Scientific Research Measures

Marco Frittelli*
Department of Mathematics
University of Milan

Loriano Mancini†
Swiss Finance Institute
and EPFL

Ilaria Peri‡
ESC Rennes
School of Business

This version: June 24, 2013
First version: May 2012

***Corresponding author**: Marco Frittelli, Department of Mathematics, University of Milan, Via Saldini 50, I-20133 Milan, Italy. E-mail: marco.frittelli@unimi.it.

†Loriano Mancini, Swiss Finance Institute at EPFL, Quartier UNIL-Dorigny, CH-1015 Lausanne, Switzerland. E-mail: loriano.mancini@epfl.ch.

‡Ilaria Peri, Finance and Operations Department, ESC Rennes, 2 rue Robert d'Arbrissel CS 76522, F-35065 Rennes Cedex, France. E-mail: ilaria.peri@esc-rennes.fr.

Scientific Research Measures

Abstract

We introduce the novel class of Scientific Research Measures (SRMs) to rank scientists' research performance. In contrast to many bibliometric indices, SRMs take into account the whole scientist's citation curve, share sensible structural properties, allow for a finer ranking of scientists, fit specific features of different disciplines, research areas and seniorities, and include several bibliometric indices as special cases. We also introduce the further general class of Dual SRMs that allows for a more informed ranking of research performances based on "journals' value." An empirical application to 173 finance scholars' citation curves shows that SRMs can be easily calibrated to actual citation curves and produce different authors' rankings than traditional bibliometric indices.

Keywords: Bibliometric Indices, Citations, Risk Measures, Scientific Impact Measures, Calibration, Duality

JEL Codes: C02, I20

1. Introduction

Over the last decades, the importance of evaluating scientific research performance has grown exponentially. Crucial decisions such as funding research projects, faculty recruitment or academic promotion depend to a large extent upon the scientific merits of the involved researchers. Selecting inappropriate valuation criteria can have obvious and deleterious effects on performance assessment of scientists and structures, such as departments or laboratories, and distort funding allocation and recruiting process.

Two approaches have been proposed to assess research performance: content valuation, based on internal committees and external panels of peer reviewers, and context valuation, based on bibliometric indices, i.e., statistics derived from citations and characteristics of journals where the research output appeared. Context valuation is far less expensive in terms of time and resources involved than content valuation and can be easily carried out on a systematic base, e.g., yearly base. In contrast, content valuation can only be carried out on a multiple-year base and used to check or fine tune the evaluation based on bibliometric indices. Moreover, content valuation is unfeasible when a large number of scientists or research outputs need to be evaluated.

Because of the fast increase in availability and quality of online databases (e.g., Google Scholar, ISI Web of Science, Scopus, MathSciNet), context valuation has become widely popular and several bibliometric indices have been suggested to assess research performance. The most popular citation-based metric is the h-index introduced by Hirsch (2005). The h-index is the largest number such that h publications have at least h citations each. This index has received a large attention from the scientific community and has been widely used to assess research performance in many fields. Recently, various authors have proposed several extensions of the h-index.[1] Most of these indices have been introduced on an ad hoc basis to improve on some of the shortcomings of the h-index and its subsequent extensions.

This paper provides three contributions to the fast growing literature on bibliometric

[1]For an overview of these proposals see Section 2 below and, e.g., Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2009), Panaretos and Malesios (2009), and Schreiber (2010).

indices. The first contribution is to introduce a novel class of scientific performance measures based on citation curves[2] that we call Scientific Research Measures (SRMs). The main feature of SRMs is to take into account the whole scientist's citation curve and to assess her research performance using predetermined performance curves. The performance curves determine the minimum threshold of citations to reach a certain performance level. Importantly, such performance curves can be chosen in a flexible way, for example reflecting the seniority of the scientists and characteristics of the specific research field, such as typical number of publications and citation rates. This flexibility is absent in virtually all existing bibliometric indices. SRMs are derived from a "calibration" approach in the sense that SRMs are informed by actual citation data. Under the premise that research performance is reflected by the whole citation curve, rather than by only part of it, SRMs provide by construction a better assessment of research performance than standard bibliometric indices.

SRMs have two desirable properties, namely monotonicity and quasi-concavity. The first property simply implies that better scientists, as reflected by their citation curves, have higher research performance measures. The second property, quasi-concavity, implies that when a given scientist is combined with a better scientist (in the sense of a convex combination of their citation curves), the performance of the two is higher than the performance of the initial scientist. Quasi-concavity is not only a sensible property of SRMs. It also allows to rank scientists whose citation curves intersect each other (see Appendix A), as is often the case in practice. In those cases, the monotonicity property alone would not lead to any ranking.

SRMs subsume many existing bibliometric indices as special cases. Woeginger (2008) provides an earlier attempt to collect bibliometric indices in a general class. However, there are significant differences between the two approaches. SRMs allow for a granular ranking of scientists' research performances, share sensible structural properties, and appear to be more flexible as SRMs reflect the seniority of scientists and specific features of the various disciplines.

[2] The citation curve is the number of citations of each publication, in decreasing order of citations.

The second contribution of this paper is to introduce a further general class of research measures called Dual SRMs. We provide a rigorous and axiomatic-based theoretical foundation of the Dual SRMs. The novelty is to describe researchers' performance using citation records, rather than citation curves. A citation record associates to each journal the citations of the researcher's publications in that journal. Then we draw an analogy between citation records, viewed as random variables defined on journals, and risky financial payoffs. This analogy allows us to formalize the mathematical properties of Dual SRMs by building on the well-established theory of risk measures. The axiomatic approach developed in the seminal paper by Artzner, Delbaen, Eber, and Heath (1999) has been very influential for the theory of risk measures.[3] Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002) extended the theory of (coherent) risk measures to the class of convex risk measures. The origin of our proposal of Dual SRMs can be traced back to the most recent developments of this theory, leading to the notion of quasi-convex risk measures introduced by Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2011) and further developed in the dynamic framework by Frittelli and Maggis (2011).[4]

We establish a duality between the primal space of citation records and its dual space given by the "value" of each journal. A journal's value could be measured for example using the impact factor of the journal or any other criterion deemed sensible by the evaluator who is assessing the scientists' research performances. Multiple criteria are possible which may induce multiple values of a given journal.[5] In general, there will be no agreement on the

[3]Rather than focusing on a specific measurement of risk carried by financial positions (e.g., variance of asset returns), Artzner, Delbaen, Eber, and Heath (1999) proposed a class of measures satisfying sensible properties, so-called coherent axioms. Ideally, each financial institution could select its own risk measure, provided it obeyed the structural coherent properties. This approach added flexibility in the selection of the risk measure and, at the same time, established a unified framework. In the same spirit, we introduce Dual SRMs to assess research performance, rather than financial risk.

[4]Among other studies in this area, Cherny and Madan (2009) introduced the concept of an acceptability index having the property of quasi-concavity, and Drapeau and Kupper (2013) analyzed the relation between quasi-convex risk measures and associated family of acceptance sets, already present in Cherny and Madan (2009).

[5]To further elaborate on the analogy between Dual SRMs and risk measures, the journals are the states of nature and each journal's value is the price of the Arrow–Debreu security associated to the given state. Multiple values of the journal are multiple prices of the Arrow–Debreu security, which may occur when financial markets are incomplete.

journals' values. Notably, the Dual SRM provides a prudential or conservative assessment of the scientist's research performance, i.e., under the least favorable journal's evaluation criterion.

Dual SRMs are relevant for three reasons. First, if two scientists have the same citation curve, any traditional bibliometric index would not be able to rank them. In contrast, Dual SRMs can rank those scientists as soon as their publications appeared in different journals. Second, additional information beyond the citation curve (such as journals' values and the evaluator's assessment of the corresponding criteria) can be taken into account in the rankings based on Dual SRMs. Consequently, Dual SRMs induce a more informed ranking than existing bibliometric indices based on citation curves. Third, and perhaps most importantly, Dual SRMs allow to achieve a meaningful aggregation of scientists' research outputs, by aggregating scientists' citation records. Thus, Dual SRMs allow to rank the research performances of teams, departments and academic institutions. The issue of ranking research institutions on the basis of their research output has been a central theme in the scientific community over the last decades, e.g., Borokhovich, Bricker, Brunarski, and Simkins (1995), and Kalaitzidakis, Stengos, and Mamuneas (2003).

The third contribution of this paper is to present an extensive empirical application of SRMs. Using Google Scholar, we construct a novel dataset of citation curves of 173 full professors affiliated to 11 main U.S. business schools or finance departments. Our empirical results show that the SRM characterized by hyperbolic-type performance curves describes well actual citation curves. We compare authors' and universities' rankings based on the SRM and other 7 existing bibliometric indices. It appears that rankings based on the SRM and the other bibliometric indices are generally quite different. The largest discrepancies in rankings concern a large portion of "average" performing scholars and universities (in relative terms), which are arguably the most difficult to rank, rather than the best and least performing scholars and universities. We recall that when research performance is reflected by the whole citation curve, rather than by only part of it, our SRMs provide a more sensible ranking than standard bibliometric indices. Finally, we provide clear operational guidance on how to implement our SRMs and develop a simple two-step algorithm to compute SRMs based on hyperbolic-type performance curves.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the SRMs. Section 4 presents the Dual SRMs. Section 5 discusses the empirical analysis. Section 6 concludes. Appendices collect technical material.

2. Related Literature

At least since the early nineties, the use of bibliometric indices have encompassed many fields, e.g., Chung and Cox (1990), May (1997), Hauser (1998), Wiberley (2003), and Kalaitzidakis, Stengos, and Mamuneas (2003). Indeed, the effectiveness of citation-based metrics is validated by correlation analyses with peer-review ratings, e.g., Lovegrove and Johnson (2008). Moreover, when a large number of scientists or research outputs need to be evaluated, bibliometric indices are essentially the only viable method.[6]

Bibliometric indices were initially introduced either to merely quantify the production of researchers, e.g., total number of publications in a given period, or to assess the impact of their publications, e.g., total number of citations. To obtain a more comprehensive evaluation of the scientist's research performance, Hirsch (2005) introduced the h-index, which is nowadays the most popular citation-based metric.

The main achievement of the h-index is to combine scientific production (given by the number of publications) and quality of research (given by the number of citations) in a single bibliometric index. The index is also simple to compute.[7] Moreover, Hirsch (2007) shows that his index predicts future individual scientific achievements more than other bibliometric indices. The main drawback of the h-index is the insufficient ability to discriminate between

[6]The Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) is currently undertaking a large evaluation process that requires the assessment of more than 200,000 research items (journal articles, monographs, book chapters, etc.) published by all Italian researchers between 2004 and 2010 in 14 different fields. Each panel, one for each field, consists of 30 members and has about 15,000 research items to evaluate in less than one year. All panels, except those in humanistic areas, decided to use bibliometric indices.

[7]For example using ISI Web of Science, http://www.webofknowledge.com, the h-index can be easily computed by ordering the scientist's publications by the field 'Time Cited.'

scientists' research performances. To illustrate this point, suppose an author has 10 publications with 10 citations each, and thus an h-index of 10. She has the same h-index as a second author who has 100 publications but only 10 of them received 10 citations, and a third author who has only 10 publications but with 100 citations each. In other words, the h-index does not reflect publications with high and low number of citations, which may well contain valuable information for assessing the scientific performance. The h-index is fully determined by the Hirsch core, which is the set of the most cited publications with at least h citations each. In the example above, the Hirsch core is 10 publications. The report by the Joint Committee on Quantitative Assessment of Research[8] (Adler, Ewing, and Taylor, 2008) strengthened this point by a well-documented example and encouraged the use of more complex measures. The report also emphasized the lack of mathematical analysis of the properties of the h-index.[9]

To overcome the limitations of the h-index, a number of scientists from different fields have proposed a wide range of indices to measure scientific performances. We classify the numerous proposals in two broad categories: h-index variants, that suggest alternative methods to determine the core size of the most cited publications, and h-index complements, that include additional information in the h-index especially concerning the further citations of the most cited publications.

Among the h-index variants, one of the first proposals is the g-index by Egghe (2006). The g-index is defined as the largest number g such that the most cited g articles received together at least g^2 citations. Following the same logic, Kosmulski (2006) suggests the h^2-index defined as the largest number h such that the h most cited publications receive at least h^2 citations each. Relative to the h-index, the h^2-index is probably more appropriate in fields where the number of citations per article is relatively high, like chemistry or physics, while the h-index seems to be more suitable, say, for mathematics or astronomy. Many studies

[8]See http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf.

[9]As any other bibliometric index, the h-index presents other limitations concerning for example the issue of self citations, the number of co-authors and the accuracy of bibliometric databases. Recently, various indices have been proposed to alleviate these limitations. However, these indices are based on transformations of the original citation data, raising new issues concerning the best practice of carrying out such transformations.

point out the arbitrariness of the definition of the citation core size. To alleviate this issue, van Eck and Waltman (2008) propose the h_a-index defined as the largest natural number h_a such that the h_a most cited publications received at least $a\,h_a$ citations each. A further variant is the w-index introduced by Woeginger (2008), and defined as follows: a scientist has index w if w of her publications have at least $w, w-1, \ldots, 1$ citations. Ruane and Tol (2008) propose the rational h-index (h_{rat}-index) whose advantage is to provide more granularity in the evaluation process since it increases in smaller steps than the original h-index. Guns and Rousseau (2009) review several real and rational variants of the h- and g-indices.

