

# Scientific Research Measures

Marco Frittelli\*  
Department of Mathematics  
University of Milan

Loriano Mancini‡  
Swiss Finance Institute  
at EPFL

Ilaria Peri‡  
ESC Rennes  
School of Business

January 29, 2015

\*Marco Frittelli, Department of Mathematics, University of Milan, Via Saldini 50, I-20133 Milan, Italy. E-mail: [marco.frittelli@unimi.it](mailto:marco.frittelli@unimi.it).

‡Loriano Mancini, College of Management, EPFL, Quartier UNIL-Dorigny, CH-1015 Lausanne, Switzerland. E-mail: [loriano.mancini@epfl.ch](mailto: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.com](mailto:ilaria.peri@esc-rennes.com).

## Abstract

The evaluation of scientific research is crucial for both the academic community and society as a whole. Numerous bibliometric indices have been proposed for the ranking of research performance, mainly on an ad hoc basis. We now introduce the novel class of Scientific Research Measures (SRMs) to rank scientists' research performance and provide a rigorous theoretical foundation for these measures. In contrast to many bibliometric indices, SRMs take into account the whole citation curve of the scientist, offer appealing structural properties, allow a finer ranking of scientists, correspond to specific features of different disciplines, research areas and seniorities, and include several bibliometric indices as special cases. Thus SRMs result in more accurate rankings than ad hoc bibliometric indices. We also introduce the further general class of Dual SRMs that reflect the “value” of journals and permit the ranking of research institutions based on theoretically sound criteria, which has been a central theme in the scientific community over recent decades. An empirical application to the citation curves of 173 finance scholars shows that SRMs can be easily calibrated to actual citation curves and generate different authors' rankings than those produced by seven traditional bibliometric indices.

**Keywords:** Research Performance, Scientific Impact Measures, Bibliometric Indices, Citations, Calibration

**JEL Codes:** C02, I20

## Introduction

Over the last decades, the evaluation of scientific research performance has become increasingly important. Crucial decisions such as the funding of research projects or faculty recruitment depend to a large extent upon the scientific merits of the researchers involved. The selection of inappropriate valuation criteria can have deleterious effects on the performance assessment of scientists and structures, such as departments or laboratories, distort funding allocation and recruiting processes, and ultimately have a negative impact on society as a whole.

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

Because of the rapid increase in availability and quality of online databases (e.g., Google Scholar, Thomson Reuters' Web of Science, Scopus, MathSciNet), bibliometric valuation has become widely popular and numerous bibliometric indices have been proposed 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 great deal of attention from the scientific community and been widely used to assess research performance in many fields, e.g., Bornmann and Daniel (2007). Various authors have recently proposed several extensions of the *h-index*, mainly on an ad hoc basis, to improve on some of its shortcomings.<sup>1</sup>

This paper provides three contributions to the fast-growing literature on research valuation. The first contribution is to introduce a novel class of scientific performance measures designated Scientific Research Measures (SRMs). The main features of SRMs are that they take into account the scientist's whole citation curve<sup>2</sup> and rank research performance using a novel tool designated *performance curves* for research performance, similar to isotherm lines for temperature. Importantly, performance curves can be chosen in a flexible way, reflecting the scientists' seniority and characteristics of the research fields, such as typical number of publications and citation rates. This flexibility is lacking in virtually all existing bibliometric indices. SRMs are derived from a "calibration" approach in the sense that they are informed by actual citation data.

The second contribution of this paper is to introduce a further general class of research measures designated Dual SRMs. We provide a rigorous theoretical foundation for Dual SRMs building on the well-established theory of risk measures.<sup>3</sup> Dual SRMs are rooted in 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).<sup>4</sup>

Dual SRMs are relevant for three reasons. First, if two scientists have the same citation curve, any traditional bibliometric index is unable to rank them, whereas Dual SRMs can rank those scientists as soon as their publications appear in different journals. Second, additional information, such as journals' impact factors, can be taken into account in a theoretically consistent manner in research rankings based on Dual SRMs. Third, and perhaps most importantly, Dual SRMs are able to achieve a meaningful aggregation of scientists' research outputs. This in turn allows the ranking of the research performance of teams, departments and academic institutions using theoretically sound criteria. The issue of ranking research institutions has been a central theme in the scientific community over recent decades, e.g., Borokhovich, Bricker, Brunarski and Simkins (1995), Kalaitzidakis, Mamuneas and Stengos (2003), and Bornmann, Stefaner, de Moya Anegón and Mutz (2014).

The third contribution of this paper is to present an extensive empirical application of SRMs. Using Google Scholar, we construct a novel data set of citation curves for 173 finance scholars affiliated to 11 major US finance departments. We observe that power law performance curves fit actual citation curves well. Therefore we compare research rankings based on the SRM characterized by power law performance curves with seven other existing bibliometric indices. We find that the rankings of authors and universities based on the SRM are quite different from those based on the other bibliometric indices. The largest discrepancies concern "average" performing scholars and universities (in relative terms) that are arguably the most difficult to rank. Under the premise that research performance is reflected by the whole citation curve rather than only part of it, SRMs provide a better ranking of research performance than traditional bibliometric indices. Finally, we develop a simple two-step algorithm to operationalize SRMs.

