

Measurable utility for scientific influence

Albert Burgos

(Universidad de Murcia)

Measurable utility for scientific influence*

Abstract

I give necessary and sufficient conditions for the existence of a cardinal utility function to represent, through summation, rankings of scientific units based on their journal articles and its citations. I discuss and interpret the meaning of those conditions, its connections with inequality theory and the theory of choice under uncertainty, and I connect the results of this approach to other performance measures provided by the literature on citation analysis.

JEL Classification: A11, C43, D63, D70, D81.

Keywords: Scientific rankings, citations, impact, additive utility, stochastic dominance.

Albert Burgos
Departamento de Fundamentos del Análisis Económico
Facultad de Economía y Empresa
Universidad de Murcia
30100 Espinardo, Murcia, Spain
e-mail: albert@um.es

*I thank Simon Grant, Thierry Marchant, María F. Morales, and Javier Ruiz-Castillo, who have provided me with useful feedback on this topic. Of course, remaining errors are my own. Financial support from the Spanish Ministry of Science and Education through MEC/FEDER grant SEJ2007-67580-C02-02 is gratefully acknowledged.

“It is quality rather than quantity that matters...”, Lucius Annaeus Seneca 5BC-65AD, Roman tragedian, philosopher, and counselor to Nero, Epistles.

1 Introduction

Although peer review is still the prevalent way by which individual research quality is judged, and most researchers prefer an explicit peer assessment of their work, those same researchers know how time-consuming peer assessment can be. The most obvious alternative available to rank the quality of scientists, and, by extension, of scientific units (interpreted here not only as individual authors, but also as university departments, journals, countries or other geographical areas) is using their citation record—that is, the number and impact of their publications. As a result, a good deal of work is currently concerned with converting those citation records into different (cardinal) measures of quality. The key is focusing not on where unit publish but on how many times other scientists cite their work, and translate this data into a scalar measure.

In this note I shall develop a model in which a decision maker, which I dub “evaluator” tackles the problem of ranking a given set of research units. My treatment of this problem borrows heavily from Camacho’s (1979a, 1979b, 1980, 1982) approach to cardinal utility in a social choice context. The reader is referred to his works, and the follow up by Wakker (1986), for a detailed description of this literature. For my purposes here, it is sufficient to know that the model provides conditions under which preferences on a set \mathcal{X} of alternatives extend to cardinal preferences on the n -fold cartesian product of \mathcal{X} for any natural n . My task here is translating this result into the context of measuring scientific output. This can be easily accomplished by: (i) taking as the set of alternatives the set \mathbb{N}_0 of possible citation scores for any publication,¹ (ii) identifying the set of all possible citation records for a unit as the n -fold cartesian product of \mathbb{N}_0 , and (iii) extending the natural order on \mathbb{N}_0 to a cardinal utility function on sequences of elements in this cartesian product under some “plausible” conditions.

I shall proceed now to briefly enumerate informally these assumptions:

1. The evaluator can consistently and completely order any given set of scientific units.
2. For any pair of units having published a single publication each, the evaluator prefers the unit having received more citations.
3. For any unit, its ranking depends only on the citation scores entering the citation record to be evaluated and not on the order in which scores are presented.

¹ For ease of exposition, I assume throughout that the citation count of any given paper is a non-negative integer. The representation results in this paper, however, hold for any set of possible citations which is countable. In particular, one can choose this set to be \mathbb{Q}_+ , allowing citations scores to be counted fractionally, for instance, in multi-authored publications.

4. The evaluator’s preferences between two units is independent of publications having the same citation score.
5. There is not a publication that is by itself necessary or sufficient to make a unit preferred over another. The ranking between any pair of units may be reversed by a sufficient number of additional publications.

The main result is that these assumptions are sufficient for the construction of an additively separable representation of the evaluator’s preferences. This representation allows for a cardinal measure of influence that may overcome some of the criticisms commonly raised against single-number criteria to evaluate scientific output, as the sensibility to a small number of “big hits”, or the penalization of high productivity.

2 Set-up and examples

I am concerned with a (binary) relation \succsim over ordered pairs in a nonempty set of scientific units \mathcal{X} represented as mappings from a set of articles to the set of possible citation scores. For each unit $\mathbf{x} \in \mathcal{X}$, I denote by $|\mathbf{x}|$ the number of articles published by this unit over a given period T and by x_i the number of citations for each article,² arranged in some pre-specified order. I focus on the case in which the set of possible citation scores x_i is \mathbb{N}_0 . (No major difficulties arise with the extension to any countable spaces). For mathematical convenience, I work with infinite sequences $\{x_i\}_{i \in \mathbb{N}}$ with “tail” -1 . Slightly abusing notation, then, I shall thereafter identify \mathcal{X} with the set of all sequences $\{x_i\}_{i \in \mathbb{N}}$ for which $x_i \in \mathbb{N}_0 \cup \{-1\}$ for all i , and $|\{i \mid x_i \neq -1\}| = |\mathbf{x}| < +\infty$.

Given $\mathbf{x} \in \mathcal{X}$, a performance index as defined in the literature of citation analysis is a mapping from \mathcal{X} into \mathbb{R} which quantifies the impact of every possible output in \mathcal{X} . What is sought are measures leading to rankings that hopefully would make sense to authors, editors, publishers, etc. having potentially different preferences. The following are some examples of rankings induced by performance indices taken from this literature:

- TOTAL NUMBER OF PAPERS: $n_{p,tot}(\mathbf{x}) = |\mathbf{x}|$.

The ranking based on the $n_{p,tot}$ score focuses on total production and completely disregards the information about impact or visibility of each article.

- TOTAL NUMBER OF CITATIONS: $n_{c,tot}(\mathbf{x}) = \sum_{i=1}^{|\mathbf{x}|} x_i$.

The ranking induced by the $n_{c,tot}$ score measures is based on the total impact of a unit’s research, and is independent of the volume of production needed to obtain a given score.

² In particular, this reduced form of the citation network implies that all citations are independent of the citing paper. This way of modelling is, however, “plug compatible” with an analysis in which citations are weighted depending on the journal on which the citing paper is published, as in Palacios-Huerta and Volij (2004) or Slutzki and Volij (2006).