Rather than concentrating on the size of the citation core, the h-index complements seek to measure the core citation intensity. For example, Jin (2006) proposes to compute the average of the Hirsch core citations (A-index). Subsequently, to reduce the penalization of the best scientists receiving many citations, Jin, Liang, Rousseau, and Egghe (2007) modify the A-index by taking the square root of the Hirsch core citations (R-index). The same approach is used by Egghe and Rousseau (2008) but they consider as citation core only a subset of the Hirsch core (h_w-index). Both the R- and h_w-index can be very sensitive to just a few publications with many citations. In an attempt to make these indices more robust, Bornmann, Mutz, and Daniel (2008) propose to calculate the median, rather than the mean, of the citations in the Hirsch core (m-index). The distribution of citations is often positively skewed, and the suggestion of using the median appears to be a sensible proposal. An interesting h-index complement is the tapered h-index, proposed by Anderson, Hankin, and Killworth (2008), which attempts to take into account all citations, not just those in the Hirsch core. Instead, Vinkler (2009) proposes the π-index defined as one hundredth of the total citations obtained by the most influential publications, which are defined as the square root of the total publications ranked in decreasing order of citations. Zhang (2009) suggests a complementary index to the h-index to summarize the excess citations in the Hirsch core by taking their square root (e-index). Further proposals to attenuate the drawbacks of the h-index include the geometric mean of the h- and g-index (hg-index by Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2010)) and the geometric mean of the h-

and m-index (q^2-index by Cabrerizo, Alonso, Herrera-Viedma, and Herrera (2010)).

The main consideration that emerges from this literature review is that most bibliometric indices have been proposed on an ad hoc basis to remedy specific issues related either to the core size of publications or to the impact of such core publications. Moreover, as pointed out by Schreiber (2010), many of these indices are highly correlated and lead to very similar rankings.

We depart from the literature by adopting a "calibration" approach to assess research performances, rather than improving existing bibliometric indices. The main motivation for our approach is that scientists of different fields and seniorities can have significantly different citation curves. These differences are important and need to be reflected in the research evaluation process. Our class of SRMs can take into account those specific features using flexible performance curves that are calibrated on the data, as discussed in the next section.

3. Scientific Research Measures

In this section we introduce our class of Scientific Research Measures based on citation curves. A scientist's *citation curve* is the number of citations of each publication, in decreasing order of citations. Figure 1 shows a citation curve of a hypothetical scientist with 8 publications. As discussed in the previous section, many existing bibliometric indices are based on citation curves.

Let X denote the citation curve of an author. We model X as a function $X : \mathbb{R}_+ \to \mathbb{R}_+$, mapping each publication into the corresponding number of citations. By construction, X is bounded, decreasing, and taking only a finite number of values on $\mathbb{R}_+$. As the number of publications and citations are positive integers, the domain and range for X appear to be the positive integers,[10] rather than the positive reals. However, in the research evaluation

[10] The theory of SRMs developed in this section also holds when the domain and range of X are positive integers. To study the properties of SRMs, we only require X to be bounded and decreasing; for instance X is never required to be continuous.

process, the domain and range of X need not be the positive integers. For example, if a publication is coauthored by two or more scientists, the evaluator may choose to count that publication as half or 1/(number of coauthors), when assessing the research performance of one of the coauthors. The evaluator may also choose to give more weight to citations of research papers, rather than review papers or books, inducing the range of X to be the positive reals, rather than positive integers. Thus, we present the theory of SRMs embedding X in the positive reals.

To assess a scientist's research performance, we compare the whole citation curve X with a theoretical performance curve f_q. The performance curve represents the required amount of citations to reach a performance level q. Intuitively, the higher the scientist's research performance, the higher the performance curve f_q she can reach, and the higher the corresponding level q is. Formally, the *performance curve* is a function $f_q : \mathbb{R}_+ \to \mathbb{R}$ that associates to each publication x a theoretical number of citations $f_q(x)$. The performance curve f_q is assumed to be increasing in q, i.e., if $q \geq r$ then $f_q(x) \geq f_r(x)$ for all x. This means that the higher the q the higher the research performance level associated to the corresponding curve f_q. Although not necessary for the theory, we assume that $f_q(x)$ is decreasing in x, as any citation curve.

When q varies in a given set $\mathcal{I} \subseteq \mathbb{R}_+$, we obtain a family of performance curves $\{f_q\}_{q\in\mathcal{I}}$. Typically, we will consider $\mathcal{I} = \{1, 2, 3, \ldots\} \subset \mathbb{N}$ or $\mathcal{I} = [1, \infty]$. Figure 1 shows two families of performance curves: square-type functions corresponding to the h-index and hyperbolic-type functions corresponding to a specific SRM that we will use in Section 5. Recall that the higher the scientist's research performance, the higher her citation curve X is. Moving from the origin $(0, 0)$ of the graph to direction North-East the performance level increases, i.e., each performance curve f_q is associated to a higher and higher level q. The 4 performance curves in Figure 1 associated to the h-index correspond to an h-index of 1, 2, 3 and 4, respectively. Similarly, each hyperbolic-type performance curve is associated to an increasing level of research performance.

We now introduce the novel class of SRMs, which is defined by the following ϕ-index:

$$\phi(X) := \sup\{q \in \mathcal{I} \mid X(x) \geq f_q(x), \forall x \in [1, \infty)\} \tag{1}$$

where := denotes definition. The ϕ-index is obtained by comparing the actual citation curve X and the family of performance curves $\{f_q\}_{q\in\mathcal{I}}$. Specifically, the ϕ-index is the highest level q of the performance curve still below the author's citation curve. Consider the hyperbolic-type performance curves in Figure 1. The ϕ-index of that citation curve is 24.75 and it is associated to the performance curve passing through the point $(3, 5)$. In general, the numerical value of the ϕ-index (24.75 in the above example) has no interpretation. The numerical values of the ϕ-index of various scientists are simply used to rank their research performances. However, the ϕ-index can be normalized in order to attach some interpretation to it.[11]

The family of performance curves $\{f_q\}_{q\in\mathcal{I}}$ plays a key role in the definition of SRMs. Performance curves create the common ground or "level the playing field" to assess research performances and rank scientists. Under the premise that scientist's research performances are reflected by the whole citation curves, performance curves should have a similar shape as citation curves.

If performance curves are unrelated to citation curves, then it is unclear how to interpret the resulting scientists' rankings. Moreover, scientists of different research fields and seniorities have obviously different citation curves. The performance curves in equation (1) can be specified in a flexible way in order to reflect specific features of the scientific field, research areas and seniority. Unfortunately, most existing bibliometric indices (Table 1 and Figure 2 provide some examples) do not reflect the shape of actual citation curves. Figure 1 suggests that hyperbolic-type performance curves provide a much better description of the citation

[11] As discussed in Section 5, the performance curves in Figure 1 are given by $f_q(x) = q/x^c - k$. An equivalent representation of these performance curves is $f_q(x) = \alpha q/x^c - k$. The parameter α can be used, for example, to normalize to 100 the ϕ-index of the scientist with the largest number of publications or citations, or any other scientist. This normalization is achieved replacing q by αq in the two-step algorithm developed in Section 5 and determining α using the citation curve of the scientist with the largest number of publications or citations.

curve X than square-type performance curves. As shown in our empirical application in Section 5, actual citation curves exhibit hyperbolic-type shapes, calling for hyperbolic-type performance curves.

An additional argument supporting performance curves that mimic actual citation curves is the following. In Figure 1, the hypothetical author has an h-index of 4. Even if the 4 most cited publications, i.e., the publications in the so-called Hirsch core, receive more citations, by definition the h-index does not change. In contrast, the ϕ-index based on hyperbolic-type curves may increase, depending on the specific shape of the performance curves and reflecting the higher research performance of the scientist.[12] A similar situation occurs when the least cited publications have a modest increase in the number of citations. These additional citations do not increase the h-index but may increase the ϕ-index.

The definition of the ϕ-index in equation (1) is based on performance curves. An equivalent approach to define the ϕ-index is based on performance sets. To visualize these sets, consider again Figure 1. Intuitively, a performance set is the area above a given performance curve. Formally, given the performance curve f_q, the associated performance set is $\mathcal{A}_q := \{X : \mathbb{R}_+ \to \mathbb{R}_+ \mid X(x) \geq f_q(x), \forall x \in [1, \infty)\}$. Since there exists a performance set associated to each performance curve, when q varies in the set $\mathcal{I} \subseteq \mathbb{R}_+$, we obtain a family of performance sets $\{\mathcal{A}_q\}_{q\in\mathcal{I}}$, which parallels the family of performance curves $\{f_q\}_{q\in\mathcal{I}}$. By definition, $\mathcal{A}_q$ is monotone[13] and convex for any q and the family of performance sets $\{\mathcal{A}_q\}_{q\in\mathcal{I}}$ is monotone decreasing. Thus, the ϕ-index in (1) can be equivalently defined as $\phi(X) := \sup\{q \in \mathcal{I} \mid X \in \mathcal{A}_q\}$.

[12]Other bibliometric indices, such as A- and R-index reviewed in Section 2 and summarized in Table 1, increase when the number of citations in the Hirsch core increases. However, A- and R-index are not SRMs because they do not take into account the whole scientist's citation curves and are not characterized by performance curves, making their applications to scientists of different fields and seniorities non-trivial.

[13]Monotonicity of $\mathcal{A}_q$ means that if $X_1 \in \mathcal{A}_q$ and $X_2 \geq X_1$ then $X_2 \in \mathcal{A}_q$. Convexity of $\mathcal{A}_q$ means that if $X_1, X_2 \in \mathcal{A}_q$ then $\lambda X_1 + (1-\lambda)X_2 \in \mathcal{A}_q, \forall \lambda \in [0,1]$. The family of performance sets $\{\mathcal{A}_q\}_{q\in\mathcal{I}}$ is monotone decreasing because if $q \geq r$ then $\mathcal{A}_q \subseteq \mathcal{A}_r$.

3.1. *Existing Bibliometric Indices as Scientific Research Measures*

The class of SRM defined in (1) includes many existing bibliometric indices. Different indices can be recovered by suitably specifying the class of performance curves. For example, the h-index is characterized by the performance curve $f_q(x) = q1_{(0,q]}(x)$, where $1_{(0,q]}(x)$ takes value 1 when $x \in (0, q]$ and zero otherwise. The h^2-, h_α- and w-index, reviewed in Section 2, fall in the class of SRMs and the respective performance curves are reported in Table 1. The maximum number of citations, as a bibliometric index, is a SRM whose performance curve is given by $f_q(x) = q1_{(0,1]}(x)$. Similarly, the number of publications with at least one citation is a SRM whose performance curve is given by $f_q(x) = 1_{(0,q]}(x)$. The rational and real h-index, also reviewed in Section 2, fall in the class of SRMs. The corresponding performance curves are the same as those of the h-index, but the q parameter varies in the rational and real numbers, respectively, rather than in the natural numbers.

In general, any bibliometric index that can be represented using performance curves is a SRM. Not all bibliometric indices proposed in the literature are SRMs. The ϕ-index in (1) has three features: it takes into account the whole citation curve, it is defined in terms of performance curves and, as discussed in the next section, it is monotone and quasi-concave. Any bibliometric index that does not possess any of these three features is not a SRM. For example, the g-, A-, R- and m-index, reviewed in Section 2, are not SRMs. The reason is that these indices do not take into account the whole scientist's citation curve and thus cannot be represented in terms of performance curves as in (1).[14] The advantage of using flexible performance curves is to reflect in the research evaluation process specific features of the scientific field and seniority of scientists, and to allow for a more granular ranking of scientists. Without resorting to flexible performance curves this task appears to be quite challenging to achieve.

Figure 2 shows six families of performance curves associated to six different bibliometric indices, namely the h-, h^2-, h_α- and w-index, the maximum number of citations, and the

[14]The tapered h-index takes into account the whole scientist's citation curve but it does so in an involved manner and it is not defined in terms of performance curves.

maximum number of publications with at least one citation; see Table 1 for a short description of the indices. As citation curves have typically hyperbolic shapes, none of these performance curves provides an adequate description of actual citation curves, making rankings of research performances based on such bibliometric indices difficult to interpret.

3.2. Monotonicity and quasi-concavity of SRMs

The class of SRMs defined in (1) has two desirable properties, namely monotonicity and quasi-concavity. The first property simply implies that better scientists, as reflected by their citation curve, have higher ϕ-index. Specifically, if Scientist 1 has a lower research performance than Scientist 2, i.e., $X_1 \leq X_2$, then $\phi(X_1) \leq \phi(X_2)$. Thus, monotonicity of the ϕ-index is well justified.[15]

If SRMs were characterized by monotonicity only, they would have been of limited use. The reason is that the monotonicity property would only allow to rank scientists whose citation curves satisfy a dominance order, like $X_1 \leq X_2$ in the previous example. In many cases, the evaluator would need to rank scientists whose citation curves are not ordered or intersect each other. Quasi-concavity of the ϕ-index allows to rank scientists in such more complex but realistic situations. Formally, quasi-concavity is expressed as follows: given two citation curves X_1 and X_2, quasi-concavity of the ϕ-index implies that, for all $\lambda \in [0, 1]$,

$$\phi(\lambda X_1 + (1 - \lambda)X_2) \geq \min(\phi(X_1), \phi(X_2)). \tag{2}$$

Quasi-concavity implies that a convex combination of two citation curves leads to a performance level which is at least as large as the research performance of the less performing scientist. In other words, when the citation curve of a less performing scientist is combined with the citation curve of a better performing scientist, the resulting research performance is at least as good as that of the less performing scientist. This appears to be a natural

[15] As pointed out by an anonymous referee, other criteria of monotonicity could be adopted. For example, one could adopt the following weak form of majorization: $X_1 \preceq X_2 \Leftrightarrow \sum_{i=1}^{n} X_1(x_i) \leq \sum_{i=1}^{n} X_2(x_i)$, for all $n \in \mathbb{N}$; see Marshall, Olkin, and Arnold (2011). This criterion would lead to a quite different theory of SRMs and we defer this topic to future research.

property of the SRM. Appendix A shows that the ϕ-index in (1) satisfies (2), i.e., it is quasi-concave. Moreover, quasi-concavity ensures an internal consistency or coherency of scientists' rankings. Appendix A provides a numerical example to illustrate this point.

