotti@unibg.it (L. Bertoli-Barsotti).

http://dx.doi.org/10.1016/j.joi.2017.05.005
1751-1577/

to different citation databases (e.g., Web of Science, Scopus), different publication and citation time windows; the JIF may also depend on the degree of overlap between these timeframes (leading to so-called synchronous and diachronous impact factors; see Bar-Ilan, 2010; Ingwersen, Larsen, Rousseau, & Davis, 2001; Ingwersen, 2012) and different document types to be considered as citing or citable items for the citation count (e.g., all documents versus articles alone versus reviews and articles alone, etc.). For this reason, the JIF is also referred to hereafter using the plural "journal impact factors" (JIFs).

The advantages and disadvantages of JIFs have been discussed from different perspectives, and many variants or adjustments have been proposed in order to account for several limitations, such as the lack of statistical significance (Vanclay, 2012; Stern, 2013), insensitivity to field differences (Moed, 2010), insensitivity to the "weight" of the citing articles (Ferrer-Sapena, Sánchez-Pérez, González, Peset, & Aleixandre-Benavent, 2015) and manipulability by editorial strategies (Moustafa, 2015), among others. In this paper, we do not consider these more advanced issues; rather, we focus on two related problems:

- poor representativeness of the citation distribution: the JIF alone says little about the shape of the citation distribution—i.e., the mean is not suitable to represent highly skewed distributions (see, e.g., Seglen, 1997); and
- poor robustness: the JIF may be strongly influenced by one highly cited publication (e.g., Foo, 2013, shows how one single article increased the JIF of *Acta Cristallographica A* from 2 to 50 in 2009) or by a few (Editorial, 2005).

In particular, the JIF is often used (in a misleading way) to approximate the actual number of citations a paper might receive, although it is well documented in the literature that citation distributions generally show quite high concentration patterns, i.e., a few articles account for most of the citations, whilst most articles produce zero or only a few citations (see, e.g., Laband, 1986; Stern, 2013). This suggests that the issue of measuring citation impact should be related not only to the JIF but also to the measurement of inequality, or concentration, as we shall propose below.

Many authors agree that indicators of impact more representative and robust than the JIF should be employed. For instance, one may use the mode (Vanclay, 2012), the median (Wall, 2009) or another quantile of the distribution (Bornmann, Leydesdorff, & Mutz, 2013; Leydesdorff, 2012), although these indices have inferior discriminating power and may yield large numbers of ties. Alternatively, one can employ a trimmed mean (Seiler & Wohlrabe, 2014), a geometric mean corrected for uncited items (Zitt, 2012; Thelwall & Fairclough, 2015), or more generally (as we shall discuss in Section 3), a power mean of order a, where $0 < a < 1$. In all of these cases, the effect is that of "downsizing" the role of highly cited items and obtaining a more robust indicator of impact. In a more general framework, as discussed by Bouyssou and Marchant (2011), a family of generalized JIFs may be defined on the basis of the mean of an increasing function u of citations, where we may emphasize or downsize extreme values if we, respectively, choose u to be convex or concave.

The approach of Bouyssou and Marchant (2011) is, in turn, related to the concept of *stochastic dominance* for ranking citation distributions. In statistics, stochastic dominance relations establish preorders in the space of distribution functions that quantify the idea of one distribution being "preferable" to another (see, e.g., Marshall, Olkin, & Arnold, 2011). Such ranking criteria consist in a functional comparison that is generally much stronger (although not always verifiable) than a simple condition on a single-valued parameter (e.g., the mean) and leave very little room for ambiguity.

In a bibliometric context, the use of dominance rules for comparison of citation data has been proposed by Carayol and Lahatte (2009) and briefly discussed by Bouyssou and Marchant (2011) and Waltman and Van Eck (2009). In the paper, we develop and widen this approach, at both theoretical and applied levels. From a methodological point of view, it should be stressed that all of the aforementioned authors have used classic dominance relations—namely, first-order and second-order stochastic dominance, which are based on the distribution function and its integral, respectively. In contrast, we demonstrate that it is definitely more advantageous, in terms of ease of interpretation and computation, to express dominance relations by using the *quantile function* and, in particular, its integral—namely, the *generalized Lorenz curve* (*GLC*, Shorrocks, 1983). The GLC provides an attractive representation of the overall impact as well as the shape of the citation distribution. For a given journal with T publications, the GLC evaluated in $p \in [0, 1]$ determines an average of the quantiles, defined as the *partial JIF of order p*, i.e., the JIF corresponding to "the set of the 100p% less-cited papers", as will be shown in Section 2.3. Put another way, the GLC represents the "distribution" of the JIF within the papers of a journal.

It is well known that stochastic dominance relations are *preorders* (see Section 2) and, in particular, that they are not *total* (*complete*), because one may find pairs of distributions (journals) that cannot be ranked. In this case, it is possible to introduce some *finer* (or *weaker*, see Section 2) criteria that conform with our preferences and increase the number of comparable pairs of journals. In Sections 2.2 and 2.3, we analyse some *strong* preorders—namely, 1) *first order stochastic dominance (1-SD)*, which requires each quantile (i.e., generally, the citations of the Tp-th ranked paper) of the dominant journal to be higher compared to that of the dominated one; and 2) the *generalized Lorenz dominance (GLD)*, which basically requires that the condition for 1-SD holds *on average* or, in other words, that the "distribution of the JIF" (i.e., the GLC) of the dominant journal is uniformly higher. However, real data comparisons show that 1-SD and GLD are rarely verified.

Therefore, in Section 2.4, we introduce a new dominance relation for measuring the impact of journals that emphasizes the "body" of the citation distribution—namely, the *second-order outward generalized Lorenz dominance (2-OGLD)*. This is accomplished by cumulating the GLC, or "averaging" the values of the partial JIFs (as will become clear in the following explanation), from the "centre". This approach may serve a twofold objective: i) to rank the pairs of journals that are not ranked by 1-SD and GLD; and ii) to reward journals whose citations are mainly concentrated in the body, rather than in the tails, according to the principle that tails do not provide a good representation of the impact of a journal. Notably, the 2-OGLD does not require the dominant journal to have greater JIF, and it is especially suitable for ranking intersecting GLCs.

Finally, the main advantage of a dominance-based approach is that, once we have identified the dominance relation most suitable for our purposes, we can obtain several classes of impact measures that are *isotonic* (or *order preserving*, Marshall et al., 2011) with this relation. The use of any of these measures is clearly sufficient to obtain unambiguous rankings.

To compensate for the aforementioned theoretical weaknesses of the JIF, we also study some alternative "finer" measures of impact (to be understood hereafter as indicators that are isotonic with finer dominance relations). In Section 3, we present some classes of indicators that are isotonic with the dominance relations defined in Section 2. In particular, we focus on an alternative index of impact, defined as *stabilized-JIF* (*s-JIF*), that is given by the JIF multiplied by one minus the Gini coefficient. In addition to being very easy to compute, the s-JIF provides a powerful graphical representation of citation impact, based on the area under the GLC, and it can be interpreted as an "average of partial JIFs" (see Section 3.2).

To test the usefulness and applicability of our theoretical approach, in Section 4, we perform an empirical analysis of 100 economics journals. First, we investigate the percentage of journal pairs that can be ranked according to the different preorders (with similar studies having been performed by Bouyssou and Marchant, 2011, with regard to a small dataset of eight journals, and by Carayol and Lahatte, 2009, with regard to French universities). Thereafter, by means of a thorough pairwise comparison, we analyse the coherence of the JIF with the dominance relations proposed.

2. Ranking journals on the basis of dominance relations

2.1. Preliminaries

In what follows, we shall be concerned with the issue of ranking scientific journals on the basis of dominance rules, i.e., preorders in the space of citation distributions. We recall that a preorder is a binary relation $\prec$ over a set S that is reflexive and transitive. In particular, observe that a preorder $\prec$ does not generally satisfy the antisymmetry property (that is, $a \prec b$ and $b \prec a$ does not necessarily imply $a = b$) and it is generally not total (that is, each pair a, b in S is not necessarily related by $\prec$). We say that the preorder $\prec$ is *stronger* than the preorder $\prec^*$ (or, equivalently, $\prec^*$ is *weaker,* or *finer,* than $\prec$) if, for every pair a, b in S, $a \prec b$ implies $a \prec^* b$.

Let us characterize a journal with T publications (items) by the *citation distribution* $N : \mathbb{R} \to [0, T]$, to be understood as a *size-frequency function* (Egghe, 2005), where $N(x)$ is the number of papers which have been cited x times at most, $x \in \mathbb{R}$. Then, in particular, $n(k) = N(k) - N(k-1)$, for $k = 1, 2, \ldots, n(0) = N(0)$, is the number of papers that have been cited exactly k times. Denote by $\mu = \frac{1}{T}\sum_{k} kn(k)$ the mean of the distribution N. Note that one can also write $\mu = \frac{1}{T}\int_0^\infty x dN(x)$ using the Lebesgue-Stieltjes notation. For the sake of simplicity, we henceforth assume that one particular database—with specified time windows for publication and citation data—is used and that the JIF is calculated on the basis of the publication and citation counts provided by this database. By so doing, we identify the mean μ with the JIF of the journal.