- MEAN CITATION RATE (IMPACT FACTOR): $MCR(\mathbf{x}) = n_{c,tot}(\mathbf{x}) / n_{p,tot}(\mathbf{x})$. [Garfield (1972)]

Ranking units according to the MCR rewards units publishing widely cited papers and penalizes those showing more diversity and sustainability in research. Initially developed for journals, nowadays it is also used as a proxy for rating individual scientists.

- NUMBER OF HIGHLY CITED PUBLICATIONS: $HCP_\alpha(\mathbf{x}) = |\{i \mid x_i \geq \alpha\}|$.

The number HCP_α gives an idea of broad and sustained impact. However, it is often criticized by arbitrariness in the choice of the threshold α .

- HIRSCH INDEX: $h(\mathbf{x}) = |\{i \mid x_i \geq i\}|$. [Hirsch (2005)]

The h-index induces a ranking that attempts to evaluate both the scientific productivity and the apparent scientific impact of a unit. In its five years of life, the h-index has been the biggest splash in a flood of rating measures for individuals.³ The main reason for its popularity seems to be: (i) it is relatively insensitive to large numbers of lowly cited publications, (ii) it is also relatively insensitive to a few very highly cited publications, (iii) the calculation is easy to understand, and (iv) recently, some authors (Woeginger (2008a,2008b), Quesada (2009a, 2009b) or Marchant (2009)) provide lists that completely characterize it, thus greatly facilitating its applicability in a given context.

In the remaining of the paper, rather than delve immediately into particular methods because they look reasonable or yield intuitive results, I start from the scratch and focus on the type of structure that one might impose on preferences guaranteeing an additive utility representation but avoiding the straitjacket of a particular functional.

3 Axioms

I am interested in finding conditions that lead to a simple representation of the evaluator's preferences. I next introduce some basic properties that \succsim may satisfy. As usual, the symbols \succ and \sim correspond to strict preference and indifference.

Throughout the paper I will assume:

A1 (WEAK ORDER) \succsim is complete and transitive.

³ Panaretos and Malesios (2009) or Alonso *et al.* (2009) offer comprehensive reviews of the bibliometrics literature on the topic. Even more recent works are Quesada (2009a, 2009b) and Marchant (2009, 2010). Among economists, Ruan and Toll (2008) apply the h-index metric to rank economics departments in the Republic of Ireland, and Harzing and van der Wal (2008) use it as an alternative metric to the ISI Impact Factor. Harzing (2009) offers a software program, *Publish or Perish*, which uses Google Scholar database to obtain several indexes of quality, among them the h-index and the closely related g-index (Egghe (2006a, 2006b)).

Although A1 is fairly standard in decision theory, in this context it is a strong axiom, and precludes situations in which the evaluator uses the intersection of several indices to rank units (which leads to an incomplete relation \succsim) or neglects small differences in citation records (which implies an intransitive \sim).

A2 (MONOTONICITY) *For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $|\mathbf{x}| = |\mathbf{y}| = 1$, then $\mathbf{x} \succ \mathbf{y}$ if and only if $x_1 > y_1$.*

A2 states that from the point of view of the evaluator, impact is desirable, therefore when comparing units containing a single publication each, the unit with more citations should be preferred. This axiom is rather weak, and it is satisfied by all the examples showed in the previous section.

In order to motivate my third axiom, notice that, without further conditions, the order in which papers are ordered might be an issue for the evaluator. In this way, I shall require that a reordering of the publications of a unit does not change its ranking. Thus, the varying orders in which at least some articles can be counted (e.g. chronologically, from high to low impact, or any other variant) should not matter. Namely,

A3 (PERMUTABILITY) *For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $n \in \mathbb{N}$, if σ is a permutation on $\{1, \dots, n\}$ such that $y_i = x_{\sigma(i)}$ for all $i \in \{1, \dots, n\}$ and $y_i = x_i$ for all $i \notin \{1, \dots, n\}$, then $\mathbf{x} \sim \mathbf{y}$.*

Axiom A3 precludes “reputation effects”, which would make the evaluation of an article dependent on the impact of preceding publications.

The following axiom explores further the separability of preferences across publications by requiring that the preference between units \mathbf{x} and \mathbf{y} should be independent of publications have identical impact:

A4 (INDEPENDENCE) *For all $\mathbf{x}, \mathbf{y}, \mathbf{x}^*, \mathbf{y}^* \in \mathcal{X}$ and $i \in \mathbb{N}$ such that*

$$\begin{cases} x_i = y_i \\ x_j^* = x_j \quad \forall j \neq i \end{cases}$$

and

$$\begin{cases} y_i^* = x_i^* \\ y_j^* = y_j \quad \forall j \neq i, \end{cases}$$

then $\mathbf{x} \succ \mathbf{y}$ if and only if $\mathbf{x}^ \succ \mathbf{y}^*$.*

A4 is undoubtedly a strong assumption. In particular, if units \mathbf{x} and \mathbf{y} share some publication i , (i.e. $x_i = y_i$), and $\mathbf{x} \succ \mathbf{y}$, then this relative position is unaffected by the impact of this publication. Interestingly, the well known h-index and its variants, as the g-index (Egghe (2006a, 2006b)) violate independence. However, life without independence can lead to somewhat paradoxical results: It may be the case that $\mathbf{x} \succ \mathbf{y}$ when units \mathbf{x} and \mathbf{y} are research groups but using the same utility preference relation at the individual level each researcher in \mathbf{x} outperforms each researcher in \mathbf{y} . In the light of this shortcoming, I regard independence as a desirable property.