4. Dual Scientific Research Measures

In this section we introduce a further general class of research measures, called Dual SRMs. To see the importance of this class, consider the following example. Two scientists have two publications each and the same total number of citations. The first scientist has 10 citations from her publication in journal A and zero citations from her publication in journal B. The second scientist is in the opposite situation, i.e., she has zero citations from her publication in journal A and 10 citations from her publication in journal B. As the distribution of citations is the same for both scientists, any bibliometric index based on citation curves would imply that the two scientists have the same research performance. If journal A is more prestigious than journal B, then the first scientist has a higher research performance than the second scientist. The Dual SRM allows to discriminate the research performance of the two scientists and rank them correctly. The reason is that the Dual SRM evaluates the scientist's *citation record*. The citation record is the random variable that associates to each journal the citations of the scientist's publications in that journal.[16] Citation records contain significantly more information than citation curves. Consequently, the Dual SRM allows for a more informed ranking of research performances. Of course, to implement this approach a richer dataset for each scientist is needed, which includes citations, in which journals publications appeared, and quality of the journals.

The Dual SRM also allows to aggregate scientists' research outputs in a sensible way. The reason is that the Dual SRM aggregates citation records, i.e., citations of publications of different scientists appeared in the same journal, not citation curves. This in turn allows to rank research teams, departments and academic institutions in a sensible way, which has

[16]Any scientist may have published more than one paper in the same journal and Dual SRMs account for this situation, as discussed in Section 4.2.

been a central theme in the scientific community over the last decades.

We first provide a numerical example and then present the theory of Dual SRMs.

4.1. Numerical Example of Dual SRMs

To illustrate the derivation of the Dual SRM we use a simple numerical example. We consider an evaluator who needs to rank scientists' research performances.

The first step is to select the evaluation criteria of the journals. The evaluator needs to quantify the "value" of the journals in which the publications appeared. This is obviously a challenging task, but it is also a necessary step to achieve a sensible ranking of scientists with the same or similar distributions of citations, and publications in different journals.

Consider the following three journals: Management Science (MS), Mathematical Finance (MF) and Finance and Stochastics (FS). One possibility to assess a journal's value is to use the 1-year and 5-year journal's impact factor.[17] For MS, MF and FS, the 1-year impact factors are 1.73, 1.25, and 1.15, and the 5-year impact factors are 3.30, 1.66, and 1.58, respectively.[18] The information in each impact factor is collected in a probability measure, Q_1 and Q_5, respectively, on the journal space $\Omega = \{\mathrm{MS}, \mathrm{MF}, \mathrm{FS}\}$. For example, $Q_1(\mathrm{MS}) = 1.73/(1.73 + 1.25 + 1.15)$, and similarly for the other journals and impact factor.

Consider a scientist with three publications, one in each journal, MS, MF, and FS, with 10, 6, and 8 citations each. These citations are collected in her citation record $X(\omega)$, where $\omega = \mathrm{MS}, \mathrm{MF}, \mathrm{FS}$. Note that the citation record is not necessarily a citation curve, as values of $X(\omega)$ need not be in decreasing order. In the current example these values are 10, 6, and 8, respectively.

The second step is to assess the scientist's research performance under each criterion or probability measure Q_1 and Q_5. Under Q_1 the research performance of a scientist with

[17] Other criteria are obviously conceivable to assess journals' quality, such as overall citations of the journal or acceptance rate of journal submissions. Such additional criteria can be easily accounted for in the Dual SRM. For illustration purposes, we only consider two impact factors for each journal in this example.

[18] The impact factors were obtained from the website of each journal in 2012. The impact factor was devised in the mid 1950's by Eugene Garfield. See Garfield (2005) for a recent account of the impact factor.

citation record $X(\omega)$ is defined as

$$\beta(Q_1, X) := \sup\{q \in \mathbb{R} \mid E_{Q_1}[X] \geq \gamma(Q_1, q)\}. \tag{3}$$

Using citation record and impact factor, the value of $E_{Q_1}[X]$ can be easily computed. In the example $E_{Q_1}[X] = 10\ Q_1(\text{MS}) + 6\ Q_1(\text{MF}) + 8\ Q_1(\text{FS}) = 8.2$.

By definition, $\gamma(Q_1, q)$ represents the smallest Q_1-average of citations to reach a performance level of q. Intuitively, if the evaluator deems the criterion Q_1 as highly important, then $\gamma(Q_1, q)$ will be large, and it will be a challenging task for each scientist to do well under that criterion. Moreover, if Q_1 is more important than Q_2, then $\gamma(Q_1, q) > \gamma(Q_2, q)$.

Suppose the evaluator is indifferent between the criteria, Q_1 and Q_5. This ranking of the journal evaluation criteria can be expressed by a constant $\gamma(Q, q)$ function, i.e., $\gamma(Q, q) := q$, for any Q. Thus, under Q_1 the assessment of the scientist's research performance is

$$\beta(Q_1, X) := \sup\{q \in \mathbb{R} \mid E_{Q_1}[X] \geq q\} = E_{Q_1}[X] = 8.2.$$

If measure Q_1 was the only measure to assess journal quality, then $\beta(Q_1, X) = 8.2$ was the scientist's final research performance. Since measure Q_5 is another journal criterion, the evaluator computes the scientist's research performance also under Q_5, obtaining a corresponding index of $\beta(Q_5, X)$. In our example, $E_{Q_5}[X] = 10\ Q_5(\text{MS}) + 6\ Q_5(\text{MF}) + 8\ Q_5(\text{FS}) = 8.5$. As $\gamma(Q_5, q) := q$ constant, $\beta(Q_5, X) = E_{Q_5}[X] = 8.5$.

Finally, the evaluator takes the smallest value between $\beta(Q_1, X)$ and $\beta(Q_5, X)$. The Dual SRM gives the scientist's final research performance as

$$\Phi(X) := \min(\beta(Q_1, X), \beta(Q_5, X)) = \min(E_{Q_1}[X], E_{Q_5}[X]) = 8.2.$$

The example shows that the Dual SRM leads to a conservative or "worst case scenario" assessment of the research performance. In other words, the evaluator assess the scientist's research performance using the least favorable journal's evaluation criterion. In general, there will be no consensus on the journals' evaluation criteria. Taking a conservative assessment appears to be a natural approach. This is a key feature of the Dual SRM and will be further discussed below. To assess the research performance of another scientist the evaluator repeats the calculations above using her citation record and the same Q_1 and Q_5 measures. Comparing the scientists' Φ-indices gives the scientists' ranking.

In the example above, the evaluator is indifferent between the criteria Q_1 and Q_5. We now extend the example by considering an additional journal evaluation criterion and introducing a new evaluator who assigns different weights to the three criteria. The additional criterion is the number of years since the first journal's edition and expressed by the probability measure Q_0.[19] Under Q_0, $E_{Q_5}[X] = 10\ Q_0(\mathrm{MS}) + 6\ Q_0(\mathrm{MF}) + 8\ Q_0(\mathrm{FS}) = 8.8$.

Suppose the evaluator assigns different weights to the three criteria. This decision can be expressed by choosing the $\gamma(Q, q)$ function as follows

$$\gamma(Q, q) := q\,\alpha(Q) \tag{4}$$

where $\alpha(Q) \geq 0$ represents the weight that the evaluator assigns to each criterion Q. A natural range for the weight appears to be the interval $[0, 1]$, but weights larger than 1 are also admissible. The more important the journal's evaluation criterion (according to the evaluator), the higher the $\alpha(Q)$ weight. If a particular criterion is controversial or deemed unimportant, it can receive a weight of zero.

Table 2 presents different cases. Evaluator 2 believes that criterion Q_5 is more important than criterion Q_1 which in turn is more important than criterion Q_0. Accordingly, this evaluator assigns decreasing weights $\alpha(Q)$ to the three criteria. Evaluators 3 and 4 have different rankings for the criteria, reflected in the respective weights. Evaluator 1 is indifferent among all the three criteria, and assigns weight 1 to all of them. In the initial example, the evaluator used only Q_1 and Q_5 and was indifferent between the two criteria, and thus assigned $\alpha(Q_1) = \alpha(Q_5) = 1$ and $\alpha(Q_0) = 0$. Clearly, $\gamma(Q, q) := q$ can be recovered by

[19] The first edition of MS, MF and FS appeared in 1954, 1991 and 1996, respectively, which implies $Q_0(\mathrm{MS}) = (2012 - 1954)/95 = 0.61$, $Q_0(\mathrm{MF}) = (2012 - 1991)/95 = 0.22$ and $Q_0(\mathrm{FS}) = (2012 - 1996)/95 = 0.17$.

setting $\alpha(Q) = 1$ in (4). The key aspect of the $\gamma(Q, q)$ function is that it is selected by the evaluator ex-ante and independently of any scientist's citation record.

As in the previous example, the evaluator determines the scientist's research performance under the different criteria Q_0, Q_1 and Q_5. Under Q_0 the research performance of a scientist with citation record $X(\omega)$ is now defined as

$$\beta(Q_0, X) := \sup\{q \in \mathbb{R} \mid E_{Q_0}[X] \geq q\, \alpha(Q_0)\} = \frac{E_{Q_0}[X]}{\alpha(Q_0)} \tag{5}$$

and similarly under Q_1 and Q_5.

The Dual SRM that provides the scientist's final research performance is

$$\Phi(X) := \min(\beta(Q_0, X), \beta(Q_1, X), \beta(Q_5, X)) = \min\left(\frac{E_{Q_0}[X]}{\alpha(Q_0)}, \frac{E_{Q_1}[X]}{\alpha(Q_1)}, \frac{E_{Q_5}[X]}{\alpha(Q_5)}\right). \tag{6}$$

Different weights $\alpha(Q)$ lead to different Dual SRMs. Table 2 shows the research performances of the hypothetical scientist when different evaluators use different Dual SRMs, i.e., assign different weights to the journal evaluation criteria.

As mentioned above, quantifying journals' quality is a necessary step to rank scientists with publications in different journals. A priori there will be no consensus on journals' evaluation criteria. Taking the minimum in (6) naturally translates the evaluator's prudential assessment of the scientist's research performance. Furthermore, there also exists a theoretical reason for taking the minimum in (6) that comes from the duality theory and is discussed in the next section.

4.2. Theoretical Derivation of Dual SRMs

In this section we present the theory of the Dual SRM, which requires additional mathematical structures.

We consider the scientist's citation record, rather than her citation curve. Formally, the citation record is a random variable $X(\omega)$ defined on the events $\omega \in \Omega$, where each event ω corresponds to the journal in which the publication appeared and Ω is the set of journals.

Thus, $X(\omega)$ is the number of citations of the papers published in journal ω.[20] The primal space is given by the set of all possible scientist's citation records.

We now fix a family of probabilities $\mathcal{P}$ defined on Ω, where for each $Q \in \mathcal{P}$, $Q(\omega)$ represents the "value" of journal $\omega \in \Omega$. The dual space is given by all possible linear valuation of the journals, i.e., the "Arrow–Debreu prices" of each journal. As mentioned above, the journals' evaluations, namely the selection of the family $\mathcal{P}$, are determined a priori and could be based on various criteria. For example, a specific measure Q can assign high value to journals with high impact factor, while another measure Q can assign high value to journals with low acceptance rate of submissions. A priori there will be no consensus on how to evaluate a journal and this is reflected in the set of $Q \in \mathcal{P}$ rather than just a single probability measure.

Then we select a family of functions $\{\gamma(Q,q)\}_{q\in\mathbb{R}}$. Each function $\gamma(\cdot,q) : \mathcal{P} \to \mathbb{R}$ associates to each measure Q the value $\gamma(Q,q)$, representing the smallest Q-average of citations which is necessary to reach a research performance level q. This research performance assessment is under the measure Q and the function $\gamma(Q,q)$ does not depend on the scientist's citation record X. The Dual SRM is given by the Φ-index

$$\Phi(X) := \inf_{Q\in\mathcal{P}} \beta(Q,X) \tag{7}$$

where $\beta(Q,X) := \sup\{q \in \mathbb{R} \mid E_Q[X] \geq \gamma(Q,q)\}$ represents the research performance of a scientist with citation record X, under the fixed measure Q. In general, the function $\gamma(Q,q)$ can be nonlinear in q, perhaps reflecting different scientist's effort to reach a certain performance level. The main feature of $\gamma(Q,q)$ is to be non decreasing in q. In the previous

[20] Over the years any scientist may have published more than one paper in the same journal. A simple approach to account for this situation is to consider her total number of citations of all publications in the given journal. Indeed, if the journal evaluation criterion, say Q_i, does not change over time, the total number of citations per journal is the only relevant quantity when assessing the scientist's research performance as $E_{Q_i}[X]$ is a linear operator. To simplify the exposition we implicitly adopted this approach in the current section. If instead the journal evaluation criterion changes over time (e.g., the impact factor of the journal), the citations of publications in a given year can be evaluated using the journal evaluation criteria for that year. Hence, citations of publications in different years can be evaluated using different journal evaluation criteria.

section, the $\gamma(Q, q)$ is either constant or specified as in (4). The Φ-index is a conservative or prudential assessment of a scientist's research performance, as it is obtained under the least favorable journal's evaluation criterion.

For each measure Q, the function $\beta(Q, X)$ in (7) resembles equation (1) defining the ϕ-index. Indeed, to derive the ϕ-index the citation curve $X(x)$ is compared with performance curves $f_q(x)$. In the dual setting, Q-averages of citation records $X(\omega)$, i.e., $E_Q[X]$, are compared with the smallest Q-average of citations to reach a performance level q, i.e., the function $\gamma(Q, q)$. The key difference between the two approaches is that the set of functions $\beta(Q, X)$ are derived under "different scenarios" $Q \in \mathcal{P}$ characterized by different criteria to assess journals' quality. This feature characterizes the Dual SRM and it is not present in the SRM or any other bibliometric index based on citation curves.