Let us introduce the *normalized citation distribution* (NCD) $F(x) = \frac{N(x)}{T}$, which gives the proportion of papers that have been cited x times at most, for every $x \in \mathbb{R}$. Similarly, $f(k) = \frac{n(k)}{T}$, for $k = 0, 1, \ldots$, gives the proportion of papers that have been cited exactly k times. Let us denote by $\mathcal{F}$ the set of all NCDs. Notably, the function $F(x)$ may also be interpreted as the cumulative distribution function of the random variable X = number of citations of a randomly chosen article from a given journal. This will enable us to employ hereinafter the terminology and notation commonly used in the theory of (discrete) probability distributions—e.g., the notions of first- and second-order stochastic dominance—without further distinguishing between our deterministic setting and a purely probabilistic setting.

Because, in this paper, we are mainly concerned with the issue of measuring the *per paper* impact, rather than the productivity, of a journal, we shall especially focus on the space $\mathcal{F}$, overlooking the total number of publications in most cases. Note that, if journal B is obtained by multiplying the number of papers of journal A for a fixed positive integer m, that is, $N_B(x) = mN_A(x)$, we obtain the result $F_B(x) = F_A(x)$. Thus, by considering NCDs, we take into account the fundamental property of a JIF generally referred to as *size-independence* (Garfield, 2006) or *homogeneity* (within the axiomatic framework of Bouyssou & Marchant, 2011), which makes it possible to compare journals with different numbers of publications.

In the next subsections, besides providing some basic definitions, we present three different approaches to ranking citation distributions by means of dominance relations of progressively higher orders.

2.2. First-order comparison

If two journals, A and B, have an equal number of publications T, we argue that A is definitely preferable to B in terms of impact if the citations obtained by the i-th ranked item of A (for i = 1., T) are never less than the citations obtained by the i-th item of B (where items are clearly ranked according to the number of citations). Apparently, this criterion is far from being applicable, mainly because it is difficult to find pairs of journals with the same numbers of publications. Nevertheless, this limitation is bypassed by i) focusing on the space $\mathcal{F}$ of NCDs and ii) comparing the corresponding *quantiles*. By so doing, the intuitive ranking criterion explained above can be extended to a more applicable framework, relaxing the assumption of equal numbers of items. For this purpose, we shall need the definition of *quantile function*.

Definition 1. Let $F \in \mathcal{F}$. The left-continuous inverse, or *quantile function*, of F is defined as

$$Q(p) = \inf\left\{x : F(x) \geq p\right\}, 0 < p \leq 1$$

$$Q(0) = \inf\left\{x : F(x) > 0\right\}.$$

Note that $Q(0) = \min\{x : n(x) > 0\}$ and $Q(1) = \max\{x : n(x) > 0\}$ give, respectively, the minimum and maximum numbers of citations obtained by papers in journal A.

The quantile function is a non-decreasing and left-continuous step function in [0, 1], which gives the quantile of order p, that is, a threshold value of citations $Q(p)$ such that the proportion of items cited at most $Q(p)$ times is greater than (or equal to) p. Stated otherwise, $Q(p)$ gives the number of citations of the item whose rank is nearest to (but not less than) Tp. We observe in passing that there is no universal agreement in the literature on how to compute quantiles for discrete distributions. For a thorough discussion of some alternative methods and their application to citation data, the reader is referred to Bornmann et al. (2013).

An unequivocal order of preference can be defined on the basis of a pairwise comparison of quantiles.

Definition 2. Let journals A and B have NCDs given by F_A and F_B. We say that F_A dominates F_B with respect to the 1-SD, and write $F_A \geq_1 F_B$, if and only if $F_A(x) \leq F_B(x), \forall x \in \mathbb{R}$ or, equivalently, $Q_A(p) \geq Q_B(p), \forall p \in [0, 1]$.

It is convenient to note that it is possible to write the mean μ as the integral of the step function $1 - F(x)$ (see, e.g., Billingsley, 1986, p.74), i.e., $\mu = \int_0^\infty (1 - F(x))\,dx$. Then, by construction, it is also possible to write $\mu = \int_0^1 Q(p)\,dp$.

2.3. Second-order comparison: generalized Lorenz dominance

As a general rule, the stronger the preorder, the less it is applicable, because we may find, in practice, many pairs of intersecting distributions that are not comparable at the first order. Hence, we consider a finer ranking criterion based on the GLC, which is the integral of the quantile function. As will be shown, the GLC, introduced by Shorrocks (1983) and referred to as the *absolute Lorenz curve* (Yitzhaki & Schechtman, 2012), has many attractive properties for the representation and measurement of citation impact. Similarly, we recall that the use of the *Lorenz curve* (LC, i.e., the GLC divided by the mean) as a tool for the graphical representation of inequality/concentration patterns of informetric data is not new (see, e.g., Burrell, 2005).

Definition 3. Let $F \in \mathscr{F}$. The *generalized Lorenz curve* $L : [0, 1] \to [0, \mu]$ of F is defined as

$$L(p) = \int_0^p Q(t)\,dt, p \in (0, 1),$$

$$L(0) = 0, L(1) = \mu.$$

(Note: With a little abuse of notation, in this paper, we denote the GLC with the symbol "L", which is generally used to indicate the classic LC.)

By construction, the GLC is a piecewise linear function, strictly increasing and convex in [0, 1]. Note that the GLC completely determines the citation distribution F, because $L'(p) = F^{-1}(p)$ almost everywhere or equivalently, $Q(p) = \lim_{t \to p^-} L'(t)$ $\forall p$ in$[0, 1)$.

Now, because $\mu = \int_0^1 Q(p)\,dp$, we can interpret the GLC curve as an "incomplete (lower) mean". Indeed, after normalization, the ratio $L(p)/p$ represents the *partial (conditional) JIF*, defined as

$$\mu_\uparrow(p) = L(p)/p, p \in (0, 1], \quad (1)$$

over "the set of the 100p% less-cited papers". Strictly speaking, this set is not always well defined (insofar as the cardinality of the set of 100p% less-cited papers is not necessarily an integer number), but formula (1) can still be reasonably interpreted. Indeed, let $x_{(1)}, x_{(2)}, \ldots, x_{(T)}$ be the citation counts of the papers of a given journal, sorted from least to greatest. Two cases are possible: 1) $p = \frac{k}{T}$, for some k, $k = 1, \ldots, T$; and 2) $p \neq \frac{k}{T}$, for every k, $k = 0, 1, \ldots, T$. In the former case, $k = pT$ is an integer value which determines the exact p-quantile rank. Hence, we obtain the partial JIF of order p—that is, the JIF of the set of the 100p% less-cited papers—

$$\mu_\uparrow(p) = \frac{x_{(1)} + x_{(2)} + \ldots + x_{(k)}}{k}. \quad (2)$$

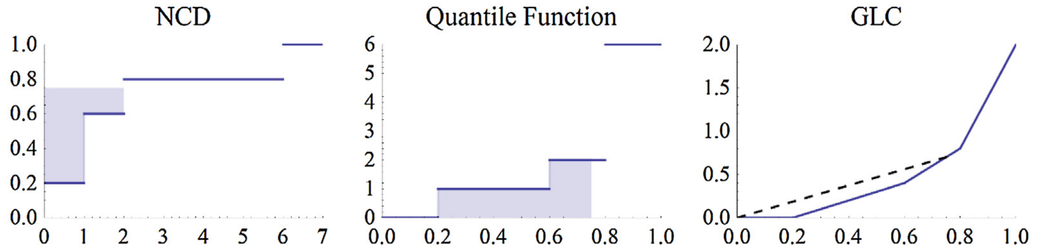

Fig. 1. NCD, Quantile function and GLC. The curves correspond to the dataset described in Example 1. The GLC evaluated in 0.75 is equivalently represented by the filled areas in the first two graphs. The slope of the dashed segment in the third graph (connecting (0, 0) and (0.75, $L(0.75)$)) represents the partial JIF of order 0.75.

In the latter case, by construction, there exists k^*, $k^* = 1, 2, \ldots, T-1$, such that $\frac{k^*}{T} < p < \frac{k^*+1}{T}$, where $\mu_\uparrow(p)$ is an interpolation between the exact values of $\mu_\uparrow\left(\frac{k^*}{T}\right)$ and $\mu_\uparrow\left(\frac{k^*+1}{T}\right)$.

Example 1

$$\text{Let} T = 5, \text{with} x_{(1)} = 0, x_{(2)} = 1, x_{(3)} = 1, x_{(4)} = 2, x_{(5)} = 6 .$$

For $p = 0.4$, we obtain $pT = 2$. The two less cited papers have citations of 0, 1. Then, because (see Fig. 1)

$$L(0.4) = \int_0^{0.4} Q(t)\,dt = 0.2,$$

the partial JIF is $\frac{L(0.4)}{0.4} = 0.5 = \mu_\uparrow(0.4) = \frac{0+1}{2}$.