Finally, I shall assume:

A5 (ARCHIMEDEAN) For all \mathbf{x} , \mathbf{y} , \mathbf{x}^* , and $\mathbf{y}^* \in \mathcal{X}$ with $\mathbf{x} \succ \mathbf{y}$ and $\mathbf{x}^* \succ \mathbf{y}^*$ there exists $m \in \mathbb{N}$ such that if

$$\begin{cases} x'_{i|\mathbf{x}|+j} = x_j & \forall 0 \leq i < m \text{ and } 1 \leq j \leq |\mathbf{x}| \\ x'_{m|\mathbf{x}|+j} = y_j^* & \forall 1 \leq j \leq |\mathbf{x}^*| \\ x'_j = -1 & \forall j > m|\mathbf{x}| + |\mathbf{y}^*| \end{cases}$$

and

$$\begin{cases} y'_{i|\mathbf{x}|+j} = y_j & \forall 0 \leq i < m \text{ and } 1 \leq j \leq |\mathbf{y}| \\ y'_{m|\mathbf{y}|+j} = x_j^* & \forall 1 \leq j \leq |\mathbf{x}^*| \\ y'_j = -1 & \text{if } j > m|\mathbf{y}| + |\mathbf{x}^*| \end{cases}$$

then $\mathbf{x}' \succ \mathbf{y}'$.

Notice that the “untailed” part of \mathbf{x}' is constructed using m replicas of the “untailed” part of \mathbf{x} , followed by one replica of the “untailed” part of \mathbf{y}^* . Unit \mathbf{y}' is built in the same way using instead m replicas of the “untailed” part of \mathbf{y} and one of the “untailed” part of \mathbf{x}^* . A5 implies that the difference between any pair of units \mathbf{x}^* and \mathbf{y}^* can be compensated by a sufficiently large number of differences between \mathbf{x} and \mathbf{y} .

Archimedean-like axioms are necessary whenever one wishes to obtain a numerical representation of a weak order. In particular, A5 would fail if preferences are “lexicographic” in the sense that some publications combined with other ones in whatever manner would always lead to a strict preference.

4 A (cardinal) representation of the evaluator’s preferences

The five axioms introduced in the preceding section lead to the following representation result:

Theorem 4.1 Preferences \succ over scientific units satisfy A1-A5 if and only if there is increasing real-valued mapping u from \mathbb{N}_0 such that, for all \mathbf{x} and $\mathbf{y} \in \mathcal{X}$,

$$\mathbf{x} \succ \mathbf{y} \iff \sum_{i \in \mathbb{N}} [u(x_i) - u(y_i)] \geq 0. \quad (1)$$

The function u is unique up to scale and location, i.e. if (1) holds, u can be replaced by v if and only if $v = \beta u + \tau$ for some $\beta, \tau \in \mathbb{R}$ with $\beta > 0$.

Proof. Necessity is easily established. Sufficiency follows from Theorem 1.1 in Wakker (1986), which guarantees that A1, A3, A4, and A5 imply the existence and the uniqueness of u up to a linear transformation such that representation (1) holds. A2 guarantees that such a u must be increasing. ■

Conditions A1-A5 lead to a representation for \succ which separate the effect of the size and the distribution of impact. I shall introduce here some additional notation to formalize this. For a given $\mathbf{x} \in \mathcal{X}$, let $\mathbf{f}_{\mathbf{x}} = \{f_{\mathbf{x}}(i)\}_{i \in \mathbb{N}_0}$ be its relative distribution of papers according to impact, i.e., for

all $i \in \mathbb{N}_0$, $f_{\mathbf{x}}(i) \in \mathbb{Q}_+$ is the proportion of papers published by unit \mathbf{x} having received exactly i citations. By setting $u(-1) = 0$, representation (1) can be written as

$$\mathbf{x} \succcurlyeq \mathbf{y} \iff |\mathbf{x}| \sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) \geq |\mathbf{y}| \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i). \quad (2)$$

where u is now unique up to scale.

4.1 Remarks on the shape of u

Some popular citation measures may be derived from representations (1) or (2) by imposing additional conditions which constraint the shape of the function u , which I dub the *impact function*. By imposing this impact function equal to a positive constant, the representation (1) implies that the evaluator ranks units in accordance with the total number of publications $n_{p,tot}$. If the impact function is linear, one obtains the ranking based on the *total number of citations* ($n_{c,tot}$), i.e. the utility of any unit $\mathbf{x} \in \mathcal{X}$ is given by the total impact of its research, and is independent of the volume of production needed to obtain a given score. Clearly, for the impact function to be linear, it is necessary that if a unit gets an additional citation, it should not matter which of its publications receives this citation. Namely,

A6 (TRANSFERABILITY) For all $\mathbf{x} \in \mathcal{X}$ and for all $i, j \in \mathbb{N}$, if

$$\begin{cases} x_i^* = x_i + 1 \\ x_k^* = x_k \end{cases} \quad \forall k \neq i, \quad \text{and} \quad \begin{cases} x_j' = x_j + 1 \\ x_k' = x_k \end{cases} \quad \forall k \neq j,$$

then $\mathbf{x}^* \sim \mathbf{x}'$.

Under A6, additional citations are perfect substitutes across different publications of the same unit. When added to the conditions used in Theorem 1, it is sufficient to obtain a ranking based in the total number of citations:

Corollary 4.2 Preferences \succcurlyeq over scientific units satisfy A1-A6 if and only if for all \mathbf{x} and $\mathbf{y} \in \mathcal{X}$,

$$\mathbf{x} \succcurlyeq \mathbf{y} \iff \sum_{i \in \mathbb{N}} [x_i - y_i] \geq 0. \quad (3)$$

Proof. Necessity is straightforward. I show sufficiency. By Theorem 1, A1-A5 imply that here is an increasing u such that $u(-1) = 0$ and $U(\mathbf{x}) = \sum_{i \in \mathbb{N}} u(x_i)$. For any pair $i, j \in \mathbb{N}$, ($j > i$) set \mathbf{x} be such that

$$\begin{cases} x_k = z + 1 & \text{if } k = i \\ y_k = z & \text{if } k = j \\ y_k = -1 & \forall k \neq i, j \end{cases}$$

where $z \in \mathbb{N}_0$. Now, let \mathbf{y} and \mathbf{z} defined as

$$\begin{cases} y_k = z + 1 & \text{if } k = i \\ y_k = z + 1 & \text{if } k = j \\ y_k = -1 & \forall k \neq i, j \end{cases} \quad \text{and} \quad \begin{cases} z_k = z + 2 & \text{if } k = i \\ z_k = z & \text{if } k = j \\ z_k = -1 & \forall k \neq i, j \end{cases}$$

respectively. By A6, $\mathbf{x} \sim \mathbf{y} \sim \mathbf{z}$, then $u(z+2) - u(z+1) = u(z+1) - u(z)$ holds for any given z hence u must be linear. ■

A6 implies constant marginal utility of the impact of citations. Of course, the axiom can be easily modified so that the marginal utility of citations below a certain threshold α is zero. This yields the ranking based on the number of *highly cited publications* (HCP_α).