By construction any Dual SRM in (7) is monotone increasing and quasi-concave, as proved in Proposition 2 in Appendix B. Both properties have a natural motivation. The monotonicity property implies that better scientists have higher research performances. The quasi-concavity induces sensible rankings of research centers. The citation records of two scientists can be meaningfully aggregated by taking their linear convex combination $\lambda X_1 + (1 - \lambda) X_2$, where $\lambda \in [0, 1]$.[21] The aggregation is meaningful because it is at the level of citation records, i.e., scientists' citations of publications in the same journal. Notably, quasi-concavity implies that the performance of the research team is at least equal to the scientific performance of the less performing member, which appears to be a natural property.

An important question is the following: Does any monotone increasing and quasi-concave map admit the representation in (7)? Essentially, the answer is positive and formalized in Theorem 1. In order to prove the theorem, we need to assume the continuity from above of Φ and to introduce some new notations.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $\mathcal{F}$ is a σ-algebra and P is a probability measure on $\mathcal{F}$. Since the citation record of an author X is a bounded function, it appears natural to take $X \in L^\infty(\Omega, \mathcal{F}, P)$, where $L^\infty(\Omega, \mathcal{F}, P)$ is the vector space of $\mathcal{F}$-measurable

[21] If $\lambda = 0.5$, the convex combination is simply the arithmetic mean of the two citation records. To aggregate n scientists' citation records one computes $\sum_{i=1}^n \lambda_i X_i$, where $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geq 0$ for all i.

functions that are P almost surely bounded. We also denote with $\Delta := \{Q \ll P\}$ the set of all probabilities absolutely continuous with respect to P.

Theorem 1 *Any map $\Phi : L^\infty(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$ that is quasi-concave, monotone increasing and continuous from above can be represented as*

$$\Phi(X) := \inf_{Q \in \Delta} \beta(Q, X), \quad \forall X \in L^\infty(\Omega, \mathcal{F}, P) \tag{8}$$

where $\beta : \Delta \times L^\infty(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$ and $\gamma : \Delta \times \mathbb{R} \to \overline{\mathbb{R}}$ are defined as

$$\begin{aligned} \beta(Q, X) :=& \quad \sup \{q \in \mathbb{R} \mid E_Q[X] \geq \gamma(Q, q)\} \\ \gamma(Q, q) :=& \quad \inf_{Y \in L^\infty} \{E_Q[Y] \mid \Phi(Y) \geq q\}. \end{aligned} \tag{9}$$

The proof is in Appendix B.[22]

There are two technical differences between the ϕ-index in (1) and the Φ-index in (7). The first difference is that the ϕ-index is law invariant while the Φ-index may not. If two scientists have the same citation curve, then they have the same ϕ-index. If they published in different journals, in general they have different Φ-indices. Thus, the Φ-index can rank scientists with the same citation curves.[23] The second difference is that the ϕ-index is not necessarily quasi-concave on the citation records (even if it is always quasi-concave on the citation curves). When the ϕ-index is quasi-concave on the citation records, the Φ-index includes the ϕ-index as a special case.

[22]In equation (8) the infimum is taken with respect to all probabilities in Δ, and not in a prescribed subset $\mathcal{P} \subseteq \Delta$. This must be the case, since in the theorem the map Φ is given a priory and only the objects that can be determined from Φ can appear in the representation (8).

[23]Frittelli, Maggis, and Peri (2013) provide a theoretical analysis of quasi-convex maps defined on distributions.

5. Empirical Analysis

This section provides an empirical application of our SRM, discusses its implementation, and compares the empirical findings to 7 existing bibliometric indices.[24]

5.1. Data

We consider 173 full professors affiliated to 11 main U.S. business schools or finance departments. Thus, we consider a rather large and homogenous set of researchers, in terms of seniority and field of research, making comparison meaningful. A Python script is run to extract their citation curves from Google Scholar (http://scholar.google.com/), which are publicly and freely available. The American Scientist Open Access Forum (2008) remarked that "Google Scholar's accuracy is growing daily, with growing content."[25] Our dataset is created on September 6, 2012. Table 3 shows the list of business schools or finance departments that for brevity we call universities. The number of faculty members varies significantly across universities, ranging from 8 members at Berkeley and Cornell to 34 members at Columbia. To provide an overview of the data, Table 3 reports summary statistics for the number of citations and publications, aggregated at the university level. Chicago has the largest average number of citations, as well as median and standard deviation, and fourth largest faculty by size. For all universities, the average number of citations is more than twice the median number of citations, suggesting that the distribution of citations is highly positive skewed. Princeton has the largest number of publications, and fifth largest faculty by size. The average number of publications is often more than twice the median number of publications.

[24] As discussed in the previous section, an empirical application of Dual SRMs requires a richer dataset, including in which journal each author's publications appeared and values of those journals. We defer such an application to future work.

[25] http://openaccess.eprints.org/index.php?/archives/417-Citation-Statistics-International-Mathematical-Union-Report.html.

5.2. *Estimates of the Scientific Research Measure*

To properly rank research performances, performance curves should mimic actual citation curves. We experimented different functional forms of performance curves. We emphasize that the flexibility in the choice of performance curves stems from the fact that our class of SRMs in (1) accommodates performance curves with largely different shapes. It turns out that a sensible specification of performance curves is given by the hyperbolic function $f_q(x) = q/x^c - k$, trading off fitting accuracy and parametric complexity of the curve. Indeed, all citation curves in our dataset display hyperbolic-type shapes. Figure 3 shows four typical citation curves. Superimposed to each citation curve is the fitted hyperbolic function. The fitting appears to be accurate; indeed all adjusted R^2's are above 97.4%. We fit the hyperbolic function $f_q(x) = q/x^c - k$ using nonlinear least squares, i.e., by minimizing $\sum_{x=1}^{p}(X(x) - f_q(x))^2$ with respect to q, c and k, where p is the author's number of publications with at least one citation each. Starting values for the nonlinear minimization are obtained as follows. We note that $\tilde{f}_q(x) = q/x^c$ is a restricted version of $f_q(x)$, imposing $k = 0$. As in the log-log space $\tilde{f}_q(x)$ is linear in $\log(x)$, i.e., $\log(\tilde{f}_q(x)) = \log(q) - c\log(x)$, the two parameter estimates $\widehat{\log}(q)$ and $\hat{c}$ can be easily obtained by regressing the log number of citations on a constant and the log number of publications, i.e., by running the least square regression of $\log(X(x))$ on 1 and $\log(x)$, where $x = 1, 2, \ldots, p$. As starting values for q, c and k we take $\exp(\widehat{\log}(q))$, $\hat{c}$ and 0, respectively.

For each author in our dataset, we fit the hyperbolic function to its citation curve, as described above. Table 4 summarizes the estimation results. For all regressions, t-statistics are significantly different from zero. All adjusted R^2 are above 95%, suggesting that hyperbolic functions provide an accurate description of citation curves. We point out that estimates of the q parameter in Table 4 are not research performance levels. The reason is that the corresponding hyperbolic functions pass through the citation curves, and not below the citation curves as required by (1).

Given the evidence above, we use the following ϕ-index to rank the authors in our dataset:

$$\phi(X) := \sup \{q \in \mathbb{R} \mid X(x) \geq q/x^c - k, \forall x \in [1, \infty)\} \tag{10}$$

where we set $c = 0.59$ and $k = 543.2$, which are the weighted averages of all estimated c and k coefficients in our sample, respectively, with weights given by the number of data points used to estimate each coefficient. We also experimented other values of c and k, such as the simple average of all estimated c and k coefficients. The empirical results are quite similar and available from the authors upon request. The ϕ-index in (10) provides the common ground to rank all authors in our dataset.

The following two-step algorithm can be used to compute the ϕ-index in (10). Given an author with p publications with at least one citation each:

1. Compute $q_x := (X(x) + k)\, x^c$, for $x = 1, 2, \ldots, p$, and $q_* := k\, p^c$.

2. Take $\phi(X) := \min(q_1, q_2, \ldots, q_p, q_*)$.

By construction, the performance curve associated to q_x passes through the point $(x, X(x))$. This holds true for each $x = 1, 2, \ldots, p$.[26] The performance curve associated to q_* crosses the horizontal axis at $(p, 0)$, i.e., $f_{q_*}(p) = 0$. To better understand the two-step algorithm, consider again Figure 1. The figure shows the $p = 8$ performance curves for the hypothetical author with 8 publications associated to $q_1, \ldots, q_p$. The performance curve associated to q_*, not shown in Figure 1, would pass through the point $(8, 0)$. Taking the minimum among all the q_x's and q_* (which is $q_3 = 24.75$ in Figure 1) ensures that the associated performance curve is below all the other performance curves (as performance curves are increasing in q) and consequently it is also below the citation curve $X(x)$. At the same time, it is also the highest performance curve still below the citation curve, as it is touching the citation curve in one point (which is $(3, 5)$ in Figure 1). The performance curve associated to q_* enters the

[26]For example, the performance curve associated to q_p is $f_{q_p}(x) = q_p/x^c - k$. When evaluated at $x = p$, it gives $f_{q_p}(p) = q_p/p^c - k = (X(p) + k)\, p^c/p^c - k = X(p)$, showing that this performance curve is passing through the point $(p, X(p))$.

calculation of the ϕ-index because the whole citation curve may lay above the performance curve f_{q_*}, and in this case f_{q_*} determines the ϕ-index in (10). This may happen when the citation curve is roughly linear and steep. Although in practice citation curves do not appear to have this shape, the performance curve f_{q_*} needs to be considered when computing the ϕ-index in (10).[27]

5.3. *Empirical Results*

For each of the 173 authors in our dataset, we compute the ϕ-index in (10) and the other 7 bibliometric indices listed in Table 1. To provide an overview of the results, we aggregate estimates of the ϕ-index and bibliometric indices at the university level. Table 5 shows mean and median for each index and university. Averages are often significantly larger than medians (e.g., for Princeton and Stanford), implying that the distribution of the indices is positively skewed. Thus, aggregated results should be interpreted cautiously.

The main message from Table 5 is that rankings of universities based on our ϕ-index and the other bibliometric indices are generally different. Under the premise that the research performance is reflected by the whole citation curve, rather than by only part of it, our ϕ-index provides by construction a better assessment of research performances than the other bibliometric indices, leading to a more sensible ranking. Differences in rankings are relatively more concentrated on "average" performing universities, rather than in most and least performing universities, in relative terms. For example, according to the average ϕ-index, Chicago ranks first and Princeton ranks second. According to the average h-index, Princeton ranks first and Chicago ranks second. Berkeley ranks tenth according to both averages. For the remaining universities, the rankings provided by the ϕ- and h-index, as well as the other indices, are quite different.

Empirical results at the level of individual full professors parallel the empirical results at

[27]More specifically when $x = p$, $X(p) \geq 1 > 0 = f_{q_*}(p)$, which suggests that the citation curve $X(x)$ may lay above $f_{q_*}(x)$ for all x when the citation curve is roughly linear and steep. As the citation curve $X(x) = 0$ for all $x > p$ and the performance curves are increasing in q, it is not necessary to consider citation curves for $q > q_*$, as all these performance curves lay above f_{q_*} and thus lay above the citation curve $X(x)$ at least for all $x > p$.

the university level. Because of space constraints, we cannot report estimates of all indices for all 173 full professors. Nonetheless, to give a sense of the empirical findings, Figure 4 shows the scatter plots of the ϕ-index versus the h-index and versus the g-index for each scholar in our sample. Scatter plots of the ϕ-index versus the other bibliometric indices (not reported) share similar patterns. The relation between the indices is clearly positive. Best performing scholars tend to have the largest ϕ- and h-index (or g-index). Scholars in the top-right corner of the scatter plots are typically Nobel laureates in economics. This obviously induces a positive correlation between the indices. However, the relation is far from monotone, meaning that different indices lead to different rankings. A closer look at the scatter plots shows that in various cases, different scholars have the same (or almost the same) h-index (plotted on the vertical axis of upper graphs), while their ϕ-indices are different. This implies that the h-index cannot be used to rank them, while the ϕ-index provides a more granular ranking of scholars. This situation occurs for example when two authors have roughly the same number of publications (and h-index) but one author has more citations per publication, especially in the Hirsch core, which implies a steeper citation curve to the left and thus higher ϕ-index. As a consequence of this phenomenon, at the university level the median h-index (rounded to the nearest integer) is 29 both for Cornell and Yale, and 25 both for Berkeley and Penn. In contrast, the corresponding median ϕ-index is quite different for these universities, as shown in Table 5.

To summarize, this empirical application shows that our ϕ-index can be easily implemented and in general leads to different rankings than other bibliometric indices.

6. Conclusion

Funding research projects, faculty recruitment, and other key decisions for universities and research institutions depend to a large extent upon the scientific merits of the involved researchers. This paper introduces the novel class of Scientific Research Measures to rank scientists' research performances. The main feature of SRMs is to take into account the whole scientists' citation curves and to reflect specific features of the scientific field, research areas, and seniority of the scientists. This task is achieved by using flexible performance curves. Most existing bibliometric indices do not appear to have such a flexibility.

We also introduce the further general class of Dual SRMs that allows for a more informed ranking of research performances using citation records. Dual SRMs allow to aggregate scientists' research outputs and induce sensible rankings of research teams, departments and universities, which has been a central theme in the scientific community over the last decades.