Instead, for p=0.75, we find k^*=3 for which 3<pT=3.75<4. Then, because $L(0.6) = \int_0^{0.6} Q(t)\,dt = 0.4$ and $L(0.8) = \int_0^{0.8} Q(t)\,dt = 0.8$, we obtain

$$\mu_\uparrow(0.75) = \frac{1}{0.75} \cdot \left(L(0.6) + (0.75 - 0.6)\frac{L(0.8) - L(0.6)}{0.2}\right)$$

$$= \frac{1}{0.75} \cdot \left(0.4 + (0.75 - 0.6)\frac{0.8 - 0.4}{0.2}\right) = 0.93 .$$

We observe (see Fig. 1) that $\mu_\uparrow(p)$ can be represented graphically by the slope of the segment connecting (0, 0) and $(p, L(p))$. Also note that, when all items are equally cited, $L(p) = \mu p$ (i.e., $\mu = \mu_\uparrow(p), \forall p$).

The GLC can be used to rank citation distributions of journals according to the *generalized Lorenz dominance* (GLD), defined as follows.

Definition 4. Let journals A and B have NCDs given by F_A, F_B. We say that F_A dominates F_B w.r.t. the GLD and write $F_A \geq_L F_B$, if and only if $L_A(p) \geq L_B(p), \forall p \in [0, 1]$.

Note that the GLD, which is a term used in the economic literature (see, e.g., Zoli, 1999; Zoli, 2002) is often referred to as the *second-order inverse stochastic dominance*, which, in turn, is well known to be mathematically equivalent (Muliere & Scarsini, 1989) to the *second-order stochastic dominance*, already proposed in the bibliometric literature, for comparisons of journals, by Carayol and Lahatte (2009) and Bouyssou and Marchant (2011).

The basic logic of the GLD is quite simple. As the 1-SD is too restrictive (and not verified in most cases), we may require that citations of journal A are higher than those of journal B *on average,* since the GLC can be seen as a partial average of the quantiles (see Eq. (1)). Indeed, $F_A \geq_L F_B$ means that the JIF of A is uniformly greater than (or equal to) the JIF of B for every set of 100p% less-cited papers. If this relation holds, we argue that journal A definitely manifests greater impact than B. Thus, its coherence with the GLD should be a fundamental property of an index of impact.

The GLD basically rewards overall impact as well as regularity, or consistency. On one hand, a necessary condition for the GLD is that the dominant distribution must have a JIF greater than (or equal to) the dominated one: that is, $F_A \geq_L F_B$ implies $L_A(1) \geq L_B(1)$. On the other hand, the GLD reflects its preference for those distributions that exhibit lesser degrees of inequality. This is easy to understand when the journals have equal JIFs because, in this case, the GLD is equivalent to the classical *Lorenz dominance* (LD; see e.g., Aaberge, 2009, we also refer to the *Lorenz order*, which is equivalent but in the opposite direction (see, e.g., Marshall et al., 2011), a tool widely used in economics to rank income distributions in terms of inequality.

2.4. Third-order comparison: the second-order outward GLD

Strong preorders, such as 1-SD and GLD, provide a normative justification for the use of the JIF, provided that they are verified, because they require the dominant distribution to have a greater mean. However, it may frequently occur that the GLCs intersect; thus, the GLD (as well as, consequently, 1-SD) is not verified, as will be demonstrated in Section 3. Thus, ranking journals based on the JIF can often be inadequate—that is, not supported by a thorough comparison of distributions. Nevertheless, to obtain a broader set of comparable pairs of journals, we may introduce a finer dominance criterion that conforms to our preferences

In the bibliometric literature, Carayol and Lahatte (2009) have proposed ranking universities on the basis of *upward dominance* relations, which basically consist of checking the conditions for 2-SD by focusing on a restricted set of the $q100\%$ top-cited items $(0 < q < 1)$. We argue that this approach is not suitable in our context (i.e., ranking journals). First, the choice of q may be tricky. Moreover, by restricting the set of papers considered, the method is weakened in terms of robustness because it would further emphasize the effect of a few highly cited papers. Hence, this approach is not consistent with our objectives. Rather, a more thorough method consists of moving to a dominance relation of higher order by cumulating the GLC. As has been shown in Section 2.3, the GLC—i.e., the integral of the quantile function—is an average of the quantiles or, equivalently, a partial average (JIF). Similarly, by integrating the GLC, we average the partial JIFs (see also Section 3.2).

The idea of integrating the GLC, or the LC, is widely used in the economics literature, in particular with regard to the measurement of income inequality, where different orderings have been proposed to address the issue of how to rank intersecting GLCs, or LCs (see, e.g., Aaberge, 2009; Chiu, 2007; Davies & Hoy, 1995; Muliere & Scarsini, 1989; Shorrocks & Foster, 1987). Such a method yields a dominance relation of a higher degree, which may be equivalently termed "third-order inverse stochastic dominance" (Muliere & Scarsini, 1989) or "second-order GLD" (Aaberge, 2009). Aaberge (2009) proposes cumulating LCs from the bottom or, alternatively, from the top (namely, the *downward* or *upward second-order Lorenz dominance*, respectively). Mathematically, by cumulating the GLC from the bottom, one places progressively more emphasis on the lower part of the distribution, whilst, by cumulating it from the top, one obtains the opposite effect (see Muliere & Scarsini, 1989 and Aaberge, 2009). This may be made more clear by the following example. Let A and B two journals with the same JIF μ and with GLCs $L_A(p) = \mu p^2$ and $L_B(p) = \mu\left(1-(1-p)^{1/2}\right)$. L_A and L_B are single-crossing; L_A concentrates its mass (to be understood as the area under the GLC) on the right tail, whilst L_B concentrates its mass on the left one. When integrating from the left to the right, one obtains $\int_0^t L_A(p)\,dp \le \int_0^t L_B(p)\,dp, \forall t \in [0, 1]$, which means that L_B second-order upward dominates L_A (Aaberge, 2009). Conversely, when integrating from the right to the left, one obtains $\int_t^1 L_A(p)\,dp \ge \int_t^1 L_B(p)\,dp, \forall t \in [0, 1]$, which means that L_A second-order downward dominates L_B. Hence, downward and upward second-order Lorenz dominance may lead to opposite rankings. Indeed, by performing an integration, we attach more weighting to the left (or right) tail of the distribution and, simultaneously, we downsize the right (or left) one. Such approach definitely makes sense in an economic context, where one is interested especially in those variations that occur at the bottom (or sometimes at the top) of the income scale. Conversely, in a bibliometric context, it would be preferable to define an ordering that emphasizes the "body" of the citation distribution, rather than just one of the two tails. This conforms with the idea that, symmetrically, tails are not truly representative of the impact of a journal. In other words, when a paper obtains a very high (or low) number of citations, compared to others, it is reasonable to assume that it is mainly the authors' responsibility, and the journal's ranking should not be unduly influenced by that (e.g., a few top papers that amass outstandingly high numbers of citation compared to others should be mainly credited to authors, rather than to the journal). Hence, the right and left tails of the GLC may represent the overall journal's impact poorly, because they may be influenced, respectively, by i) a few highly cited papers or ii) a few uncited ones. Conversely, the "central" values of the GLC are generally more robust with respect to situations i) and ii), such that it seems to be a clever solution to define a ranking criterion that emphasizes such values and downsizes the tails' values. To formalize such concept, we propose a new dominance relation, namely, the *second-order outward generalized Lorenz dominance* (*2-OGLD*), which consists of cumulating the GLC from the centre to the tails (i.e., from the "inside" to the "outside"). By so doing, as we shall discuss also in Section 3.2 (see Eq. (8)), we basically average the partial JIFs, starting from the central values. For the sake of simplicity, we identify the "centre" of the GLC with the value $p = 0.5$. Note that the 2-OGLD is related to the *disparity dominance* proposed by Lando and Bertoli-Barsotti (2016). This new approach is complementary with respect to those represented by the so-called *k-th order inverse stochastic dominance* (Muliere & Scarsini, 1989) or by the second-order Lorenz dominance (upward or downward) of Aaberge (2009), which emphasize one of the two tails.

Definition 5. Let journals A and B have NCDs given by F_A, F_B. We say that F_A dominates F_B w.r.t. the 2-OGLD and write $F_A \ge_{OL}^2 F_B$, if and only if

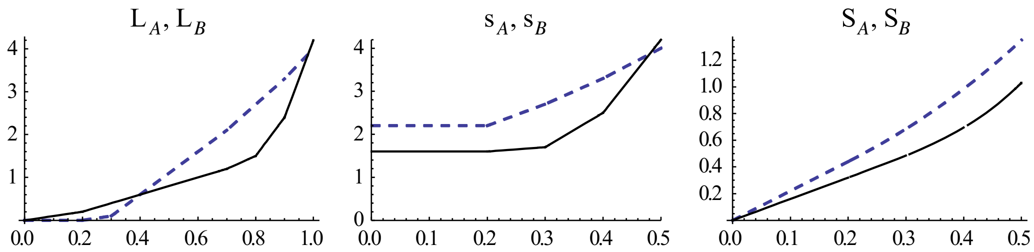

Fig. 2. GLCs, SGLCs and cumulated SGLCs. The curves correspond to the dataset of Example 2 (dashed for journal A, solid for journal B).

$$S_A(p) = \int_0^p s_A(t)\,dt \geq \int_0^p s_B(t)\,dt = S_B(p),\ \forall p \in [0, 0.5],$$

where

$$s(p) = L(0.5+p) + L(0.5-p),\ 0 \leq p \leq 0.5.$$

In particular, drawing inspiration from the statistical literature (see, e.g., Balanda & MacGillivray, 1990) we denote the curve $s(p)$ as the *(outward) spread function of the GLC (SGLC)*. It can be easily seen that $s(p)$ is a non-negative, continuous, increasing and convex function for $0 \leq p \leq 0.5$, where $s(0) = 2L(0.5) = \mu_{\uparrow}(0.5)$ and $s(0.5) = \mu$.