The case in which this marginal utility is decreasing deserves further analysis. Although there is a sense of arbitrariness in asking about the possibility of having $\{u(x_i)\}_{i=1}^{|\mathbf{x}|}$ concave,⁴ and there seems to be no simple defense of global concavity (or global convexity for that matter) in this context, the next section offers some justification for its use: I shall come to that soon. Meanwhile, I remark that the concavity of the impact function attenuates the influence of one highly cited article which may not be representative of the broad impact of the unit, hence capturing one of the most appealing features of the h-index but still retaining the property of independence.

A7 (CONCAVITY) For all $\mathbf{x} \in \mathcal{X}$ and for all $i, j \in \mathbb{N}$ with $x_i > x_j$, if

$$\begin{cases} x_i^* = x_i + 1 \\ x_k^* = x_k & \forall k \neq i, \end{cases} \quad \text{and} \quad \begin{cases} x_j' = x_j + 1 \\ x_k' = x_k & \forall k \neq j, \end{cases}$$

then $\mathbf{x}^* \prec \mathbf{x}'$.

If we add A7 to axioms A1-A5, we get

Corollary 4.3 Preferences \succsim over scientific units satisfy A1-A5 and A7 if and only if there is increasing and concave real-valued mapping u from \mathbb{N}_0 such that, for all \mathbf{x} and $\mathbf{y} \in \mathcal{X}$,

$$\mathbf{x} \succsim \mathbf{y} \iff \sum_{i \in \mathbb{N}} [u(x_i) - u(y_i)] \geq 0. \quad (4)$$

If (2) holds, u can be replaced by a mapping v if and only if $v = \beta u + \tau$ for some $\beta, \tau \in \mathbb{R}$ with $\beta > 0$.

Proof. Essentially the same as the proof of Corollary 4.2 ■

⁴ A finite sequence $S = \{z_1, z_2, \dots, z_m\}$ in \mathbb{R} is concave iff it is strictly increasing and $z_1 - z_2 \geq z_2 - z_3 \geq \dots \geq z_m - z_{m+1}$.

4.2 Remarks on size

It is clear that under A1-A5, the evaluator's preferences are not *size-invariant*, in the sense that two different units having the same relative distribution of citations are never indifferent. In fact, A1-A5 imply a multiplicative scaling behavior. To see this, define, for all $\mathbf{x} \in \mathcal{X}$, and $t \in \mathbb{N}$, the unit $t\mathbf{x}$ in \mathcal{X} constructed by repeating t times the sequence of papers in \mathbf{x} , so that the untailed part of $t\mathbf{x}$ consists of t replicas of the untailed part of \mathbf{x} . Since $t\mathbf{x}$ and \mathbf{x} have the same relative distribution of papers, using representation (2) one can think of a utility representation U such that $U(t\mathbf{x}) = t|\mathbf{x}| \sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) = tU(\mathbf{x})$, i.e. U is homogeneous of degree 1 in t . This property raises questions about the nature of the scientific units under evaluation:

On the one hand, a unit that publishes a larger number of papers has a higher likelihood of generating higher utility, since every article presents another chance for citations. Nothing wrong with that if think that a unit that publishes a larger number of high impact papers has indeed a bigger influence on the field. If this influence is what the evaluator is attempting to measure, one can argue that the scaling feature implied by A1-A5 is an appealing property.

On the other hand, the presence of scaling-effects would be a disadvantage when evaluating, for instance, the standing of individual articles in a journal (as this measure should not be dependent on the number of articles published in that journal). Moreover, when preferences are size-dependent, a unit might try to manipulate its ranking by splitting all their papers into smaller pieces. It was this kind of perverse effects that were in the mind of Eugene Garfield more than 50 years ago when developing the *Impact factor* (IF) as a size-independent index (for a throughout discussion on this topic see Garfield (2006) and the references therein).

Clearly, if the evaluator wishes to capture the spirit of the IF, it must be through the normalization of the number of citations received by the size of the sequence of publications, so that the utility of a unit becomes homogeneous of degree 0 in the number of replicas t of any given distribution. That is,

A8 (SIZE INVARIANCE) For all $\mathbf{x} \in \mathcal{X}$ and for all $t \in \mathbb{N}$, $t\mathbf{x} \sim \mathbf{x}$.

A8 is incompatible with A1-A5. Thus, if one accepts A8 as a desirable property we know that at least one of the five axioms in Theorem 4.1 will be violated. It is straightforward to check that the IF and similar average indices satisfy A1-A4 and A8, but violate the Archimedean axiom A5. They are compatible, however, with the following variant of the archimedean axiom, first proposed in Bouyssou and Marchant (2010):

A5* (ARCHIMEDEAN*) For all \mathbf{x} , \mathbf{y} , and $\mathbf{z} \in \mathcal{X}$ with $|\mathbf{x}| = |\mathbf{y}| = |\mathbf{z}|$ there exist n , m , n^* , and $m^* \in \mathbb{N}$ such that if

$$\begin{cases} x_{i|\mathbf{x}|+j}^* = x_j & \forall 0 \leq i < n \text{ and } 1 \leq j \leq |\mathbf{x}| \\ x_{n|\mathbf{x}|+j}^* = z_j & \forall 0 \leq i < m \text{ and } 1 \leq j \leq |\mathbf{z}| \\ x_j^* = -1 & \forall j > n|\mathbf{x}| + m|\mathbf{z}|. \end{cases}$$

and

$$\begin{cases} z_{i|\mathbf{x}|+j}^* = x_j & \forall 0 \leq i < n^* \text{ and } 1 \leq j \leq |\mathbf{x}| \\ z_{n^*|\mathbf{x}|+j}^* = z_j & \forall 0 \leq i < m^* \text{ and } 1 \leq j \leq |\mathbf{z}| \\ z_j^* = 0 & \forall j > n^*|\mathbf{x}| + m^*|\mathbf{z}|. \end{cases}$$

then

$$\mathbf{x} \succ \mathbf{y} \succ \mathbf{z} \implies \mathbf{x}^* \succ (n+m)\mathbf{y} \text{ and } (n^*+m^*)\mathbf{y} \succ \mathbf{z}^*.$$