Finally, we provide an empirical application of the SRM using on a novel dataset of 173 finance scholars' citation curves. We develop a simple two-step algorithm to compute the SRM. The empirical results show that hyperbolic performance curves describe well actual citation curves. Authors' ranking based on the SRM characterized by hyperbolic performance curves and authors' rankings based on other 7 existing bibliometric indices are generally quite different. Under the premise that the research performance is reflected by the whole citation curve, rather than by only part of it, the SRM provides by construction a more accurate ranking of research performances.

A. Quasi-concavity of Scientific Research Measures

In this section we provide a formal discussion of the quasi-concavity property of SRMs. The following proposition states that the SRM defined in (1) is quasi-concave by construction.

Proposition 1 *Let $\{f_q\}_q$ be a family of performance curves and ϕ be the associated SRM defined in (1). Then ϕ is quasi-concave: for all citation curves X_1 and X_2*

$$\phi(\lambda X_1 + (1-\lambda)X_2) \geq \min(\phi(X_1), \phi(X_2)), \quad \lambda \in [0,1].$$

Proof. Let $m := \min(\phi(X_1), \phi(X_2))$, so $\phi(X_1) \geq m$ and $\phi(X_2) \geq m$. By definition of ϕ, $\forall \varepsilon > 0$ $\exists q_i$ s.t. $X_i \geq f_{q_i}$ and $q_i > \phi(X_i) - \varepsilon \geq m - \varepsilon$. Then $X_i \geq f_{q_i} \geq f_{m-\varepsilon}$, as $\{f_q\}_q$ is an increasing family, and therefore $\lambda X_1 + (1-\lambda)X_2 \geq f_{m-\varepsilon}$. As this holds for any $\varepsilon > 0$, we conclude that $\phi(\lambda X_1 + (1-\lambda)X_2) \geq m$ and ϕ is quasi-concave. ■

As mentioned in Section 3.2, the reason to adopt quasi-concavity is an internal consistency or coherency property that we now explain with an example.

Consider the citation curves of two scientists A and B, having 3 publications each and the same total number of citations:

$$A = \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix}, \qquad B = \begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}.$$

The citations of A are more "concentrated" while the citations of B are less concentrated or more "diversified," with the obvious intuition of the meaning of these terms.

We evaluate A and B using a ϕ-index associated to the family of performance curves $\{f_q\}_q$. Depending on the performance curves $\{f_q\}_q$ that reflect the specific features of their scientific field, research areas, seniorities, etc., one could obtain either

$$\phi\begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix} \geq \phi\begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix} \tag{11}$$

which means that the SRM assigns more importance to concentration, or

$$\phi\begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix} \geq \phi\begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix} \tag{12}$$

which on the contrary means that the SRM attributes more relevance to diversification. These two different rankings do not depend on the property of quasi-concavity but on the features of the performance curves $\{f_q\}_q$, that in turn reflect the characteristics of the citations of the particular field. Both rankings (11) or (12) are possible.[28]

Now consider the following three citation curves:

$$C_1 = \begin{pmatrix} 22 \\ 20 \\ 18 \end{pmatrix}; \quad C_2 = \begin{pmatrix} 25 \\ 20 \\ 15 \end{pmatrix}; \quad C_3 = \begin{pmatrix} 28 \\ 20 \\ 12 \end{pmatrix}.$$

These three citations curves have the same total number of citations as A and B, but the citations of C_1, C_2 and C_3 are more concentrated than B. Notice that none of the five citation curves A, B, C_1, C_2 and C_3 dominates any other. Quasi-concavity will induce a conditional ranking among them that we now illustrate. Note that the h-index is 3 for all scientists, so it would not induce any ranking.

Case I: Suppose that the SRM attributes more importance to *concentration*, meaning that (11) is satisfied. Then we should expect that:

$$\phi\begin{pmatrix} 22 \\ 20 \\ 18 \end{pmatrix} \geq \phi\begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}, \quad \phi\begin{pmatrix} 25 \\ 20 \\ 15 \end{pmatrix} \geq \phi\begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}, \quad \phi\begin{pmatrix} 28 \\ 20 \\ 12 \end{pmatrix} \geq \phi\begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}. \tag{13}$$

This natural and minimal coherency property (the coherency is between (11) and (13)) is satisfied because of the quasi-concavity of ϕ. In other words, the quasi-concavity of ϕ implies

[28]For example, if $f_q(x) = -10\,x + 10\,q$, $q \in \mathbb{N}$, then $\phi(A) = 4 > 3 = \phi(B)$. As another example, if $f_q(x) = 5\,q\,1_{(0,q]}(x)$, $q \in \mathbb{N}$, then $\phi(B) = 4 > 2 = \phi(A)$.

that (13) holds true whenever (11) is satisfied.[29] Indeed,

$$\lambda \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix} + (1-\lambda) \begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}$$

is equal to C_1, C_2, C_3, respectively, for $\lambda = 0.2$; $\lambda = 0.5$; $\lambda = 0.8$; and so

$$\phi\left(\lambda \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix} + (1-\lambda) \begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}\right) \geq \min \left\{ \phi \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix}, \phi \begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix} \right\} = \phi \begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}.$$

In the above expression the inequality is due to quasi-concavity and the equality to (11).

Case II: Suppose that the SRM attributes more importance to *diversification*, meaning that (12) is satisfied. Note that the citations of C_1, C_2 and C_3 are less concentrated than A. Using the same argument as above, we see that quasi-concavity implies:

$$\phi \begin{pmatrix} 22 \\ 20 \\ 18 \end{pmatrix} \geq \phi \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix}, \quad \phi \begin{pmatrix} 25 \\ 20 \\ 15 \end{pmatrix} \geq \phi \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix}, \quad \phi \begin{pmatrix} 28 \\ 20 \\ 12 \end{pmatrix} \geq \phi \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix}. \tag{14}$$

as we should expect from a SRM that emphasizes diversification.

[29]Quasi-convexity of ϕ would be that $\phi(\lambda X_1 + (1-\lambda)X_2) \leq \max(\phi(X_1), \phi(X_2))$. This property of ϕ would induce quite different internal consistency conditions that are much less relevant than those induced by quasi-concavity. This is mainly due to the monotonicity property of the ϕ-index, which reflects the natural request that a better scientist should have a higher ϕ-index. The monotonicity property provides a natural direction to compare citation curves. Specifically, we are interested in conditions guaranteeing that a citation curve is "at least as good as" another citation curve, which are precisely the type of conditions in (13) or (14). In contrast, the quasi-convexity of ϕ would only imply that a citation curve is "at most as good as" another citation curve, that is:

$$\begin{array}{lll} (11) & \Rightarrow & \phi(C_i) \leq \phi(A), \quad i = 1,2,3 \\ (12) & \Rightarrow & \phi(C_i) \leq \phi(B), \quad i = 1,2,3. \end{array}$$

The conclusion about the selection of the quasi-concavity property is the following: given *some citations curves* and a ranking among them (ranking which depends on the specific features of the performance curves of that particular field) quasi-concavity is a coherent requirement – in terms of concentration or diversification – for the *many other citation curves* that can be obtained by convex combination of the given ones.

B. Proof of Dual Representation of SRMs

As a direct consequence of its definition in (7), any Dual SRM is monotone increasing and quasi-concave. We use the notation introduced before the statement of Theorem 1.

Proposition 2 *Given a subset of probabilities $\mathcal{P} \subseteq \Delta$ and a map $\gamma(Q,q) : \mathcal{P} \times \mathbb{R} \to \mathbb{R}$ let*

$$\widetilde{\beta}(Q,t) := \quad \sup\{q \in \mathbb{R} \mid t \geq \gamma(Q,q)\} \tag{15}$$

$$\Phi(X) := \quad \inf_{Q \in \mathcal{P}} \widetilde{\beta}(Q, E_Q[X]) = \inf_{Q \in \mathcal{P}} \beta(Q,X). \tag{16}$$

Then the map $\Phi : L^{\infty}(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$ is quasi-concave and monotone increasing.

Proof. Notice that Φ may be equivalently expressed using the map β or the map $\widetilde{\beta}$, as it is explicitly written in (16). One immediately also checks that the function $\widetilde{\beta}(Q,\cdot) : \mathbb{R} \to \overline{\mathbb{R}}$ is monotone increasing. As any monotone function from $\mathbb{R}$ to $\overline{\mathbb{R}}$ is also quasi-concave, we obtain that, for each $Q \in \mathcal{P}$ and X_1, $X_2 \in L^{\infty}(\Omega, \mathcal{F}, P)$,

$$\begin{aligned} \widetilde{\beta}(Q, E_Q[\lambda X_1 + (1-\lambda) X_2]) &= \widetilde{\beta}(Q, \lambda E_Q[X_1] + (1-\lambda) E_Q[X_2]) \\ &\geq \min\left(\widetilde{\beta}(Q, E_Q[X_1]),\, \widetilde{\beta}(Q, E_Q[X_2])\right). \end{aligned}$$

Taking the infimum with respect to $Q \in \mathcal{P}$ on both sides and exchanging the infimum and the minimum, we deduce that also $\Phi : L^{\infty}(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$ is quasi-concave. The monotonicity of $\widetilde{\beta}(Q,\cdot)$ and of the map $E_Q[\cdot] : L^{\infty}(\Omega, \mathcal{F}, P) \to \mathbb{R}$ guarantee the monotonicity of $\Phi : L^{\infty}(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$. ■

The goal of this section is the answer, anticipated in Theorem 1, to the reverse implication of the above Proposition: does *any* quasi-concave monotone increasing map $\Phi : L^\infty(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$ admit a representation in the form (16)? Before proving Theorem 1, we need some topological structure. For simplicity we denote $L^\infty := L^\infty(\Omega, \mathcal{F}, P)$ and with $L^1 := L^1(\Omega, \mathcal{F}, P)$ the space of integrable random variables.

If we endow L^∞ with the weak topology $\sigma(L^\infty, L^1)$ then $L^1 = (L^\infty, \sigma(L^\infty, L^1))'$ is its topological dual. In the dual pairing $(L^\infty, L^1, \langle \cdot, \cdot \rangle)$ the bilinear form $\langle \cdot, \cdot \rangle : L^\infty \times L^1 \to \mathbb{R}$ is given by $\langle X, Z \rangle = E[ZX]$. The linear function $X \mapsto E[ZX]$, with $Z \in L^1$, is $\sigma(L^\infty, L^1)$ continuous and $(L^\infty, \sigma(L^\infty, L^1))$ is a locally convex topological vector space.

Definition 1 *A map $\Phi : L^\infty \to \overline{\mathbb{R}}$ is $\sigma(L^\infty, L^1)$-upper semicontinuous if the upper level sets $\{X \in L^\infty \mid \Phi(X) \geq q\}$ are $\sigma(L^\infty, L^1)$-closed for all $q \in \mathbb{R}$.*

For the class of maps that we are considering, $\sigma(L^\infty, L^1)$-upper semicontinuity is equivalent to the continuity from above. This fact can be proved in a similar way to the convex case; see, e.g., Föllmer and Schied (2004).

Lemma 1 *Let $\Phi : L^\infty \to \overline{\mathbb{R}}$ be quasi-concave and monotone increasing. Then the following are equivalent:*

Φ is $\sigma(L^\infty, L^1)$-upper semicontinuous;

Φ is continuous from above: $X_n, X \in L^\infty$ and $X_n \downarrow X$ imply $\Phi(X_n) \downarrow \Phi(X)$.

Proof. Let Φ be $\sigma(L^\infty, L^1)$-upper semicontinuous and suppose that $X_n \downarrow X$. The monotonicity of Φ implies $\Phi(X_n) \geq \Phi(X)$ and $\Phi(X_n)$ is decreasing and therefore $q := \lim_n \Phi(X_n) \geq \Phi(X)$. Hence $\Phi(X_n) \geq q$ and $X_n \in B_q := \{Y \in L^\infty \mid \Phi(Y) \geq q\}$ which is $\sigma(L^\infty, L^1)$-closed by assumption. As the elements in L^1 are order continuous, from $X_n \downarrow X$ we get $X_n \overset{\sigma(L^\infty, L^1)}{\longrightarrow} X$ and therefore $X \in B_q$. This implies that $\Phi(X) = q$ and that Φ is continuous from above.

Conversely, suppose that Φ is continuous from above. We have to show that the convex set B_q is $\sigma(L^\infty, L^1)$-closed for any q. By the Krein Smulian Theorem it is sufficient to prove that $C := B_q \cap \{X \in L^\infty \mid \quad \| X \|_\infty \leq r\}$ is $\sigma(L^\infty, L^1)$-closed for any fixed $r > 0$. As $C \subseteq L^\infty \subseteq L^1$ and as the embedding

$$(L^\infty, \sigma(L^\infty, L^1)) \hookrightarrow (L^1, \sigma(L^1, L^\infty))$$

is continuous it is sufficient to show that C is $\sigma(L^1, L^\infty)$-closed. Since the $\sigma(L^1, L^\infty)$ topology and the L^1 norm topology are compatible, and C is convex, it is sufficient to prove that C is closed in L^1. Take $X_n \in C$ such that $X_n \to X$ in L^1. Then there exists a subsequence $\{Y_n\}_n \subseteq \{X_n\}_n$ such that $Y_n \to X$ a.s. and $\Phi(Y_n) \geq q$ for all n. Set $Z_m := \sup_{n \geq m} Y_n \vee X$. Then $Z_m \in L^\infty$, since $\{Y_n\}_n$ is uniformly bounded, and $Z_m \geq Y_m$, $\Phi(Z_m) \geq \Phi(Y_m)$ and $Z_m \downarrow X$. From the continuity from above we conclude: $\Phi(X) = \lim_m \Phi(Z_m) \geq \limsup_m \Phi(Y_m) \geq q$. Thus $X \in B_q$ and consequently $X \in C$. ■

Now we provide a dual representation of SRMs in the same spirit as Volle (1998), Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2011), and Drapeau and Kupper (2013). We first provide the representation of Φ in Theorem 1 in terms of the dual function H defined below in (17) and then we show that Φ can also be represented in terms of the right continuous version H^+ (defined in (20)) of H, which can be written in a different way as in (15) or (21).