Moreover, Eq. (1) yields

$$s(p) = (0.5+p)\,\mu_{\uparrow}(0.5+p) + (0.5-p)\,\mu_{\uparrow}(0.5-p). \quad (3)$$

Hence, the SGLC is a weighted average of partial JIFs, symmetrically spaced with respect to 0.5, whilst its integral, $S(p)$, is a weighted average of all partial JIFs between $0.5 - p$ and $0.5 + p$.

The SGLC captures the "magnitude" of the GLC and maintains sensitivity to some specific features related to concentration. Therefore, the 2-OGLD makes it possible to rank intersecting GLCs by further emphasizing citations in the body, as well as reducing the effect of citations in the tails. This might lead us to prefer a journal that has slightly inferior JIF but shows much more consistency throughout its citation distribution, as shown in the following example (and in Section 4 as well).

Example 2. Let us consider journals A and B with $T_A = T_B = 10$. Papers in A have been cited 7, 6, 6, 5, 5, 5, 5, 1, 0, 0 times ($\mu_A = 4$), whilst papers in B obtained 18, 9, 3, 2, 2, 2, 2, 2, 1, 1 citations ($\mu_B = 4.2$). Citations of A are concentrated in the body of the distribution, whilst citations of B are more concentrated in the two tails. Indeed, the partial JIF of A is higher than that of B, from 0.4 to 0.98. Fig. 2 shows that L_A, L_B cross twice and that s_A and s_B cross once (close to 0.5), whilst S_A is uniformly higher than S_B because of the greater weight attached to the citations of the central papers—thus, $F_A \geq^2_{OL} F_B$. We also observe that $\mu_A = 4 < 4.2 = \mu_B$, whilst $\bar{\mu}_A = 2.7 > 2.06 = \bar{\mu}_B$.

Our empirical analysis in Section 4 will show that the 2-OGLD is especially suitable for ranking intersecting GLCs.

3. Isotonic indicators

3.1. Some classes of indicators

In the mathematics literature, many well-known results of majorization theory make it possible to determine functionals that are said to be *isotonic*, *consistent* or *order-preserving* with a given preorder. The reader is mainly referred to Marshall et al. (2011) for a detailed analysis of majorization and its applications. The following theorem consists of a different formulation of some important results of majorization theory (see, e.g., Marshall et al., 2011, Proposition B.19.c, p.710). This alternative formulation is necessary in our case because we need to apply the following results to functions that are not necessarily distribution functions, as we shall see below.

Theorem 1. *Let H, K be non-decreasing, non-negative, left-continuous and integrable functions defined on a bounded set [0,a] (0< a < ∞).*

1. $H(p) \geq K(p),\ \forall p \in [0, a]$ iff $\int_0^a u(H(p))\,dp \geq \int_0^a u(K(p))\,dp$ for every non-decreasing function u (such that the integrals exist).

2. $\int_0^p H(t)\,dt \geq \int_0^p K(t)\,dt, \forall p \in [0, a]$ iff $\int_0^a u(H(p))\,dp \geq \int_0^a u(K(p))\,dp$ for every non-decreasing and concave function u (such that the integrals exist).

Proof. Without loss of generality, we assume $a = 1$. Let us introduce the right-continuous inverse functions $\tilde{H}$ and $\tilde{K}$:

$$\tilde{H}(z) = \sup\{p \in [0, a] : H(p) \leq z\}, -\infty < z < \infty,$$

with the convention $\sup(\phi) = 0$ (and similarly for g). Note that $\tilde{H}$ and $\tilde{K}$ are probability (for $a = 1$) distribution functions (i.e., having bounded variation and being non-decreasing and right-continuous). The function $H(p)$ can be equivalently expressed as $H(p) = \inf\left\{z : H(p) \leq z\right\}$, where $H(p) \leq z$ implies that $\tilde{H}[H(p)] \leq \tilde{H}(z)$ (because $\tilde{H}$ is non-decreasing). By construction, $\tilde{H}(z) \geq p$ iff $H(p) \leq z$ and, in particular, $\tilde{H}[H(p)] \geq p$, which yields $H(p) = \inf\{z : p \leq \tilde{H}(z)\}$. This result shows that H is the (left-continuous) quantile function of $\tilde{H}$ (and, similarly, K is the quantile function of $\tilde{K}$).

1. By applying the formula of integration by substitution (Hoffman-Jorgensen, 1994, p. 205), the statement can be equivalently expressed as $\int u(z)d\tilde{H}(z) \geq \int u(z)d\tilde{K}(z)$, iff $H(p) \geq K(p), \forall p \in [0, a]$), where the latter is clearly equivalent to $\tilde{H}(z) \leq \tilde{K}(z), \forall z \in \mathbb{R}$. Hence, the proof can be straightforwardly derived from the classic characterization theorem of 1-SD (see, e.g., Marshall et al., 2011, Proposition B.19.c, p.710).
2. Using integration by substitution, the statement becomes $\int u(z)d\tilde{H}(z) \geq \int u(z)d\tilde{K}(z)$, iff $\int_0^p H(t)\,dt \geq \int_0^p K(t)\,dt, \forall p \in [0, a]$,where the latter is equivalent to $\int_{-\infty}^z \tilde{H}(t)dt \leq \int_{-\infty}^z \tilde{K}(t)dt, \forall z \in \mathbb{R}$ (Muliere & Scarsini, 1989). Hence, the proof can be straightforwardly derived from the classic characterization theorem of 2-SD (see, e.g., Marshall et al., 2011, Proposition B.19.c, p.710). ■

Given the preorder $\prec$ (defined over a set S), a mapping $M : S \to \mathbb{R}$ is said to be *isotonic* with respect to < whenever, for every $a, b \in S$ such that $a \prec b$, we obtain $M(a) \leq M(b)$. Clearly, if the preorder $\prec^*$ is finer than the preorder $\prec$, we find that every mapping that is isotonic with $\prec^*$ is isotonic with $\prec$ as well. Moreover, we say that M is *strictly isotonic* with $\prec$ if $a \prec b$ and $b \not\prec a$ imply $M(a) < M(b)$.

By means of Theorem 1, we introduce three general classes of functionals that are isotonic with respect to the following preorders: 1-SD, GLD, 2-OGLD. We shall see that many indicators that have been proposed and used in the bibliometric literature belong to some of these classes. We recall that the relations between the aforementioned preorders are

$$\geq_1 \Rightarrow \geq_L \Rightarrow \geq_{OL}^2.$$

Thus, every indicator that is isotonic with GLD is clearly isotonic with 1-SD, etc.

a) *Isotonicity with 1-SD*

1-SD is a quite strong condition; thus, every measure of impact must be isotonic with it. It is evident that every quantile $Q(p), p \in [0, 1]$, is isotonic with 1-SD. Moreover, Theorem 1 (point 1) yields that every functional of the form

$$\mu^u = \int_0^1 u(Q(t))\,dt = \sum u(x)f(x), \quad (4)$$

where u is an increasing function, is isotonic with 1-SD. From this result, it can be straightforwardly verified that every power mean with positive exponent a (thus, the arithmetic mean, or JIF, as well) is isotonic with 1-SD. In the bibliometric literature, the class of indices of the form μ^u has been studied, within an axiomatic framework, by Bouyssou and Marchant (2011), and it is referred to as the class of generalized impact factors.

b) *Isotonicity with GLD*

We may define two main classes of measures that are isotonic with GLD by applying Theorem 1 to i) the quantile distribution (point 2) or ii) the GLC (point 1).

In the former case (i), we obtain that μ^u—that is, the generalized impact factor—is isotonic with GLD, provided that u is increasing and concave. Therefore, every power mean of order $a, \left(\sum x^a f(x)\right)^{1/a}$, with $0 < a \leq 1$, belongs in this class. Note that also the geometric-JIF—that is, a sort of geometric mean corrected for the uncited papers (Thelwall & Fairclough, 2015), given by $\tilde{\mu}^0 = \exp\{\mu^u\} - 1$, where $u(x) = \ln(x + 1)$ is isotonic with GLD.

Table 1
Classes of isotonic indicators, identified by Theorem 1, according to the different form (μ^u, Λ^u, Ψ^u) and the choice of u, which can be increasing (inc), increasing concave (icv), or the identity function (id).

	1-SD	GLD	2-OGLD
μ^u	u inc	u icv	–
Λ^u	u inc	u inc	u id
Ψ^u	u inc	u inc	u icv

In the latter case (ii), we obtain that every functional of the form

$$\Lambda^u = \int_0^1 u(L(t))\,dt, \tag{5}$$

where u is an increasing function, is isotonic with GLD.

c) Isotonicity with 2-OGLD

With regard to the 2-OGLD, we may apply Theorem 1 again, which yields that every functional of the form

$$\Psi^u = \int_0^{0.5} u(s(t))\,dt, \tag{6}$$

is isotonic with 2-OGLD if u is increasing and concave (point 2).

All these results are summarized in Table 1.