A5* implies that for any three units of the same size, by merging a sufficiently large number (n) of replicas of the highest ranked unit with a sufficiently small number of replicas (m) of the lowest ranked unit we obtain a new composed unit which has a higher ranking than $n+m$ replicas of the intermediate unit. However, we can reverse this preference by selecting other naturals n^* and m^* where n^* is small compared to m^* .

Replacing A5 by A5* yields the following result:

Theorem 4.4 *Preferences \succcurlyeq over scientific units satisfy A1-A4, A5*, and A8 if and only if there is an increasing real-valued mapping u from \mathbb{N}_0 such that, for all \mathbf{x} and $\mathbf{y} \in \mathcal{X}$,*

$$\mathbf{x} \succcurlyeq \mathbf{y} \iff \sum_{i \in \mathbb{N}} \left[\frac{u(x_i)}{|\mathbf{x}|} - \frac{u(y_i)}{|\mathbf{y}|} \right] \geq 0. \quad (5)$$

The function u is unique up to scale and location, i.e. if (1) holds, u can be replaced by v if and only if $v = \beta u + \tau$ for some $\beta, \tau \in \mathbb{R}$ with $\beta > 0$.

Proof. Necessity is trivial. I show sufficiency. This result is essentially Bouyssou and Marchant (2010, Thm. 1), and uses the representation result in Shepherdson (1980, Thm. 5.2) for the set Δ of probability distributions with rational support. From preferences \succcurlyeq , I induce a binary relation \succcurlyeq^* on Δ as follows: For any pair $\boldsymbol{\pi}, \boldsymbol{\mu} \in \Delta$, $\boldsymbol{\pi} \succcurlyeq^* \boldsymbol{\mu} \iff \mathbf{x} \succcurlyeq \mathbf{y}$ where $\mathbf{f}_{\mathbf{x}} = \boldsymbol{\pi}$ and $\mathbf{f}_{\mathbf{y}} = \boldsymbol{\mu}$. I claim that this relation is well defined. To prove this, notice first that $\mathbf{f}_{\mathbf{x}} \sim \mathbf{f}_{\mathbf{y}}$ implies $\mathbf{x} \sim \mathbf{y}$ (from A3, $|\mathbf{x}|\mathbf{x} \sim |\mathbf{y}|\mathbf{y}$, and from A8 $\mathbf{x} \sim |\mathbf{x}|\mathbf{x}$ and $\mathbf{y} \sim |\mathbf{y}|\mathbf{y}$, hence $\mathbf{x} \sim \mathbf{y}$ by transitivity). To prove the converse, suppose $\mathbf{x} \succcurlyeq \mathbf{y}$ and $\mathbf{f}_{\mathbf{x}} \prec^* \mathbf{f}_{\mathbf{y}}$. This implies that $\mathbf{f}_{\mathbf{x}} = \mathbf{f}_{\mathbf{x}'}$, $\mathbf{f}_{\mathbf{y}} = \mathbf{f}_{\mathbf{y}'}$ and $\mathbf{x}' \prec \mathbf{y}'$ for some $\mathbf{x}', \mathbf{y}' \in \mathcal{X}$. Thus, $\mathbf{x} \sim \mathbf{x}'$ and $\mathbf{y} \sim \mathbf{y}'$. Since by hypothesis $\mathbf{x} \succcurlyeq \mathbf{y}$ and $\mathbf{x}' \prec \mathbf{y}'$, transitivity of \succcurlyeq leads to $\mathbf{x}' \succcurlyeq \mathbf{y}'$ and $\mathbf{x}' \prec \mathbf{y}'$, which proves the claim by contradiction. It follows from Bouyssou and Marchant (2010, Thm. 1) that A1, A3, A4 and A5* together imply an expected utility representation of \succcurlyeq^* , and thus $\mathbf{x} \succcurlyeq \mathbf{y} \iff \sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) \geq \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i)$ where u is unique up to an affine transformation. The fact that A2 implies u increasing completes the proof. ■

Applying additional restrictions on the shape of u , the following corollaries are now available:

Corollary 4.5 Preferences \succcurlyeq over scientific units satisfy A1-A4, A5*, A6 and A8 if and only if units are ranked according to their IF, i.e. for all \mathbf{x} and $\mathbf{y} \in \mathcal{X}$,

$$\mathbf{x} \succcurlyeq \mathbf{y} \iff \sum_{i \in \mathbb{N}} \frac{x_i}{N_{\mathbf{x}}} \geq \sum_{i \in \mathbb{N}} \frac{y_i}{N_{\mathbf{y}}}. \quad (6)$$

Proof. Apply the result from Corollary 4.5. ■

Corollary 4.6 Preferences \succcurlyeq over scientific units satisfy A1-A4, A5*, A7 and A8 if and only if there is increasing and concave real-valued mapping u from \mathbb{N}_0 such that, for all \mathbf{x} and $\mathbf{y} \in \mathcal{X}$,

$$\mathbf{x} \succcurlyeq \mathbf{y} \iff \sum_{i \in \mathbb{N}} \left[\frac{u(x_i)}{|\mathbf{x}|} - \frac{u(y_i)}{|\mathbf{y}|} \right] \geq 0. \quad (7)$$

If (6) holds, u can be replaced by a mapping v if and only if $v = \beta u + \tau$ for some $\beta, \tau \in \mathbb{R}$ with $\beta > 0$.