Proof of Theorem 1.

Step 1 We show that

$$\Phi(X) = \inf_{Z \in L^1_+} H(Z, E[ZX])$$

where $H : L^1 \times \mathbb{R} \to \overline{\mathbb{R}}$ is defined by

$$H(Z, t) := \sup_{\xi \in L^\infty} \{\Phi(\xi) \mid E[Z\xi] \leq t\}. \tag{17}$$

Fix $X \in L^\infty$. As $X \in \{\xi \in L^\infty \mid E[Z\xi] \leq E[ZX]\}$, by the definition of $H(Z, E[ZX])$

we deduce that, for all $Z \in L^1$, $H(Z, E[ZX]) \geq \Phi(X)$, hence

$$\inf_{Z \in L^1} H(Z, E[ZX]) \geq \Phi(X). \tag{18}$$

We prove the opposite inequality. Let $\varepsilon > 0$ and define the set

$$C_\varepsilon := \{\xi \in L^\infty \mid \Phi(\xi) \geq \Phi(X) + \varepsilon\}.$$

As Φ is quasi-concave and $\sigma(L^\infty, L^1)$-upper semicontinuous (by Lemma (1)), C is convex and $\sigma(L^\infty, L^1)$-closed. Since $X \notin C_\varepsilon$, (if $\Phi(X) = -\infty$, we may take $C_M := \{\xi \in L^\infty \mid \Phi(\xi) \geq -M\}$ and the following argument would hold as well) the Hahn–Banach theorem implies the existence of a continuous linear functional that strongly separates X and C_ε, that is there exist $Z_\varepsilon \in L^1$ such that

$$E[Z_\varepsilon \xi] > E[Z_\varepsilon X] \quad \forall \xi \in C_\varepsilon. \tag{19}$$

Hence $\{\xi \in L^\infty \mid E[Z_\varepsilon \xi] \leq E[Z_\varepsilon X]\} \subseteq C_\varepsilon^c := \{\xi \in L^\infty \mid \Phi(\xi) < \Phi(X) + \varepsilon\}$ and from (18)

$$\begin{aligned}
\Phi(X) &\leq \inf_{Z \in L^1} H(Z, E[ZX]) \leq H(Z_\varepsilon, E[Z_\varepsilon X]) \\
&= \sup\{\Phi(\xi) \mid \xi \in L^\infty \text{ and } E[Z_\varepsilon \xi] \leq E[Z_\varepsilon X]\} \\
&\leq \sup\{\Phi(\xi) \mid \xi \in L^\infty \text{ and } \Phi(\xi) < \Phi(X) + \varepsilon\} \leq \Phi(X) + \varepsilon.
\end{aligned}$$

Therefore, $\Phi(X) = \inf_{Z \in L^1} H(Z, E[ZX])$. To show that the inf can be taken over the positive cone L^1_+, it is sufficient to prove that $Z_\varepsilon \subseteq L^1_+$. Let $Y \in L^\infty_+$ and $\xi \in C_\varepsilon$. Given that Φ is monotone increasing, $\xi + nY \in C_\varepsilon$ for every $n \in \mathbb{N}$ and, from (19), we have:

$$E[Z_\varepsilon(\xi + nY)] > E[Z_\varepsilon X] \Rightarrow E[Z_\varepsilon Y] > \frac{E[Z_\varepsilon(X - \xi)]}{n} \to 0, \text{ as } n \to \infty.$$

As this holds for any $Y \in L^\infty_+$ we deduce that $Z_\varepsilon \subseteq L^1_+$. Therefore, $\Phi(X) =$

$\inf_{Z\in L^1_+} H(Z,E[ZX])$.

Step 2 We show that

$$\inf_{Z\in L^1_+} H(Z,E[ZX]) = \inf_{Z\in L^1_+} H^+(Z,E[ZX])$$

where $H^+(Z,\cdot)$ is the right continuous version of H:

$$H^+(Z,t) := \inf_{s>t} H(Z,s). \tag{20}$$

Since $H(Z,\cdot)$ is increasing and $Z\in L^1_+$ we obtain

$$H^+(Z,E[ZX]) := \inf_{s>E[ZX]} H(Z,s) \le \lim_{X_m\downarrow X} H(Z,E[ZX_m])$$

$$\begin{aligned}
\Phi(X) &= \inf_{Z\in L^1_+} H(Z,E[ZX]) \le \inf_{Z\in L^1_+} H^+(Z,E[ZX]) \le \inf_{Z\in L^1_+} \lim_{X_m\downarrow X} H(Z,E[ZX_m]) \\
&= \lim_{X_m\downarrow X} \inf_{Z\in L^1_+} H(Z,E[ZX_m]) = \lim_{X_m\downarrow X} \Phi(X_m) \overset{(CFA)}{=} \Phi(X)
\end{aligned}$$

where in the last equality we used the continuity from above (CFA) of Φ.

Step 3 We prove that

$$H^+(Z,t) = \widetilde{\beta}(Z,t) \tag{21}$$

where, for $Z\in L^1$ and $t,q\in\mathbb{R}$,

$$\begin{aligned}
\gamma(Z,q) &= \inf_{Y\in L^\infty} \{E[ZY] \mid \Phi(Y)\ge q\} \\
\widetilde{\beta}(Z,t) &= \sup\{q\in\mathbb{R} \mid \gamma(Z,q)\le t\}.
\end{aligned}$$

Clearly γ and $\widetilde{\beta}$ coincide with (9) and (15) when $Q\in\Delta$ and $Z=dQ/dP$.

Note that $\widetilde{\beta}(Z,\cdot)$ is the right inverse of the increasing function $\gamma(Z,\cdot)$ and therefore $\widetilde{\beta}(Z,\cdot)$ is right continuous. To prove that $H^+(Z,t)\le\widetilde{\beta}(Z,t)$ it is sufficient to show that for all $p > t$ we have:

$$H(Z,p) \leq \widetilde{\beta}(Z,p). \tag{22}$$

Indeed, if (22) is true

$$H^+(Z,t) = \inf_{p>t} H(Z,p) \leq \inf_{p>t} \widetilde{\beta}(Z,p) = \widetilde{\beta}(Z,t)$$

as both H^+ and $\widetilde{\beta}$ are right continuous in the second argument.

Writing explicitly the inequality (22)

$$\sup_{\xi \in L^\infty} \{\Phi(\xi) \mid E[Z\xi] \leq p\} \leq \sup \{q \in \mathbb{R} \mid \gamma(Z,q) \leq p\}$$

and letting $\xi \in L^\infty$ satisfying $E[Z\xi] \leq p$, we see that it is sufficient to show the existence of $q \in \mathbb{R}$ such that $\gamma(Z,q) \leq p$ and $q \geq \Phi(\xi)$. If $\Phi(\xi) = \infty$ then $\gamma(Z,q) \leq p$ for any q and therefore $\widetilde{\beta}(Z,p) = H(Z,p) = \infty$.

Suppose now that $\infty > \Phi(\xi) > -\infty$ and define $q := \Phi(\xi)$. As $E[\xi Z] \leq p$ we have:

$$\gamma(Z,q) := \inf \{E[Z\xi] \mid \Phi(\xi) \geq q\} \leq p.$$

Then $q \in \mathbb{R}$ satisfies the required conditions.

To obtain $H^+(Z,t) := \inf_{p>t} H(Z,p) \geq \widetilde{\beta}(Z,t)$ it is sufficient to prove that, for all $p > t$, $H(Z,p) \geq \widetilde{\beta}(Z,t)$, that is:

$$\sup_{\xi \in L^\infty} \{\Phi(\xi) \mid E[Z\xi] \leq p\} \geq \sup \{q \in \mathbb{R} \mid \gamma(Z,q) \leq t\}. \tag{23}$$

Fix any $p > t$ and consider any $q \in \mathbb{R}$ such that $\gamma(Z,q) \leq t$. By the definition of γ, for all $\varepsilon > 0$ there exists $\xi_\varepsilon \in L^\infty$ such that $\Phi(\xi_\varepsilon) \geq q$ and $E[Z\xi_\varepsilon] \leq t + \varepsilon$. Take ε such that $0 < \varepsilon < p - t$. Then $E[Z\xi_\varepsilon] \leq p$ and $\Phi(\xi_\varepsilon) \geq q$ and (23) follows.

Step 4 Normalization

From the above Steps 1, 2 and 3 we then deduce:

$$\begin{aligned}\Phi(X) &= \inf_{Z\in L^1_+} H(Z, E[ZX]) = \inf_{Z\in L^1_+} H^+(Z, E[ZX]) = \inf_{Z\in L^1_+} \widetilde{\beta}(Z, E[ZX]) \\ &= \inf_{Z\in L^1_+} \beta(Z, X).\end{aligned}$$

To conclude the thesis we only need to normalize the elements $Z \in L^1_+$. This is possible since, by definition of $H(Z,t)$,

$$H(Z, E[ZX]) = H(\lambda Z, E[(\lambda Z)X]), \quad \forall Z \in L^1_+, Z \neq 0, \lambda \in (0,\infty)$$

and so, by setting $dQ/dP = Z/E[Z]$, $Q \in \Delta$, we obtain

$$\Phi(X) = \inf_{Z\in L^1_+(\mathbb{R})} H(Z, E[ZX]) = \inf_{Q\in\Delta} H(Q, E_Q[X]) = \inf_{Q\in\Delta} \beta(Q, X).$$

Index	Description	Author(s)	$f_q(x)$
h	A scientist has index h if h of her papers have at least h citations	Hirsch (2005)	$q1_{(0,q]}(x)$
h^2	A scientist has index h^2 if h of her papers have at least h^2 citations	Kosmulski (2006)	$q^2 1_{(0,q]}(x)$
h_α	A scientist has index h_α if h of her papers have at least αh citations	van Eck and Waltman (2008)	$\alpha q 1_{(0,q]}(x), \alpha > 0$
w	A scientist has index w if w of her papers have at least $w, w-1, \ldots, 1$ citations	Woeginger (2008)	$(-x+q+1)1_{(0,q]}(x)$
A	A scientist has index $A = \sum_{x=1}^{h} X(x)/h$, where h is the h-index and $X(x)$ the citation curve	Jin (2006)	not SRM
R	A scientist has index $R = \sqrt{\sum_{x=1}^{h} X(x)}$, where h is the h-index and $X(x)$ the citation curve	Jin, Liang, Rousseau, and Egghe (2007)	not SRM
g	A scientist has index g, where g is the highest number of papers that together have at least g^2 citations	Egghe (2006)	not SRM

Table 1. Bibliometric indices. The table lists popular bibliometric indices, a short description of each index, and the author(s) who introduced the index. The h-, h^2-, h_α-, and w-index are Scientific Research Measures (SRMs) as defined in (1). The last column provides the functional form of the corresponding performance curves, $f_q(x)$. The A-, R-, and g-index are not SRMs.

	Evaluator 1		Evaluator 2		Evaluator 3		Evaluator 4	
	$\alpha(Q)$	$\beta(Q,X)$	$\alpha(Q)$	$\beta(Q,X)$	$\alpha(Q)$	$\beta(Q,X)$	$\alpha(Q)$	$\beta(Q,X)$
Q_0	1	8.8	1/3	26.3	1/3	26.3	1	8.7
Q_1	1	8.2	2/3	12.3	1	8.2	2/3	12.3
Q_5	1	8.5	1	8.5	1	8.5	1/3	25.5
$\Phi(X)$	8.2		8.5		8.2		8.7	

Table 2. Numerical example of Dual SRMs. The journal space Ω consists of three journals, $\Omega = \{\mathrm{MS}, \mathrm{MF}, \mathrm{FS}\}$. The scientist has three publications, one in each journal, with 10, 6, and 8 citations, respectively, i.e., her citation record $X(\omega)$ is given by $X(\mathrm{MS}) = 10$, $X(\mathrm{MF}) = 6$, and $X(\mathrm{FS}) = 8$. To assess her research performance, three criteria are expressed by the probability measures Q_0, Q_1, and Q_5. Criterion Q_0 expresses the number of years since the first journal's edition (the first journal's edition for MS, MF, and FS is 1954, 1991, and 1996, respectively). Criterion Q_1 expresses the 1-year impact factor of the journal (the 1-year impact factor for MS, MF, and FS is 1.73, 1.25, and 1.15, respectively). Criterion Q_5 expresses the 5-year impact factor of the journal (the 5-year impact factor for MS, MF, and FS is 3.3, 1.66, and 1.58, respectively). Each evaluator assigns different weights $\alpha(Q)$ to the three criteria and thus uses a different SRM. For each criterion Q, $\beta(Q, X) := \sup\{q \in \mathbb{R} \mid E_Q[X] \geq \gamma(Q, q)\}$, where $\gamma(Q, q) := q\,\alpha(Q)$. The Dual SRM gives the final research performance as $\Phi(X) = \min(\beta(Q_0, X), \beta(Q_1, X), \beta(Q_5, X))$.