From Table 1, we can observe that the JIF is isotonic with 1-SD. Because the function $u(x) = x$ is also convex (although not strictly), JIF is (not strictly) isotonic with the GLD in a manner different from the power mean with exponent a, $0 < a < 1$, which is actually strictly isotonic with GLD and 1-SD.

3.2. The stabilized JIF

Interestingly, both classes Λ^u, Ψ^u yield, for $u(x) = x$ (that is an increasing and concave—as well as convex—function), the area under the GLC or, equivalently, the area under the SGLC, which can be expressed as

$$\int_0^1 L(p)\,dp = \int_0^{0.5} s(p)\,dp = 0.5\mu(1-G), \tag{7}$$

where $G = 2\left(\frac{1}{2} - \int_0^1 \frac{L(p)}{\mu}dp\right)$ is the well-known *Gini coefficient* (Gini, 1912), given by twice the area between the 45° line and the LC, $\frac{L(p)}{\mu}$. Note that the area under the GLC can also be expressed as

$$\int_0^1 L(p)\,dp = \int_0^1 p\mu_\uparrow(p)\,dp = 0.5\sum_{k=1}^{T}\left[L\left(\frac{k}{T}\right) + L\left(\frac{k-1}{T}\right)\right]\frac{1}{T} =$$

$$= 0.5\sum_{k=1}^{T}\left[\mu_\uparrow\left(\frac{k}{T}\right)\frac{k}{T} + \mu_\uparrow\left(\frac{k-1}{T}\right)\frac{k-1}{T}\right]\frac{1}{T}. \tag{8}$$

Finally, the following definition can be given.

Definition 6. We define the *stabilized-JIF (s-JIF)* as

$$\bar{\mu} = 2\int_0^1 L(p)dp = \mu(1-G).$$

From Eq. (8), the s-JIF can be interpreted as a *weighted average of the partial JIFs*, where the use of the normalization constant 2 in Definition 6 is necessary because $\int_0^1 L(p)\,dp \le \frac{\mu}{2}$ (i.e., the slope of the segment connecting $(0, 0)$ and $(p, L(p))$

Table 2
Proportion of times that the curves *Q*, *L*, *S* intersect.

No. of intersections	*Q*	*L*	*S*
0	0.49	0.75	0.94
1	0.32	0.21	0.06
2	0.11	0.03	0
≥3	0.09	0.01	0

cannot exceed μ). The s-JIF corrects the JIF by taking into account the shape and the concentration of the citation distribution and gives the mean value of the "distribution of the JIF" (i.e., the GLC). Such an index may represent a useful improvement upon the JIF because of three characteristics: i) its mathematical simplicity (it is a simple function of two basic indices); ii) its powerful graphical representation (namely, the area under the GLC); and iii) its interpretation (namely, an average partial JIF). Moreover, the s-JIF is strictly isotonic with 1-SD and the GLD and (non-strictly) isotonic with 2-OGLD, because it belongs to the classes Λ^u, Ψ^u. Thus, all the orders studied in Section 2 require a higher (or equal) s-JIF.

4. Bibliometric analysis

4.1. The dataset

To compare and apply to real data the ranking criteria proposed in Section 2 and some of the indicators proposed in Section 3, we perform an empirical analysis. The data consist of the citation distributions of the top 100 journals within the Scopus subject area of "Economics, Econometrics and Finance" ranked according to the Scopus JIF, that is, the IPP (impact per publication) of 2014. The list may be found at the website http://www.journalindicators.com and it consists of journals with a minimum number of 50 publications.

We recall that the IPP 2014 of a journal is basically the average number of citations received from papers published in 2014 (registered in the Scopus database), to papers published by the same journal from 2011 until 2013. In particular, Scopus takes account of the following types of citable items and citing sources: articles, reviews and conference papers. All other documents (e.g., notes, letters, articles in press, erratum, etc.) are excluded from the computation.

We downloaded from Scopus the citation distributions of all 100 journals within the aforementioned list during April 2016. We excluded all non-citable items (e.g., notes, etc.) in order to obtain sets of citable publications as close as possible to those employed by Scopus for the computation of IPPs (i.e., articles, reviews and conference papers). Once the set of papers for each journal has been selected, it is possible to request a citation report ("view citation overview") and download the citations (per paper) received in 2014 (i.e., the citation distributions involved in the computation of the JIF. We found some positive differences between the official JIFs (i.e., IPPs) and the observed JIFs that may have been due i) to a delayed update of the database (the IPPs were published by Scopus in June 2015) and ii) to a larger set of citing sources and documents (with Scopus, it is not possible to limit the citation report to particular citing sources or documents). Similar differences between official and observed values have been found and discussed, for instance, by Leydesdorff and Opthof (2010), Stern (2013) and Seiler and Wohlrabe (2014).

4.2. Rates of completeness

When ranking a set of elements, the "best" preorder to use is generally a compromise between two fundamental features: i) the unambiguousness of its preference relation; and ii) the percentage of pairs that it can rank. The ratio of the different preorders (1-SD, GLD, 2-OGLD) has been discussed in Section 2 such that our prime objective in this section is to analyse the number (percentage) of journal pairings (i.e., the corresponding distributions) that can be ranked accordingly. It is a matter of fact that the stronger the preorder, the smaller the number of ranked pairs; this is a logical relation. However, it seems particularly useful to quantify this simple concept with an empirical investigation.

With the help of a software program, it is possible to count the number of times that a function crosses another in a given interval. In particular, we are interested in the number of intersections among i) quantile distributions (there are no intersections iff 1-SD holds); ii) the GLCs (there are no intersections iff GLD holds); and iii) the integrals of the SGLC (there are no intersections iff 2-OGLD holds). The results are shown in Table 2.

Note that the total number of couples that we can obtain from a set of 100 elements is 4950. The first row of Table 2 (i.e., 0 crossings) determines the proportion of couples that are ranked according to the different preorders. Carayol and Lahatte (2009) refer to this proportion as the *rate of completeness* of the dominance relation. It is apparent that the finer dominance criteria significantly increase the percentage of ranked couples. Our empirical findings show that the finer criteria have a significantly higher discriminating power. On the one hand, the 1-SD ranks only 49% of the pairs; thus, we argue that this order is too strong and restrictive for comparing journals. On the other hand, the 2-OGLD yields an almost complete ordering of the set of journals considered, ultimately ranking 94% of the 4950 couples (see Table 2). With regard to the GLD, 75% seems to be a quite poor rate of completeness; thus, the GLD might also be too restrictive a requirement for ranking journals.

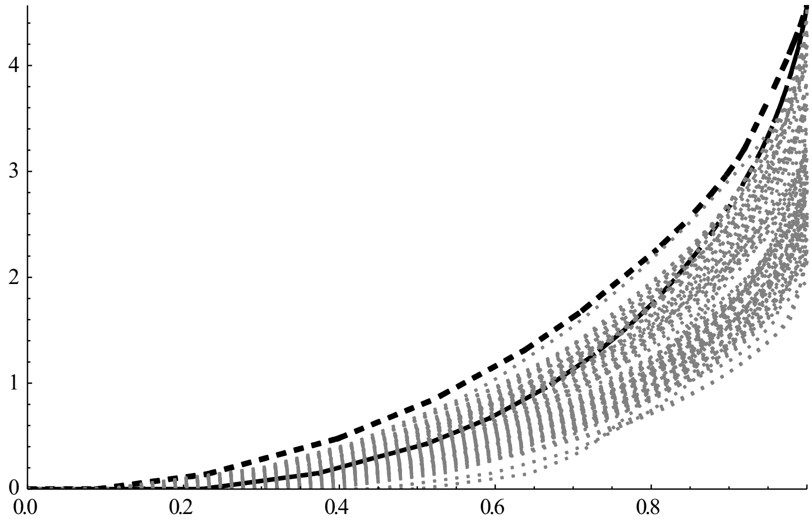

Fig. 3. Generalized Lorenz dominance. GLCs of AEJ (dashed-thick), AER (thick) and all other journals (dotted) GLD-dominated by AEJ.

4.3. Real cases supporting or contradicting the JIF

As a second step of our empirical study, we perform a different kind of analysis, aimed at understanding when the JIF is (or is not) justified by an underlying dominance relation. As discussed in Section 3, the JIF is isotonic with 1-SD and the GLD (but not isotonic with the 2-OGLD). However, in several cases, 1-SD and the GLD do not hold; thus, the JIF may be generally misleading. In this regard, it is possible to count the numbers (and percentages) of pairs for which the JIF is supported by one of these dominance relations (noting that this is a quite different computation than that reported in Table 2). We find that ranking based on JIF is supported by

- 1-SD, in 46% of the couples;
- GLD, in 69% of the couples; and
- 2-OGLD, in 75% of the couples.

Therefore, we determine that, in 25% of the pairs, the journal with a higher JIF does not 2-OGLD-dominate the other one, and in particular:

- In 10% of the pairs (i.e., 483 couples) the journal with the higher JIF is 2-OGLD-dominated by the journal with the inferior JIF.

The above results highlight the limitations of the JIF, in terms of representativeness and robustness, because they reveal that i) in 31% of the cases, the JIF is not supported by the GLD (i.e., the ordering related to the JIF via the isotonicity property); ii) in 25% of the cases, the JIF is not supported by *any* of the dominance relations studied in this paper; and iii) in 10% of the cases, the 2-OGLD even "contradicts" the ranking based on the JIF.