Proof. Apply the result from Corollary 4.6. ■

5 Relation between stochastic dominance and rankings

By identifying a citation distribution with a probability distribution,⁵ readers familiar with the theory of choice under risk will readily recognize the clear analogy between the impact function u and the Bernoulli function in expected utility theory.⁶ I will make use of this formal resemblance to borrow results on stochastic dominance from the economic literature on finance and poverty analysis. This transposition will allow for unit comparisons without actually selecting a specific impact function.

Formally, for any unit $\mathbf{x} \in \mathcal{X}$ with relative distribution of citations $f_{\mathbf{x}}$, let $F_{\mathbf{x}}(i) = \sum_{j=1}^i f_{\mathbf{x}}(j)$ be its cumulative distribution function. I say that \mathbf{x} first order stochastically dominates \mathbf{y} (and I write $\mathbf{x} \succ_{FSD} \mathbf{y}$) if and only if $D^1(i) = F_{\mathbf{x}}(i) - F_{\mathbf{y}}(i) \leq 0$ for all $i \in \mathbb{N}_0$. It then follows:

Proposition 5.1 $\mathbf{x} \succ_{FSD} \mathbf{y}$ if and only if $\mathbf{x} \succcurlyeq \mathbf{y}$ for all rankings satisfying A1-A4, A5*, and A8.

Proof. Let \mathbf{x} and $\mathbf{y} \in \mathcal{X}$ with $\mathbf{x} \neq \mathbf{y}$. Suppose that $\mathbf{x} \succ_{FSD} \mathbf{y}$. Theorem 4.4 guarantees that for any given ranking satisfying A1-A4, A5*, and A8, $\mathbf{x} \succcurlyeq \mathbf{y}$ if and only if $\sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) \geq \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i)$ for some u increasing. Now Abel's lemma⁷ yields, for $N^* = \max(|\mathbf{x}|, |\mathbf{y}|)$,

⁵ It is also possible to identify a citation distribution with an income distribution. This is the approach in Albarrán *et al.* (2009), which introduce into citation analysis a methodology borrowed from measurements of economic poverty.

⁶ Boyssou and Marchant (2010) exploit this similarity to analyze and offer an extension of the IF factor for journals.

⁷ Abel's lemma can be viewed as a discrete version of the formula of integration by parts. It states that given $x_1, \dots, x_n, y_1, \dots, y_n$ real numbers, then $\sum_{i=1}^n x_i y_i = \sum_{i=1}^{n-1} \left[\sum_{j=1}^i x_j \right] (y_i - y_{i+1}) + y_n \sum_{j=1}^n x_j$. (For an inductive proof, see, e.g., <http://planetmath.org/encyclopedia/ProofOfAbelsLemmaByInduction.html>).

$$\sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) - \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i) = \sum_{i=1}^{N^*} [f_{\mathbf{x}}(i) - f_{\mathbf{y}}(i)] u(i) = \sum_{i=1}^{N^*-1} D^1(i) [u(i) - u(i+1)] + D(N^*) u(N^*) = \sum_{i=1}^{N^*-1} D^1(i) [u(i) - u(i+1)] \geq 0.$$

To prove the “only if” part reasoning by contradiction, suppose $\mathbf{x} \succ_{FSD} \mathbf{y}$ does not hold. Then, there exists j such that $D^1(j) > 0$. Let us consider $u : \mathbb{N}_0 \rightarrow \mathbb{R}$ where $u(i) = u(i+1) \forall i \neq j$, and $u(j+1) > u(j)$. Since u is increasing, by Theorem 6.1 it induces a ranking \succ which satisfies A1-A4, A5*, and A7, thus by hypothesis $\mathbf{x} \succ \mathbf{y}$ and therefore, $\sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) - \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i) = D^1(j) [u(j) - u(j+1)] > 0$ a contradiction. ■

A more restricted class of rankings is given when in addition to A1-A4, A5* and A8, we impose A7. It turns out that all rankings of this more restricted class give an unanimous order between two units if one dominates the other at second order. Formally, for any \mathbf{x} and $\mathbf{y} \in \mathcal{X}$, let $D^2(i) = \sum_{j=1}^i D^1(j) = \sum_{j=1}^i [F_{\mathbf{x}}(j) - F_{\mathbf{y}}(j)]$. By definition, \mathbf{x} stochastically dominates \mathbf{y} at second order (written $\mathbf{x} \succ_{SSD} \mathbf{y}$) if $D^2(i) \leq 0$ for all arguments $i \in \mathbb{N}_0$. Then,

Proposition 5.2 $\mathbf{x} \succ_{SSD} \mathbf{y}$ if and only if $\mathbf{x} \succ \mathbf{y}$ for all rankings satisfying A1-A4, A5*, A7 and A8.

Proof. From Proposition 5.1, $\sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) - \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i) = \sum_{i=1}^{N^*-1} D^1(i) [u(i) - u(i+1)]$. I shall use again Abel’s lemma to get $\sum_{i=1}^{N^*-1} D^1(i) [u(i) - u(i+1)] =$

$$\begin{aligned} &= \sum_{i=1}^{N^*-2} \left[\sum_{j=1}^i D^1(j) \right] [u(i+1) - u(i) - u(i+2) + u(i+1)] + \left[\sum_{j=1}^{N^*-2} D^1(j) \right] [u(N^*) - u(N^*-1)] = \\ &= \sum_{i=1}^{N^*-2} D^2(i) [u(i+1) - u(i) - u(i+2) + u(i+1)] + D^2(N^*-1) [u(N^*) - u(N^*-1)]. \end{aligned}$$

It is clear that if $\mathbf{x} \succ_{SSD} \mathbf{y}$ then the expression above is positive for any u increasing and concave, and therefore by Theorem 4.4 $\mathbf{x} \succ \mathbf{y}$ for all rankings satisfying A1-A4, A5*, A7 and A8.