	Full name	Faculty	Citations Avg	Citations Med	Citations Std	Publications Avg	Publications Med	Publications Std
Berkeley	Haas School of Business	8	90	21	179	85	61	61
Chicago	Chicago Booth	17	279	68	562	104	85	97
Columbia	Columbia Business School	34	121	40	237	101	63	108
Cornell	Johnson	8	72	23	142	181	82	302
Harvard	Harvard Business School	15	152	45	283	59	44	61
NYU	Stern School of Business	26	110	27	285	81	73	44
Penn.	Wharton, Univ. of Pennsylvania	21	138	48	267	77	52	67
Princeton	Bendheim Center for Finance	13	127	37	268	207	82	271
Stanford	Graduate School of Business	11	131	52	226	43	31	32
UCLA	Anderson School of Management	9	145	39	281	81	74	36
Yale	Yale School of Management	11	161	55	333	75	54	45
Total		173						

Table 3. Summary statistics of citations and publications. The dataset includes full professors affiliated to business schools or finance departments of 11 main U.S. universities. Faculty is the number of full professors for each university in our dataset. For each author, the dataset includes the number of publications with at least one citation each and the number of citations per publication, i.e., the citation curve. Avg, Med and Std are average, median and standard deviation, respectively, of citations and publications of all full professors for each university. Entries are rounded to the nearest integer. The dataset of citations and publications is from Google Scholar and extracted on September 6, 2012.

	Estimates			t-statistics			R^2
	q	c	k	q	c	k	
Berkeley	1,219.40	0.67	164.76	56.12	30.01	5.83	95.88
Chicago	5,245.49	0.54	1,411.10	49.81	25.93	7.76	95.94
Columbia	1,673.92	0.63	185.16	60.25	30.51	7.51	97.22
Cornell	1,267.08	0.63	209.31	56.95	42.16	4.89	97.11
Harvard	1,591.17	0.59	221.40	41.52	15.38	6.30	96.16
NYU	2,656.87	0.64	321.18	45.40	23.47	4.77	96.81
Penn.	1,975.27	0.61	672.08	55.40	27.21	7.05	97.13
Princeton	2,742.20	0.58	204.09	61.19	38.69	8.83	96.33
Stanford	5,156.73	0.35	4,159.60	26.01	11.19	4.81	96.56
UCLA	2,377.92	0.44	577.17	31.42	14.09	4.83	95.56
Yale	1,948.72	0.46	753.68	33.50	16.17	5.55	96.62

Table 4. Estimates of hyperbolic functions. For each author, the hyperbolic function $f_q(x) = q/x^c - k$ is fitted to the citation curve using nonlinear least squares, i.e., parameter estimates are obtained by minimizing $\sum_{x=1}^{p}(X(x) - f_q(x))^2$ with respect to q, c and k, where p is the number of publications with at least one citation each. R^2 is the adjusted R^2 in percentage. Table entries are obtained by averaging the corresponding values of all authors in each university. The dataset of citations includes 173 full professors affiliated to business schools or finance departments of 11 main U.S. universities. The dataset is extracted from Google Scholar on September 6, 2012.

	ϕ		h		h^2		h_α		w		A		R		g	
	Avg	Med	Avg	Med	Avg	Med	Avg	Med	Avg	Med	Avg	Med	Avg	Med	Avg	Med
Berkeley	1,504	1,489	27	25	10	10	35	32	47	41	185	160	69	65	56	44
Chicago	3,007	2,338	41	38	16	15	48	42	66	60	479	331	135	117	83	85
Columbia	1,778	1,415	38	30	12	11	46	36	63	50	229	186	88	79	73	57
Cornell	1,495	1,446	32	29	10	11	43	36	57	51	171	146	73	71	69	56
Harvard	1,705	1,626	28	23	11	10	33	26	45	34	247	239	79	68	48	44
NYU	1,768	1,428	33	31	12	12	40	38	57	52	238	186	85	81	71	67
Penn.	1,653	1,516	31	25	11	10	39	32	52	42	235	174	79	64	58	52
Princeton	2,404	1,853	51	49	14	13	67	62	91	82	296	278	119	94	106	82
Stanford	1,536	1,109	23	19	11	9	28	20	37	28	219	218	70	61	43	31
UCLA	1,962	2,267	37	34	13	15	45	40	63	53	286	320	100	113	76	61
Yale	1,726	1,706	30	29	12	11	36	33	49	44	274	244	85	84	59	54

Table 5. The Scientific Research measure and bibliometric indices. For each scholar in our dataset, ϕ-, h-, h^2-, h_α (with $\alpha = 0.5$), w-, A-, R-, and g-index are computed. The ϕ-index is based on hyperbolic performance curves and defined in Section 5. For each index, Avg and Med are average and median of the index of all scholars in a given university. Entries are rounded to the nearest integer. The dataset of citations covers 173 full professors affiliated to business schools or finance departments of 11 main U.S. universities. The dataset is extracted from Google Scholar on September 6, 2012.

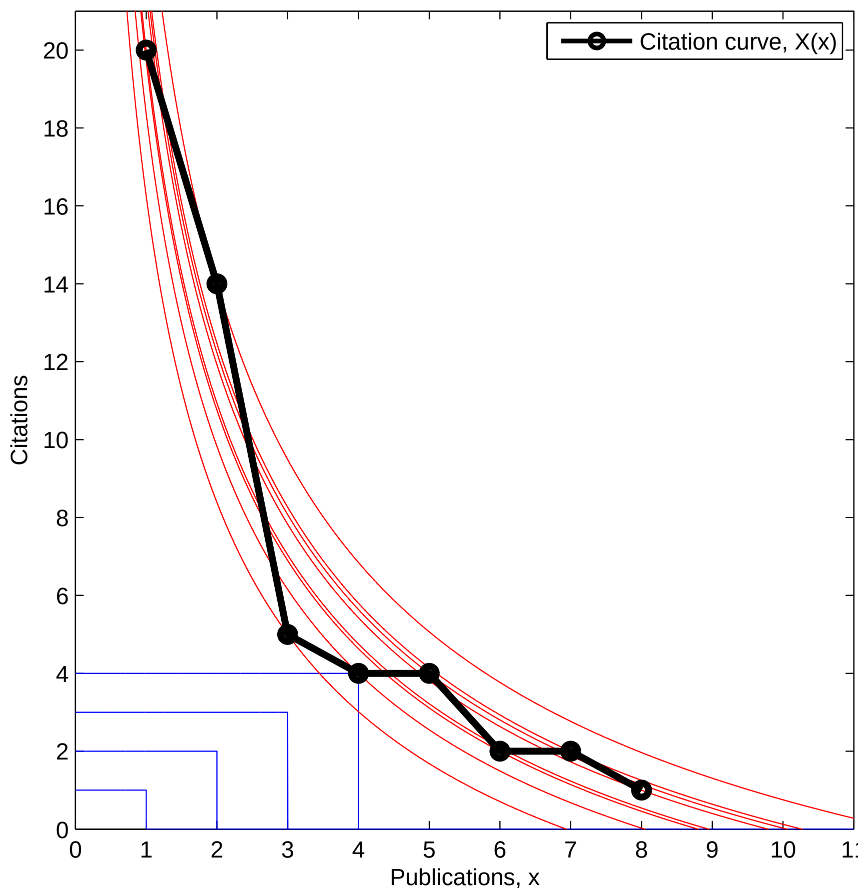

Figure 1. Citation curve. The graph shows the citation curve of a hypothetical scientist with 8 publications, as well as performance curves based on h-index and ϕ-index. The citation curve is obtained by ordering publications in decreasing order of citations. Performance curves of the h-index are given by $f_q(x) = q\,1_{(0,q]}(x)$, with $q = 1, 2, 3, 4$. Performance curves of the ϕ-index are given by $f_q(x) = q/x^{0.55} - 8.49$, with each curve passing through a different point $(x, X(x))$, with $x = 1, 2, \ldots, 8$. The hypothetical scientist has a h-index of 4 and ϕ-index of 24.75. The ϕ-index is determined by the performance curve passing through the point $(3, 5)$. This performance curve is the highest curve still below the citation curve X and indeed $24.75/3^{0.55} - 8.49 = 5$.

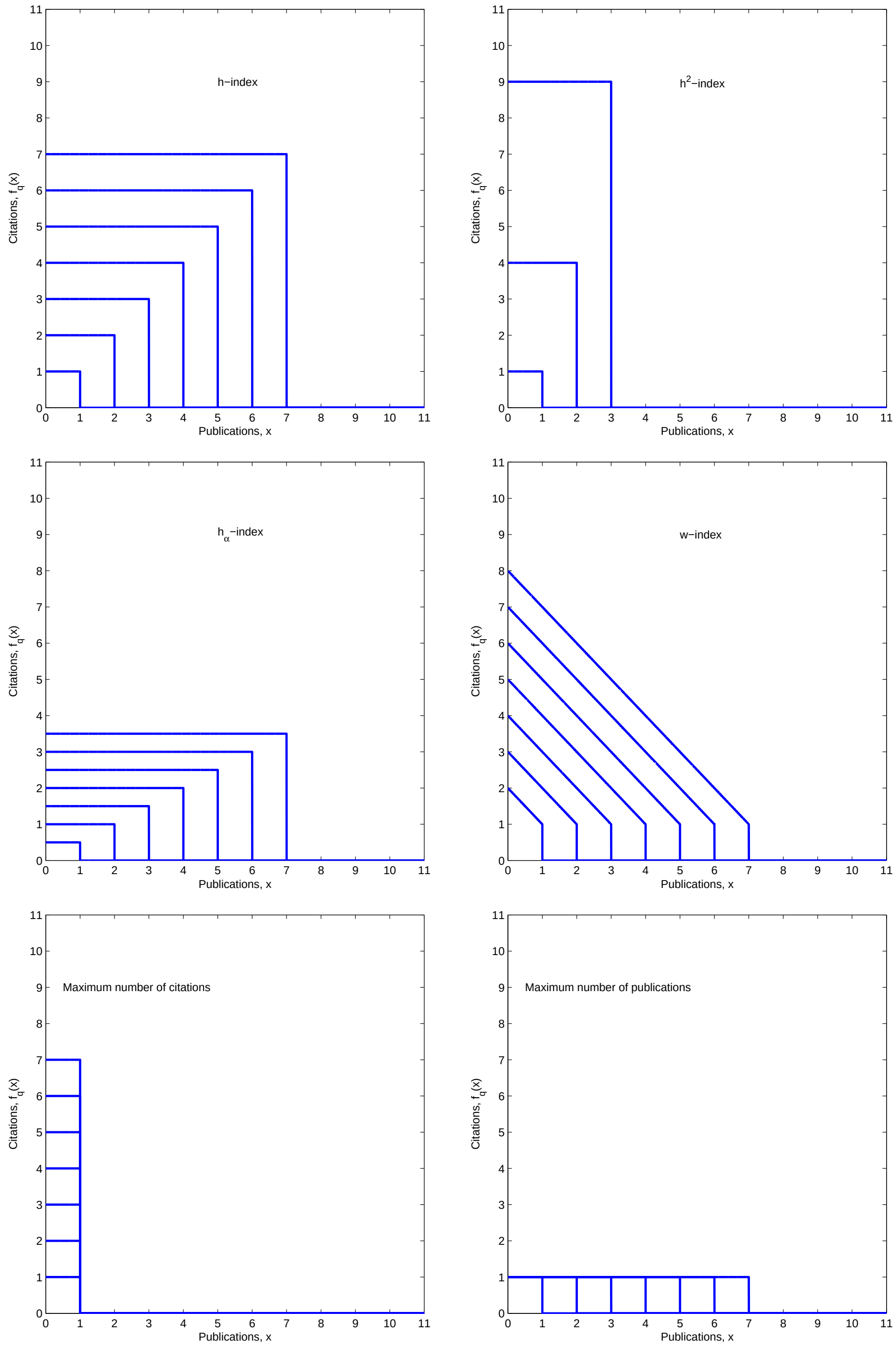

Figure 2. Performance curves. The graph shows six families of performance curves corresponding to six different bibliometric indices: h-index with performance curve $f_q(x) = q1_{(0,q]}(x)$ and $q = 1, 2, \ldots, 7$; h^2-index with performance curve $f_q(x) = q^2 1_{(0,q]}(x)$ and $q = 1, 2, 3$; h_α-index with performance curve $f_q(x) = \alpha q 1_{(0,q]}(x)$, $\alpha = 0.5$ and $q = 1, 2, \ldots, 7$; w-index with performance curve $f_q(x) = (-x + q + 1)1_{(0,q]}(x)$ and $q = 1, 2, \ldots, 7$; maximum number of citations with performance curve $f_q(x) = q1_{(0,1]}(x)$ and $q = 1, 2, \ldots, 7$; number of publications with at least one citation each with performance curve $f_q(x) = 1_{(0,q]}(x)$ and $q = 1, 2, \ldots, 7$.