Some real data examples may be helpful for better understanding the possible advantages in ranking journals on the basis of a dominance relation that is finer than the 1-SD and the GLD—namely, the 2-OGLD.

Fist, to depict some particular cases of the GLD, let us consider a couple of journals that are substantially equivalent in terms of JIF—namely, *American Economic Journal: Applied Economics* (AEJ, JIF = 4.56) and *American Economic Review* (AER, JIF = 4.56). As shown in Fig. 3, the GLC (i.e., the partial JIF) of AEJ is never lower than the GLC of AER. Hence, we obtain that AEJ dominates AER w.r.t. the GLD. *A fortiori*, we find that the GLD holds when the JIF is strictly higher. Indeed, Fig. 3 also shows all journals that are GLD-dominated by AEJ. In particular, AEJ GLD-dominates 78 journals, among which 29 journals are also dominated w.r.t. 1-SD. In these cases, the higher JIF of AEJ is actually supported and justified by a dominance relation.

Fig. 4 shows the GLC of *Experimental Economics* (EE, JIF = 3.43) and the GLCs of 34 journals that, besides having slightly inferior JIFs, dominate EE w.r.t. 2-OGLD. It can be easily seen that the partial JIF of EE is uniformly lower than that of the other journals, at least until 0.9. Then, the GLC of EE starts crossing and exceeding the others, due to a few highly cited papers. Note that EE published 99 papers, and its top-four papers absorb more than 50% of the total citations, whilst, as we can see also from the figure, the first 95% of its papers have a partial JIF that is 1.52/0.95 = 1.6. Thus, EE more than doubles its JIF on account of just a few papers (4 papers, approximately 5% of its total number of publications). Conversely, the other 34 journals show more consistency, as their citations are definitely more concentrated in the body. We argue that, in this case,

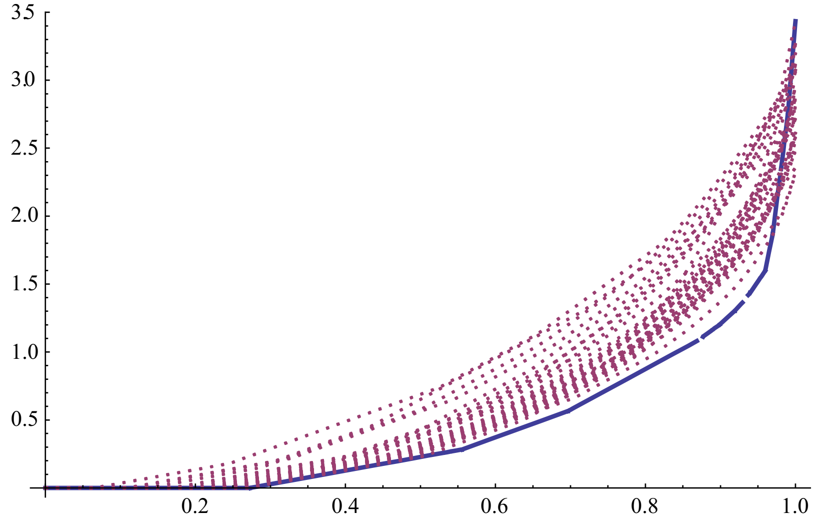

Fig. 4. 2-OGLD vs. JIF. GLC of EE (thick) and GLCs of the 34 journals, with inferior JIF, that 2-OGLD-dominate EE.

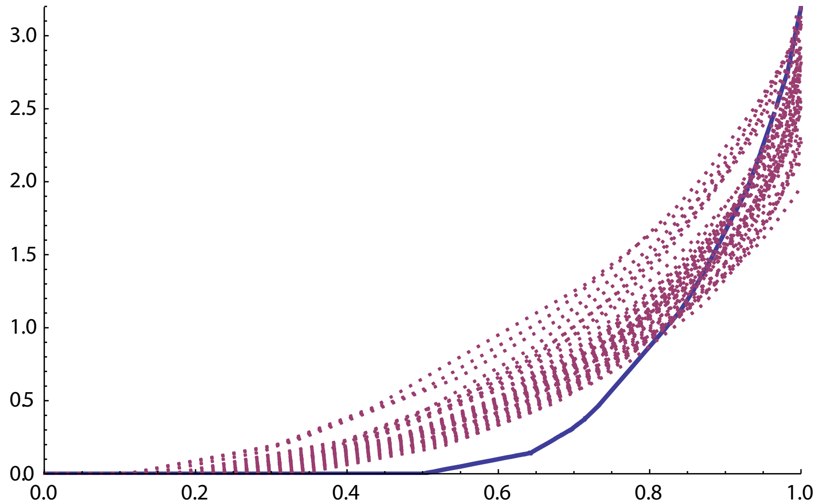

Fig. 5. 2-OGLD vs. JIF(II). GLC of BPEA (thick) and GLCs of the 41 journals, with inferior JIF, that 2-OGLD-dominate BPEA.

ranking according to the JIF would be completely misleading. A far better indicator might be the area under the GLC, i.e., the s-JIF. Indeed, it is mathematically and graphically apparent that the s-JIF ranks all the 34 journals above EE.

Another interesting case is *Brookings Papers on Economic Activity* (BPEA, JIF = 3.19). We can find 41 journals (over 100) with lower JIF that 2-OGLD-dominate BPEA. Some of these journals almost GLD-dominate BPEA, because their GLCs cross that of BPEA very close to 1 (see Fig. 5). However, also in the other, less apparent, cases, the 41 journals show much more evenly spread distributions and greater partial JIFs, at least until 0.8. Observe that the GLC of BPEA, when evaluated at 0.5, is actually equal to 0, meaning that half of the papers (from a total of 56 papers) obtained exactly 0 citations, which definitely is not meritorious. We conclude that, also in this case, the JIF is misleading, whilst other indices, such as the s-JIF, yield far more appropriate rankings.

It is possible to count and analyse how often contradicting situations, such as the ones reported in Figs. 4 and 5, occur in our dataset (let us denote by the term "contradiction" those particular cases). For a given journal, say A, we compute the number of journals that present JIF < μ_A but, at the same time, dominate A w.r.t. 2-OGLD. Then, we obtain that the total number of contradictions is 526, meaning that, for each journal, we find, on average, more than 5 journals such that 2-OGLD contradicts the JIF. Moreover, the journals without contradictions are only 22 (over 100). Thus, situations similar to those in Fig. 5 and 6 are not rare.

4.4. Indicators' performance

Last, it is interesting to compare some different indicators of impact that are isotonic with a given preorder. In this manner, we can determine whether the finer indices significantly modify the journal ranking. We consider the three families μ^u, Λ^u, Ψ^u discussed in Section 3 and, in particular, the following four indicators: 1) the JIF, which is given by $\mu = \mu^u$ with $u(x) = x$ (increasing); 2) the geometric JIF (g-JIF) used by Thelwall and Fairclough (2015), given by $\tilde{\mu}^0 = \exp\{\mu^u\} - 1$ with $u(x) = \ln(x+1)$ (increasing and concave); 3) the s-JIF $\bar{\mu} = 2\Lambda^u = 2\Psi^u$ with $u(x) = x$ (increasing); and 4) the index $\Psi^{0.5} = \Psi^u$ with $u(x) = \sqrt{x}$ (increasing and concave). Note that we use two isotonic indicators for GLD, namely, $\tilde{\mu}^0$, $\bar{\mu}$. We also compare these indices to the median (*Me*), which is clearly a robust measure, although its main drawback is that it yields a very large number of ties. The results and rankings are presented in Table A.1. We note that the most remarkable differences, in terms of ranking, can be found among the central group of journals, rather than among elite journals, e.g., the established top-five journals remain unchanged across all indicators used. These findings confirm the empirical results obtained by Stern (2013), who proposed to measure the "uncertainty" of the JIF with confidence intervals (although citation distributions are not random samples, so that the use of such a methodology is not really justified). We also note that the g-JIF and the s-JIF yield similar rankings and are highly correlated ($\rho(\tilde{\mu}^0, \bar{\mu}) = 0.998$), although the s-JIF is slightly less correlated with the JIF ($\rho(\mu, \bar{\mu}) = 0.956$, $\rho(\tilde{\mu}^0, \mu) = 0.964$). The s-JIF is more sensitive to the shape of the distribution, as we can state from both theoretical (it is strictly isotonic with the GLD and non-strictly isotonic with the 2-OGLD) and empirical viewpoints. In addition to its mathematical properties of isotonicity, we argue that the s-JIF has three main advantages compared to the g-JIF: i) its formulation is much simpler from a mathematical point of view, as it is given by the difference between two basic indicators (i.e., the JIF and Gini's mean difference); ii) it does not require any adjustment for non-cited papers; and iii) its graphical interpretation, determined by the area under the GLC, provides further justification for its use. On the other hand, the index $\Psi^{0.5}$ modifies the ranking in a more substantial way compared to the g-JIF and the s-JIF ($\rho\left(\Psi^{0.5}, \mu\right) = 0.92$), and it is more sensitive to the shape of the distribution; however, it is slightly more complicated to compute and to interpret. Because the simplicity of an index seems to be among the main requirements demanded by the bibliometric literature, we argue that the most suitable index for enhancing the measurement of journal impact is the s-JIF.