To see the converse, suppose $\sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) - \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i) \geq 0$ for any u increasing and concave. If there exists j such that $D^2(j) > 0$, then consider $u : \mathbb{N}_0 \rightarrow \mathbb{R}$ where $u(i) = u(i+1)$, $\forall i > j+1$, $u(j+1) > u(j)$ and $u(i+2) - u(i+1) = u(i+1) - u(i)$, $\forall i \leq j-1$. We have u increasing and concave and $\sum_{i \in \mathbb{N}_0} f_{\mathbf{x}}(i) u(i) - \sum_{i \in \mathbb{N}_0} f_{\mathbf{y}}(i) u(i) = D^2(j) [u(j) - u(j+1)] < 0$ which contradicts the hypothesis. ■

6 The connection between rankings and prospects

In the previous sections, uncertainty is not present. Therefore, the evaluator is assumed to know with certainty the impact of each of the publications of any given unit. However, as the propositions 5.1 and 5.2 show, there is a formal analogy between concepts expressed in terms of citation records and uncertainty counterparts. Indeed, it is possible to go beyond this formal resemblance and link both models as follows:

- (i) Let $\pi = (\pi_1, \dots, \pi_i, \dots)$ and $\pi^* = (\pi_1^*, \dots, \pi_i^*, \dots)$ be prospects (i.e. finite probability distributions, $\pi_i, \pi_i^* \in [0, 1]$, $\sum_{i \in \mathbb{N}} \pi_i = \sum_{i \in \mathbb{N}} \pi_i^* = 1$), and let x_1, \dots, x_i, \dots and $x_1^*, \dots, x_i^*, \dots$ be sequences of outcomes in \mathbb{N}_0 generated, respectively, by π and π^* .
- (ii) From the binary relation \succsim over \mathcal{X} , we can obtain utility indices $u_i = u(x_i)$, which are constant up to positive scale transformations, and such that for any given $k \in \mathbb{N}$ the sequence $\mathbf{x}_k = (x_1, \dots, x_k, -1, \dots)$ is at least as good as the sequence $\mathbf{x}_k^* = (x_1^*, \dots, x_k^*, -1, \dots)$ if and only if $\sum_{i=1}^k u_i(x_i) \geq \sum_{i=1}^k u_i(x_i^*)$.
- (iii) By the weak law of large numbers, random variables $Z_k = \frac{1}{k} \sum_{i=1}^k u(x_i)$ and $Z_k^* = \frac{1}{k} \sum_{i=1}^k u(x_i^*)$ converge in probability respectively to $\sum_{i \in \mathbb{N}_0} \pi_i u_i(x_i)$ and $\sum_{i \in \mathbb{N}_0} \pi_i^* u_i(x_i^*)$ as $k \rightarrow +\infty$.
- (iv) Given two prospects P and P^* giving prizes $u_i(x_i)$ and $u_i(x_i^*)$ with probabilities π_i and π_i^* respectively, the rule of maximizing the expected value prescribes choosing P over P^* whenever $\sum_{i \in \mathbb{N}_0} p_i u_i(x_i) > \sum_{i \in \mathbb{N}_0} p_i^* u_i(x_i^*)$.

Therefore, we can conclude $\lim_{k \rightarrow +\infty} \Pr[Z_k > Z_k^*] = 1$. But $\Pr[Z_k > Z_k^*] = \Pr[kZ_k > kZ_k^*] = \Pr\left[\sum_{i=1}^k u(x_i) > \sum_{i=1}^k u(x_i^*)\right] = \Pr[\mathbf{x}_k \succ \mathbf{x}_k^*]$. Thus, $\sum_{i \in \mathbb{N}_0} p_i u_i(x_i) > \sum_{i \in \mathbb{N}_0} p_i^* u_i(x_i^*)$ implies that $\lim_{k \rightarrow +\infty} \Pr[\mathbf{x}_k \succ \mathbf{x}_k^*] = 1$.

How can we interpret these results? The rationale underlying all citation analysis is that citation data should be strongly correlated such that a “good” unit has a far higher probability of publishing a highly cited paper than a “bad” unit. This hypothesis appears to be fulfilled in practice when units are authors,⁸ and seems also plausible in the case in which units are journals. I think likely that authors may consider that the past distribution of citations of papers published in a journal gives some information on the future selection process in this journal and its visibility in the scientific community. In such a case, the author’s risk attitudes may play a role in the choice of a journal to which submit a paper. Provided that the impact function u can be interpreted as a Bernoulli function⁹ and that the relative distribution of citations remains the same along the time, I show that both approaches are consistent: Namely, as the sample of papers increases, preferences satisfying A1-A4, A5*, and A8 approach the ranking obtained by applying the expected utility rule.

7 Wrapping up and suggestions for further research

At the end of the day, all measurements of research quality are hard to shallow; citation analysis, no matter how sophisticated it may be, cannot possibly be a perfect substitute for critical reading

⁸ Lehmann *et al.* (2003) offer significant correlations from data in high energy physics.

⁹ A word of caution: This is a rather strong assumption. In this paper, the utility function u is derived in a riskless framework. In expected utility theory, however, the Bernoulli function is derived from axioms regarding the ordering of probability distributions or uncertain prospects.

and expert judgement. However, the bottom line is that quality measures are here to stay, and will remain a hot topic in science during the coming years. Probably, not because of any truth waiting to be discovered. Rather, because sciences (and this includes economics) are subject to market forces and the internet has enabled a plethora of low-cost measures to rank research output. Schools and labs are using such rankings as a shorthand assessment of quality to help them make grants, bestow tenure, award bonuses or hire postdocs. An analogy which I have in mind when thinking about these measures is that they do for scientists what U.S. News & World Report does for Colleges and the QB rating does for American football.

So far, the merits of most measures rely largely on intuitive arguments and value judgements. In fact, several are often only vaguely related to the intuitive ideas they purport to index, and many are so complex that it is difficult to discover what, if anything, they are measuring. Admittedly, I have written this paper as a reaction to this flow of work. This work is unlikely to change the mind of those scientists whose disbelief in the benefits of influence measures is unshakable, but it may help those who are still undecided about these benefits by bringing economic methodology to bear on the ranking problem. Rather than identify a weakness on some existing ranking and proposing a new method which avoids the same weakness, I present a utility theory for citations records that allows to discuss rankings in terms of a set of preference constraints that entirely characterize a family of rankings. Whether or not this utility theory is better or worse than others remains to be seen. I am aware of the limitations of its applicability and I indicated this fact in Section 3 when I discussed the axioms. It is not my purpose to claim that a given method based on citation records is the correct way of measuring quality.¹⁰ However, without investigating characterizing properties, we will not have a complete picture of the situation.