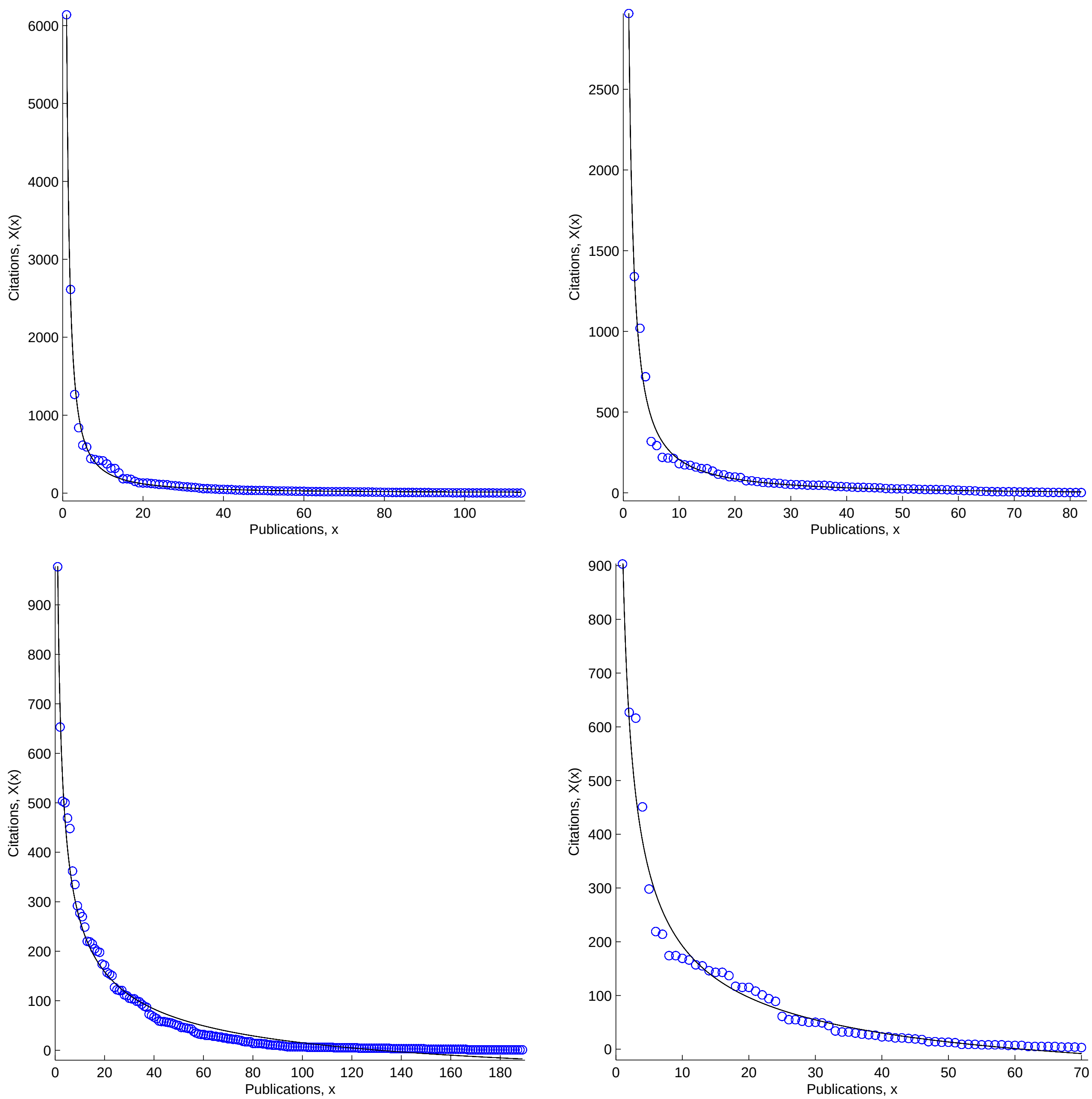

Figure 3. Typical fit of citation curves. Each graph shows one citation curve $X(x)$ (circles), i.e., the author's citations in decreasing order of citations, where $x = 1, \ldots, p$ and p is the number of publications with at least one citation. Superimposed is the fitted hyperbolic curve $f_q(x) = q/x^c - k$. For each author, parameter estimates of the hyperbolic curve are obtained using nonlinear least squares, i.e., by minimizing $\sum_{x=1}^{p}(X(x) - f_q(x))^2$ with respect to q, c and k. All adjusted R^2's are above 97.4%. Table 4 provides an overview of the fitting of such hyperbolic functions to all scholars' citation curves in our dataset.

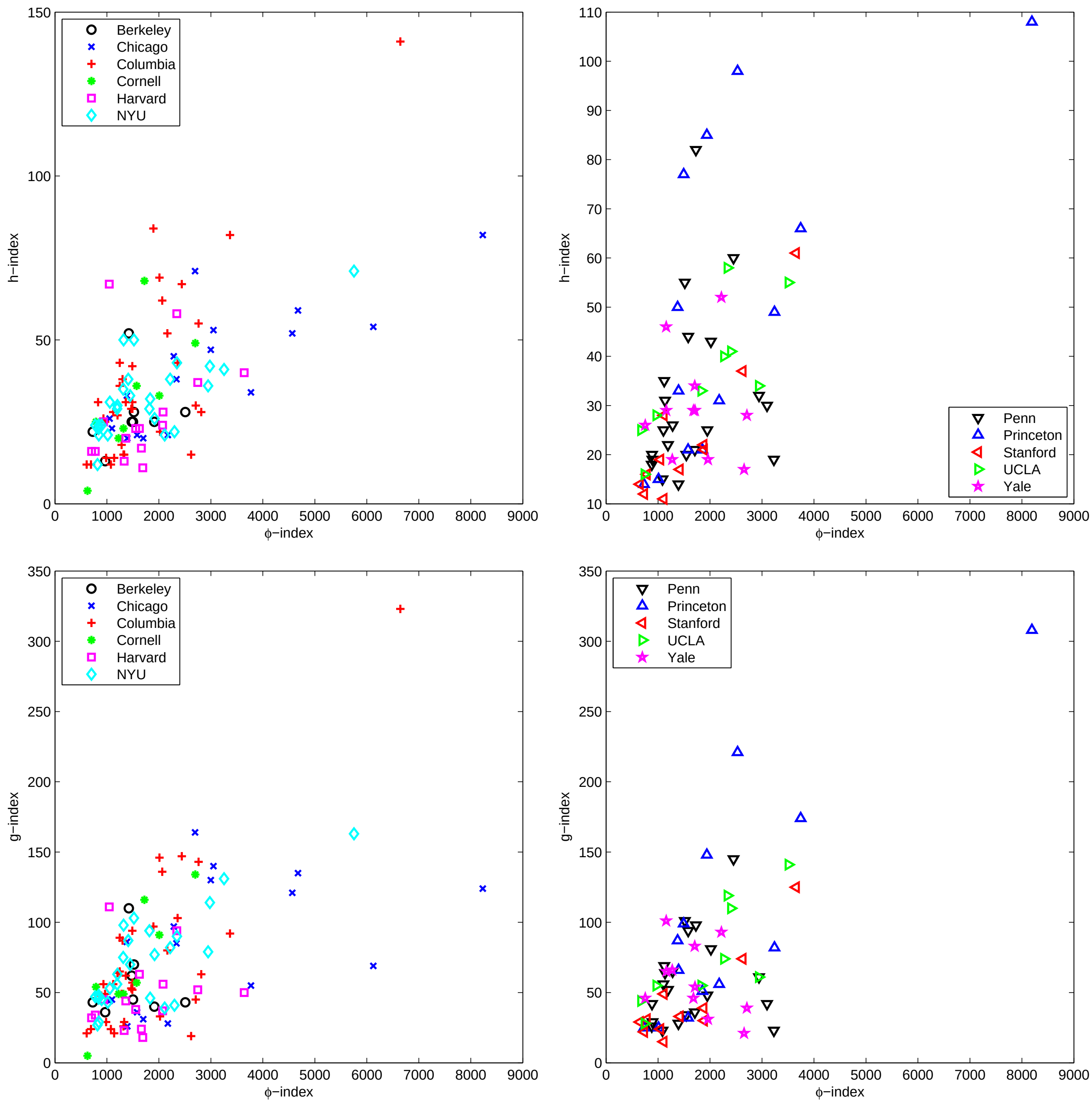

Figure 4. Scatter plots of ϕ-index versus h- and g-index. Upper graphs: scatter plot of ϕ-index (horizontal axis) versus h-index (vertical axis). Lower graphs: scatter plot of ϕ-index (horizontal axis) versus g-index (vertical axis). Those indexes are computed for each one of the 173 scholars in our dataset and displayed according to the university the scholar belongs to. Relations between ϕ-index and h-index (respectively g-index) are not monotone. Thus rankings of scholars given by the ϕ-index and h-index (respectively g-index) are different. The dataset of citations covers 173 full professors affiliated to business schools or finance departments of 11 main U.S. universities. The dataset is extracted from Google Scholar on September 6, 2012.

References

Adler, R., J. Ewing, and P. Taylor, 2008, "Citation Statistics," Joint Committee on Quantitative Assessment of Research, International Mathematical Union.

Alonso, S., F. 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.

——— , 2010, "hg-index: A New Index to Characterize the Scientific Output of Researchers Based on the h- and g-indices," *Scientometrics*, 82, 391–400.

American Scientist Open Access Forum, 2008, *Comments to the Citation Statistics Report*. Citation Statistics: International Mathematical Union Report, U.S.

Anderson, T., K. Hankin, and P. Killworth, 2008, "Beyond the Durfee Square: Enhancing the h-index to Score Total Publication Output," *Scientometrics*, 76, 577–588.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath, 1999, "Coherent Measure of Risk," *Mathematical Finance*, 9, 203–228.

Bornmann, L., R. Mutz, and H.-D. Daniel, 2008, "Are There Better Indices for Evaluation Purposes than the h-index? A Comparison of Nine Different Variants of the h-index Using Data from Biomedicine," *Journal of the American Society for Information Science and Technology*, 59, 830–837.

Borokhovich, K., R. Bricker, K. Brunarski, and B. Simkins, 1995, "Finance Research Productivity and Influence," *Journal of Finance*, 50, 1691–1717.

Cabrerizo, F., S. Alonso, E. Herrera-Viedma, and F. Herrera, 2010, "q^2-index: Quantitative and Qualitative Evaluation Based on the Number and Impact of Papers in the Hirsch Core," *Journal of Informetrics*, 4, 23–28.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio, 2011, "Risk Measures: Rationality and Diversification," *Mathematical Finance*, 21, 743–774.

Cherny, A., and D. Madan, 2009, "New Measures for Performance Evaluation," *Review of Financial Studies*, 22, 2571–2606.

Chung, K., and R. Cox, 1990, "Patterns of Productivity in the Finance Literature: A Study of the Bibliometric Distributions," *Journal of Finance*, 45, 301–309.

Drapeau, S., and M. Kupper, 2013, "Risk Preferences and Their Robust Representation," *Mathematics of Operations Research*, 38, 28–62.

Egghe, L., 2006, "Theory and Practice of the g-index," *Scientometrics*, 69, 131–152.

Egghe, L., and R. Rousseau, 2008, "An h-index Weighted by Citation Impact," *Information Processing and Management*, 44, 770–780.

Föllmer, H., and A. Schied, 2002, "Convex Measures of Risk and Trading Constraints," *Finance and Stochastics*, 6, 429–447.

——— , 2004, *Stochastic Finance: An Introduction in Discrete Time*. de Gruyter Studies in Mathematics 27, Berlin, 2nd edn.

Frittelli, M., and M. Maggis, 2011, "Dual Representation of Quasiconvex Conditional Maps," *SIAM Journal on Financial Mathematics*, 2, 357–382.

Frittelli, M., M. Maggis, and I. Peri, 2013, "Risk Measures on $\mathcal{P}(\mathbb{R})$ and Value at Risk with Probability/Loss Function," *Mathematical Finance*, forthcoming.

Frittelli, M., and E. Rosazza Gianin, 2002, "Putting Order in Risk Measures," *Journal of Banking and Finance*, 26, 1473–1486.

Garfield, E., 2005, "The Agony and the Ecstasy – The History and Meaning of the Journal Impact Factor," International Congress on Peer Review and Biomedical Publication (Chicago).

Guns, R., and R. Rousseau, 2009, "Real and Rational Variants of the h-index and the g-index," *Journal of Informetrics*, 3, 64–71.

Hauser, J., 1998, "Research, Development, and Engineering Metrics," *Management Science*, 44, 1670–1689.

Hirsch, J., 2005, "An Index to Quantify an Individual's Scientific Research Output," *Proceedings of the National Academy of Sciences*, 102, 16569–16572.

——— , 2007, "Does the h Index Have Predictive Power?," *Proceedings of the National Academy of Sciences*, 104, 19193–19198.

Jin, B., 2006, "h-index: An Evaluation Indicator Proposed by Scientist," *Science Focus*, 1, 8–9.

Jin, B., L. Liang, R. Rousseau, and L. Egghe, 2007, "The R- and AR-indices: Complementing the h-index," *Chinese Science Bulletin*, 52, 855–863.

Kalaitzidakis, P., T. Stengos, and T. Mamuneas, 2003, "Rankings of Academic Journals and Institutions in Economics," *Journal of the European Economic Association*, 1, 1346–1366.

Kosmulski, M., 2006, "A New Hirsch-type Index Saves Time and Works Equally Well as the Original h-index," *ISSI Newsletter*, 2, 4–6.

Lovegrove, B., and S. Johnson, 2008, "Assessment of Research Performance in Biology: How Well Do Peer Review and Bibliometry Correlate?," *BioScience*, 58, 160–164.

Marshall, A. W., I. Olkin, and B. C. Arnold, 2011, *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York, 2nd edn., Springer Series in Statistics.

May, R., 1997, "The Scientific Wealth of Nations," *Science*, 275, 793–796.

Panaretos, J., and C. Malesios, 2009, "Assessing Scientific Research Performance and Impact with Single Indices," *Scientometrics*, 81, 635–670.

Ruane, F., and R. Tol, 2008, "Rational (Successive) h-indices: An Application to Economics in the Republic of Ireland," *Scientometrics*, 75, 395–405.

Schreiber, M., 2010, "Twenty Hirsch Index Variants and Other Indicators Giving More or Less Preference to Highly Cited Papers," *Annals of Physics*, 522, 536–554.

van Eck, N., and L. Waltman, 2008, "Generalizing the h- and g-indices," *Journal of Informetrics*, 2, 263–271.

Vinkler, P., 2009, "The π-index: A New Indicator for Assessing Scientific Impact," *Journal of Information Science*, 35, 602–612.

Volle, M., 1998, "Duality for the Level Sum of Quasiconvex Functions and Applications," *Control, Optimisation and Calculus of Variations*, 3, 329–343.

Wiberley, S., 2003, "A Methodological Approach to Developing Bibliometric Models of Types of Humanities Scholarship," *Library Quarterly*, 73, 121–159.

Woeginger, G., 2008, "An Axiomatic Characterization of the Hirsch-index," *Mathematical Social Sciences*, 56, 224–232.

Zhang, C.-T., 2009, "The e-Index, Complementing the h-Index for Excess Citations," *PLoS ONE*, 4, e5429.