5. Conclusions

We proposed different dominance relations with which to rank citation distributions journals in terms of impact. We have shown that the GLC may serve as a powerful tool for representation and comparison of citation impact because it determines the distribution of the JIF within the publications of a journal. We have analysed the relations among the ranking criteria proposed and introduced families of isotonic indicators that may be used alternatively to the JIF in order to compensate for its limitations, i.e., its lack of robustness and insensitivity to the shape of the distribution.

The empirical analysis has shown that the use of the JIF for ranking journals may sometimes be unjustified. Indeed, we have deduced that, in 25% of the pairs, the journal with the higher JIF does not dominate the other w.r.t. any of the dominance relations considered. In particular, in 10% of the cases, the 2-OGLD contradicts the ranking based on the JIF, where the average number of contradictions per journal is 5.26 (5.26%). Conversely, the 2-OGLD yields an almost complete ranking of the journals, so that the use of the corresponding isotonic indicators seems to be amply justified.

To summarize, our empirical findings have shown that citation distributions may have completely different shapes, and citation impact can be distributed across paper in many different ways, as represented by the GLCs. Such differences in the distributions lead to proposals for some alternative methods for ranking journals—namely, 1-SD, GLD and 2-OGLD. Among these dominance relations, the 2-OGLD is probably the most questionable, although, on the other hand, it is the most sensitive to some distributional aspects that cannot be captured and compared by using 1-SD and GLD.

Finally, the results obtained certainly provide a solid justification for ranking journals on the basis of indicators that are finer than the JIF, such as the s-JIF, which may represent a simple and reliable improvement upon the JIF.

Funding

This paper has been supported by the Italian funds ex MURST 60% 2015 and the Italian Talented Young researchers project Italy® 2015/16. The research was also supported through the Czech Science Foundation (GACR) under project 17-23411Y (to T.L.).

Author contributions

Conceived and designed the analysis: Lucio Bertoli-Barsotti.
Collected the data: Tommaso Lando.
Contributed data or analysis tools: Lucio Bertoli-Barsotti.
Performed the analysis: Tommaso Lando.
Wrote the paper: Tommaso Lando.

Acknowledgments

The authors thank the two anonymous referees for their constructive comments.

Appendix A

See Table A1

Table A1
Top 100 economic journals according to Scopus IPP and their corresponding indicators (T = total number of items, Me = median). The ranks yielded by the indicators of impact JIF, g-JIF, s-JIF and $\Psi^{0.5}$, are given in brackets. Each journal is identified by its ISSN code.

ISSN	T	JIF	g-JIF	s-JIF	$\Psi^{0.5}$	Me	ISSN	T	JIF	g-JIF	s-JIF	$\Psi^{0.5}$	Me
0022–0515	69	10.101 [1]	5.451 [2]	4.106 [2]	0.986 [3]	4	0306–9192	291	3.268 [51]	1.929 [59]	1.265 [60]	0.547 [59]	2
1557–1211	193	9.187 [2]	5.282 [3]	4.019 [4]	0.984 [4]	5	2041-417X	55	3.236 [52]	1.612 [80]	1.007 [82]	0.47 [85]	1
1531–4650	127	9.142 [3]	5.645 [1]	4.367 [1]	1.025 [1]	6	0022–1996	247	3.215 [53]	2.023 [53]	1.37 [54]	0.57 [55]	2
1540–6261	190	8.047 [4]	5.281 [4]	4.025 [3]	0.989 [2]	5	1533–4465	56	3.196 [54]	1.358 [97]	0.866 [94]	0.399 [99]	0
1547–7185	153	7.817 [5]	5.11 [5]	3.867 [5]	0.968 [5]	5	1873–5991	540	3.194 [55]	2.078 [52]	1.425 [51]	0.582 [52]	2
0895–3309	133	7.481 [6]	4.414 [7]	3.468 [6]	0.911 [7]	5	1392–8619	117	3.145 [56]	1.711 [72]	1.101 [74]	0.502 [78]	2
0092–0703	140	7.250 [7]	3.882 [8]	2.758 [9]	0.81 [10]	3	1572–3089	86	3.116 [57]	1.979 [56]	1.322 [58]	0.564 [56]	2
1741–6248	55	6.564 [8]	4.564 [6]	3.404 [7]	0.914 [6]	4	1932-443X	53	3.075 [58]	2.371 [40]	1.752 [37]	0.656 [33]	3
1540–6520	175	6.069 [9]	3.631 [12]	2.715 [11]	0.805 [12]	4	0265–1335	82	3.073 [59]	2.099 [51]	1.513 [46]	0.603 [46]	3
0304-405X	412	5.857 [10]	3.724 [10]	2.704 [12]	0.809 [11]	4	1075–4253	80	3.063 [60]	2.336 [42]	1.72 [39]	0.648 [36]	2
1086–4415	57	5.684 [11]	3.454 [14]	2.518 [15]	0.777 [15]	4	0969–5931	213	3.061 [61]	2.106 [50]	1.474 [49]	0.595 [47]	2
1468–0262	187	5.422 [12]	3.715 [11]	2.751 [10]	0.818 [9]	4	1537–2707	139	3.036 [62]	1.433 [91]	0.86 [96]	0.436 [96]	1
1523–2409	81	5.358 [13]	3.74 [9]	2.856 [8]	0.835 [8]	5	1464–5114	119	3.008 [63]	2.271 [44]	1.634 [42]	0.632 [40]	2
1537-534X	92	5.250 [14]	3.309 [16]	2.429 [16]	0.767 [16]	4	0301–4207	165	2.970 [64]	1.746 [69]	1.124 [72]	0.513 [73]	2
1478–6990	155	5.129 [15]	3.461 [13]	2.529 [14]	0.784 [14]	4	0304–3878	295	2.969 [65]	1.769 [67]	1.143 [69]	0.518 [70]	2
1465–7368	288	4.823 [16]	3.215 [18]	2.337 [18]	0.751 [18]	3	0038–0121	74	2.946 [66]	1.52 [86]	0.927 [89]	0.458 [92]	1
1537–5277	234	4.641 [17]	3.395 [15]	2.553 [13]	0.79 [13]	4	1475–1461	83	2.940 [67]	1.831 [62]	1.205 [65]	0.534 [65]	2
1945–7715	91	4.637 [18]	2.794 [25]	1.973 [25]	0.688 [25]	3	0047–2727	331	2.897 [68]	1.676 [75]	1.079 [77]	0.505 [76]	2
1945–7790	113	4.567 [19]	3.231 [17]	2.36 [17]	0.758 [17]	3	1468–1218	716	2.888 [69]	1.75 [68]	1.142 [70]	0.517 [71]	2
0002–8282	723	4.567 [20]	2.554 [30]	1.769 [35]	0.642 [39]	2	1467–6419	121	2.884 [70]	1.776 [66]	1.161 [66]	0.521 [69]	2
0925–5273	1036	4.515 [21]	2.807 [22]	1.977 [24]	0.689 [24]	3	1386–4181	68	2.824 [71]	1.553 [84]	0.958 [86]	0.467 [86]	1
1547–7193	213	4.451 [22]	2.859 [21]	2.02 [21]	0.7 [21]	3	1573–0697	934	2.813 [72]	1.824 [63]	1.218 [63]	0.537 [64]	2
0165–4101	118	4.381 [23]	2.885 [20]	2.111 [20]	0.711 [20]	3	1911–3846	156	2.801 [73]	1.627 [78]	1.031 [80]	0.488 [81]	2
1741–2900	54	4.333 [24]	2.163 [47]	1.421 [53]	0.572 [54]	2	1467–0895	57	2.789 [74]	1.974 [57]	1.424 [52]	0.584 [51]	2
1542–4774	148	4.331 [25]	2.595 [29]	1.838 [30]	0.663 [30]	3	1573–0913	261	2.755 [75]	1.81 [64]	1.217 [64]	0.537 [63]	2
0024–6301	58	4.276 [26]	2.371 [39]	1.632 [43]	0.618 [45]	3	0217–4561	148	2.716 [76]	1.962 [58]	1.368 [55]	0.576 [53]	2
1469–5758	65	4.185 [27]	2.016 [54]	1.327 [57]	0.542 [61]	2	1537–5307	79	2.709 [77]	1.86 [61]	1.281 [59]	0.553 [58]	2
1066–2243	85	4.176 [28]	2.803 [23]	2.018 [22]	0.699 [22]	3	1532–8007	102	2.647 [78]	1.663 [76]	1.072 [78]	0.505 [77]	2
1090–9516	154	4.149 [29]	2.8 [24]	2.005 [23]	0.696 [23]	3	0308–5147	88	2.625 [79]	1.553 [83]	0.985 [83]	0.474 [84]	1
1468–2710	142	4.127 [30]	2.684 [27]	1.895 [27]	0.677 [28]	3	1877–8585	64	2.609 [80]	1.804 [65]	1.232 [61]	0.544 [60]	2
0921–3449	421	4.093 [31]	2.72 [26]	1.935 [26]	0.683 [26]	3	1756–6916	167	2.593 [81]	1.684 [74]	1.109 [73]	0.514 [72]	2
1873–6181	667	4.022 [32]	2.638 [28]	1.872 [28]	0.67 [29]	3	1095–9068	126	2.579 [82]	1.728 [70]	1.149 [68]	0.522 [68]	1
1547–7215	60	3.983 [33]	2.928 [19]	2.185 [19]	0.729 [19]	3	0304–4076	404	2.567 [83]	1.661 [77]	1.087 [75]	0.508 [74]	2
0921–8009	675	3.881 [34]	2.487 [36]	1.734 [38]	0.645 [38]	3	0928–7655	133	2.556 [84]	1.481 [89]	0.92 [90]	0.463 [90]	1
1945-774X	109	3.872 [35]	2.529 [32]	1.801 [31]	0.659 [31]	3	1096–0473	87	2.529 [85]	1.536 [85]	0.973 [85]	0.475 [83]	1
0001–4826	217	3.857 [36]	2.342 [41]	1.609 [44]	0.62 [44]	2	0304–3932	154	2.526 [86]	1.405 [94]	0.852 [97]	0.444 [95]	1
1052-150X	70	3.786 [37]	2.463 [37]	1.754 [36]	0.646 [37]	3	1389–9341	325	2.514 [87]	1.706 [73]	1.139 [71]	0.522 [67]	2
1941–1383	66	3.727 [38]	2.187 [46]	1.49 [47]	0.592 [48]	2	1566–0141	110	2.482 [88]	1.718 [71]	1.159 [67]	0.529 [66]	2
1096–0449	183	3.678 [39]	2.322 [43]	1.638 [41]	0.625 [42]	3	1572–3097	107	2.477 [89]	1.599 [81]	1.044 [79]	0.495 [79]	1
1530–9290	242	3.665 [40]	2.513 [33]	1.787 [32]	0.658 [32]	2	0168–7034	74	2.378 [90]	1.253 [98]	0.752 [98]	0.393 [100]	1
1530–9142	292	3.647 [41]	2.376 [38]	1.658 [40]	0.631 [41]	2	1548–8004	101	2.356 [91]	1.623 [79]	1.085 [76]	0.507 [75]	2
1468–0297	210	3.619 [42]	2.493 [35]	1.783 [33]	0.656 [34]	3	0176–2680	179	2.324 [92]	1.565 [82]	1.029 [81]	0.495 [80]	1
1389–5753	64	3.609 [43]	2.506 [34]	1.781 [34]	0.654 [35]	2	1043-951X	194	2.314 [93]	1.516 [87]	0.977 [84]	0.483 [82]	2
1475-679X	111	3.586 [44]	2.157 [48]	1.475 [48]	0.587 [49]	2	1099–1255	163	2.282 [94]	1.419 [93]	0.885 [93]	0.454 [93]	1
1467-937X	192	3.583 [45]	2.142 [49]	1.456 [50]	0.586 [50]	2	0929–1199	244	2.266 [95]	1.454 [90]	0.928 [88]	0.467 [87]	1
1526-548X	182	3.484 [46]	1.914 [60]	1.231 [62]	0.539 [62]	2	1747-762X	91	2.253 [96]	1.24 [100]	0.734 [100]	0.411 [98]	1
1573–6938	99	3.434 [47]	1.491 [88]	0.909 [92]	0.46 [91]	1	0378–4266	893	2.232 [97]	1.391 [95]	0.864 [95]	0.449 [94]	1
0013–0095	51	3.431 [48]	2. [55]	1.329 [56]	0.557 [57]	2	1179–1896	127	2.142 [98]	1.429 [92]	0.931 [87]	0.465 [89]	1
1096–1224	61	3.279 [49]	2.546 [31]	1.871 [29]	0.679 [27]	2	1096–6099	113	2.142 [99]	1.249 [99]	0.741 [99]	0.415 [97]	1
0308-597X	475	3.278 [50]	2.264 [45]	1.592 [45]	0.621 [43]	2	1432–1122	89	1.966 [100]	1.39 [96]	0.919 [91]	0.466 [88]	1