The analysis of the discussed model may be extended and modified in numerous ways. Since Theorems 4.1 and 4.4 single out a cardinal utility function u for the number of citations which does not involve the ordering of uncertain prospects, the problem of the interpretation of such function as a Bernoulli function raises subtle questions. More concretely, if we suppose that the evaluator satisfies the axioms in Theorem 4.4 and the axioms necessary for an expected utility representation with Bernoulli function u^* , what is the relationship between u and u^* ? Since no uncertainty is involved in obtaining u , and, on the other hand, u^* is derived from the ordering of probability distributions, this relationship seems to capture the risk attitude of the evaluator with respect to citation records. On a tentative basis, if $u^* = \phi(u)$, it makes sense to say that the evaluator is risk neutral (respectively risk averse, risk loving) if ϕ is linear (respectively strictly concave, strictly convex). In such a case, only risk neutral behavior satisfies the convergence result in Section 6.

The connection between the problem of comparing distributions of citations and the problem of comparing probability distributions may give rise to new concepts and models in both fields. For

¹⁰ This objective is appearing more and more like an unattainable “holy grail” for the bibliometric literature. See for instance Adler *et al.* (2009).

instance, models that separate tastes from beliefs in decision under uncertainty may offer interesting approaches to cope with distributions of citations. I think this question deserves more analysis, and I hope to deal with it in another paper.

References

- ADLER, R., J. EWING, AND P. TAYLOR (2009): "Citation statistics," *Statistical Science*, **24** (1) 1-14.
- ALBARRÁN, P., I. ORTUÑO, AND J. RUIZ-CASTILLO (2009): "The measurement of low- and high-impact in citation distributions: technical results," Working Paper 09-57, Economics Series 35, Universidad Carlos III, forthcoming in the *Journal of Informetrics*.
- ALONSO, F.J. CABRERIZO, E. HERRERA-VIEDMA AND F. HERRERA (2009): "h-index: A review focused in its variants, computation and standardization for different scientific fields," *Journal of Informetrics*, **3**, 273-289.
- BOUYSSOU, D., AND T. MARCHANT (2010): "Bibliometric rankings of journals based on Impact Factors: An axiomatic approach," *Journal of Informetrics*, forthcoming.
- CAMACHO, A. (1979a): "Maximizing expected utility and the rule of long run success," in *Expected Utility and the Allais Paradox* (M. Allais and O. Hagen, Eds.) Dordrecht: D. Reidel.
- (1979b): "On cardinal utility," *Theory and Decision*, **10** (2) 131-145.
- (1980): "Approaches to cardinal utility," *Theory and Decision*, **12** (2) 359-379.
- (1982): *Societies and Social Decision Functions* Dordrecht: D. Reidel.
- EGGHE, L. (2006a): "An improvement of the h-index: The g-index," *ISSI Newsletter*, **2** 8-9.
- (2006b): "Theory and practice of the g-index," *Scientometrics*, **69**, 131-152.
- GARFIELD, E. (1972): "Citation analysis as a tool in journal evaluation," *Science*, **178**, 471-479.
- (2006): "The history and meaning of the journal impact factor," *Journal of the American Medical Association*, **295**, 90-93.
- HIRSCH, J.E. (2005): "An index to quantify an individual scientific research output," *Proceedings of the National Academy of Sciences*, **102** (46), 16569-16572.
- HARZING, A.W. (2009): "Publish or Perish, version 2.8.3644 ," available at <http://www.harzing.com/pop.htm>
- HARZING, A.W., AND R. VAN DER WAL (2008): "A Google Scholar h-index for journals: An alternative metric to measure journal impact in Economics & Business ," forthcoming in the *Journal of the American Society for Information Science and Technology*.
- LEHMANN, S., B.E. LAUTRUP, AND A.D. JACKSON (2003): "Citation networks in high energy physics," *Physical Review*, **E 68**, 026113.

- MARCHANT, T. (2009): “An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors,” *Scientometrics*, **80** (2) 325-342. DOI 10.1007/s11192-008-2075-y.
- PALACIOS-HUERTA, I., AND O. VOLIJ (2004): “The measurement of intellectual influence,” *Econometrica*, **72** (3) 963-977.
- PANARETOS, C., AND C. MALESIOS (2009): “Assessing scientific research performance and impact with single indices,” *Scientometrics*, **81** (3) 635-670. DOI 10.1007/s11192-009-k0026-x.
- QUESADA, A. (2009a): “Monotonicity and the Hirsch index,” *Journal of Informetrics*, **3** (2), 158-160.
- (2009b): “More axiomatics for the Hirsch index,” *Scientometrics*, DOI 10.1007/s11192-008-2174-9.
- RUANE, F., AND R.S.J. TOL (2008): “Rational (successive) h-indices: An application to economics in the republic of Ireland,” *Scientometrics*, **75** (3) 395-405.
- SLUTZKI, G., AND O. VOLIJ (2006): “Scoring of web pages and tournaments—axiomatizations,” *Social Choice and Welfare*, **26** (3) 75-92.
- WAKKER, P. (1986): “The Repetitions Approach to characterize cardinal utility,” *Theory and Decision*, **17** (1) 33-40.
- WOEGINGER, G.J. (2008a): “An axiomatic characterization of the Hirsch-index,” *Mathematical Social Sciences*, **56**, 224–232.
- (2008b): “A symmetry axiom for scientific impact indices,” *Journal of Informetrics*, **2**, 298-303.

This working paper is part of the DIGITUM open-access repository of the Universidad de Murcia.

Departamento de Fundamentos del Análisis Económico

Facultad de Economía y Empresa

E-30100 Campus de Espinardo

Murcia (SPAIN)

Tel.: (34) 868 883 784

E-mail: nsb@um.es

Website: www.um.es/analisiseco