References

Aaberge, R. (2009). Ranking intersecting Lorenz curves. *Social Choice and Welfare*, *33*(2), 235–259.
Balanda, K. P., & MacGillivray, H. L. (1990). Kurtosis and spread. *Canadian Journal of Statistics*, *18*(1), 17–30.
Bar-Ilan, J. (2010). Ranking of information and library science journals by JIF and by h-type indices. *Journal of Informetrics*, *4*, 141–147.
Billingsley, P. (1986). *Probability and measure*. John Wiley & Sons.
Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of Informetrics*, *7*(1), 158–165.
Bouyssou, D., & Marchant, T. (2011). Bibliometric rankings of journals based on impact factors: An axiomatic approach. *Journal of Informetrics*, *5*(1), 75–86.
Burrell, Q. L. (2005). Symmetry and other transformation features of Lorenz/Leimkuhler representations of informetric data. *Information Processing & Management*, *41*(6), 1317–1329.
Carayol, N., & Lahatte, A. (2009). *Dominance relations and universities ranking*. pp. 2. Cahier du GREThA.
Chiu, W. H. (2007). Intersecting Lorenz curves, the degree of downside inequality aversion, and tax reforms. *Social Choice and Welfare*, *28*(3), 375–399.
Davies, J., & Hoy, M. (1995). Making inequality comparisons when Lorenz curves intersect. *The American Economic Review*, 980–986.
Editorial. (2005). Not-so-deep impact. Research assessment rests too heavily on the inflated status of the impact factor. *Nature*, *435*, 1003–1004.
Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Elsevier.
Ferrer-Sapena, A., Sánchez-Pérez, E. A., González, L. M., Peset, F., & Aleixandre-Benavent, R. (2015). Mathematical properties of weighted impact factors based on measures of prestige of the citing journals. *Scientometrics*, *105*(3), 2089–2108.
Foo, J. Y. A. (2013). Implications of a single highly cited article on a journal and its citation indexes: A tale of two journals? *Accountability in Research*, *20*(2), 93–106.
Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, *14*(3), 195–201.
Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*(4060), 471–479.
Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, *295*(1), 90–93.
Gini, C. (1912). Variabilità e mutabilità. Reprinted in Memorie di metodologica statistica. In E. Pizetti, & T. Salvemini (Eds.), (p. 1). Rome: Libreria Eredi Virgilio Veschi.
Hoffman-Jorgensen, J. (1994). . *Probability with a view towards statistics* (vol. 1) CRC Press.
Ingwersen, P., Larsen, B., Rousseau, R., & Davis, M. (2001). The publication-citation matrix and its derived quantities. *Chinese Science Bulletin*, *46*, 524–528.
Ingwersen, P. (2012). The pragmatics of a diachronic journal impact factor. *Scientometrics*, *92*, 319–324.
Laband, D. N. (1986). Article popularity. *Economic Inquiry*, *24*(1), 173–180.
Lando, T., & Bertoli-Barsotti, L. (2016). Weak orderings for intersecting Lorenz curves. *Metron*, *74*(2), 177–192.
Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology*, *61*(11), 2365–2369.
Leydesdorff, L. (2012). Alternatives to the journal impact factor: I3 and the top-10% (or top-25%?) of the most-highly cited papers. *Scientometrics*, *92*(2), 355–365.
Marshall, A. W., Olkin, I., & Arnold, B. C. (2011). *Inequalities: Theory of majorization and its applications* (2nd ed.). Springer.
Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, *4*(3), 265–277.
Moustafa, K. (2015). The disaster of the impact factor. *Science and Engineering Ethics*, *21*(1), 139–142.
Muliere, P., & Scarsini, M. (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory*, *49*(2), 314–323.
Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, *314*(7079), 498.
Seiler, C., & Wohlrabe, K. (2014). How robust are journal rankings based on the impact factor? Evidence from the economic sciences. *Journal of Informetrics*, *8*(4), 904–911.
Shorrocks, A. F., & Foster, J. E. (1987). Transfer sensitive inequality measures. *The Review of Economic Studies*, *54*(3), 485–497.
Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, *50*(197), 3–17.
Stern, D. I. (2013). Uncertainty measures for economics journal impact factors. *Journal of Economic Literature*, 173–189.
Thelwall, M., & Fairclough, R. (2015). Geometric journal impact factors correcting for individual highly cited articles. *Journal of Informetrics*, *9*(2), 263–272.
Vanclay, J. K. (2012). Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics*, *92*(2), 211–238.
Wall, H. J. (2009). Don't get skewed over by journal rankings. *The BE Journal of Economic Analysis & Policy*, *9*(1).
Waltman, L., & Van Eck, N. J. (2009). *A taxonomy of bibliometric performance indicators based on the property of consistency ERIM report series research in management*. No. ERS-2009-014-LIS.
Yitzhaki, S., & Schechtman, E. (2012). . *The Gini Methodology: A primer on a statistical methodology* (vol. 272) Springer Science & Business Media.
Zitt, M. (2012). The journal impact factor: angel, devil, or scapegoat? A comment on JK Vanclay's article 2011. *Scientometrics*, *92*(2), 485–503.
Zoli, C. (1999). Intersecting generalized Lorenz curves and the Gini index. *Social Choice and Welfare*, *16*(2), 183–196.
Zoli, C. (2002). Inverse stochastic dominance, inequality measurement and Gini indices. *Journal of Economics*, *77*(1), 119–161.