## Related Literature

Since the early nineties, the use of bibliometric indices has encompassed many fields, e.g., Chung and Cox (1990), May (1997), Hauser (1998), Wiberley (2003), and Kalaitzidakis et al. (2003). The effectiveness of citation-based metrics has been 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.<sup>5</sup>

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 the number of publications and the number of citations in a single bibliometric index. One limitation of the *h*-index is its insufficient ability to discriminate between scientists' research performances. The *h*-index is fully determined by the *Hirsch core*, i.e., the set of the most cited publications with at least *h* citations each. While the *h*-index is insensitive to publications with high and low numbers of citations, these publications may well contain valuable information for ranking scientific performances. The report by the Joint Committee on Quantitative Assessment of Research<sup>6</sup> (Adler, Ewing, & Taylor, 2008) confirmed this point by a well-documented example and encouraged the use of more complex measures.<sup>7</sup>

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 these numerous proposals into two broad categories:  $h$ -index variants, which suggest alternative methods to determine the core size of the most cited publications, and  $h$ -index complements, which include additional information in the  $h$ -index especially concerning the most cited publications.

Among the  $h$ -index variants, one of the first proposals was the  $g$ -index by Egghe (2006). The  $g$ -index is defined as the largest number  $g$  such that the most cited  $g$  articles received 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. Compared to the  $h$ -index, the  $h^2$ -index is probably more appropriate in fields where the number of citations per article is relatively high, such as chemistry or physics, while the  $h$ -index seems to be more appropriate for mathematics or astronomy for instance. Many studies point out the arbitrariness of the definition of the citation core size. To attenuate this problem, 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 receive 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 their publications have at least  $w, w-1, \dots, 1$  citations. Ruane and Tol (2008) propose the rational  $h$ -index ( $hrat$ -index) that has the advantage of providing more granularity in the evaluation process since it increases in smaller increments than the original  $h$ -index. Rousseau (2006, 2014) proposed the real (or interpolated)  $h$ -index, and Guns and Rousseau (2009) review several real and rational variants of the  $h$ - and  $g$ -indices.

The  $h$ -index complements seek to measure the core citation intensity. For example, Jin (2006) proposed computing 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  $h_w$ -indices can be very sensitive to only a few publications with many citations. In an attempt to make these indices more robust, Bornmann, Mutz, and Daniel (2008) propose calculating 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 seems 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 and not just those in the Hirsch core. As alternative, Vinkler (2009) proposes the  $\pi$ -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  $g$ -indices ( $hg$ -index by Alonso, Cabrerizo, Herrera-Viedma and Herrera (2010)), and the geometric mean of the  $h$ - and  $m$ -indices ( $q^2$ -index by Cabrerizo, Alonso, Herrera-Viedma and Herrera (2010)).

The main conclusion that emerges from this literature review is that many bibliometric indices have been proposed 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 lead to very similar rankings.

Recently, Deineko and Woeginger (2009), Marchant (2009), and Ravallion and Wagstaff (2011) introduced new families of bibliometric indices, generalizing  $h$ -index type indices, using scoring rules, and using influence functions, respectively, in well-defined mathematical frameworks.

We depart from the literature above by adopting a “calibration” approach, i.e., starting from observed citation curves, rather than improving existing bibliometric indices. The main motivation for our approach is that scientists of different fields and seniorities have significantly different citation curves. These differences are important and need to be reflected in the ranking process. Our SRMs take into account those specific features using flexible performance curves calibrated to actual data, as discussed in the next section.

### Scientific Research Measures

We now introduce the class of Scientific Research Measures. 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 eight publications.

Insert Figure 1 here

We model the scientist’s citation curve,  $X$ , as a function  $X: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , mapping each publication into the corresponding number of citations. By construction,  $X$  is bounded, non-increasing, and non-negative. While the theory of SRMs holds when the domain and range of  $X$  are positive integers, we embed  $X$  in the positive reals for simplicity.<sup>8</sup>

To rank scientists’ research performances we introduce a new tool designated *performance curves*. The performance curve represents the amount of citations required to reach a performance level  $q$ . Intuitively, the higher the scientist’s research performance, the higher the performance curve  $f_q$  that they can reach, and the higher the corresponding level  $q$ . Formally, the performance curve is a function  $f_q: \mathbb{R}_+ \rightarrow \mathbb{R}$  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  $x$ . Although not necessary for the theory, we also assume that  $f_q(x)$  is non-increasing in  $x$ , like any citation curve.

When  $q$  varies in a given set  $I \subseteq \mathbb{R}_+$ , we obtain a family of performance curves  $\{f_q\}_{q \in I}$ . Typically, we will consider  $I = \{1, 2, 3, \dots\} \subset \mathbb{N}$  or  $I = [1, \infty]$ . Figure 1 shows two families of performance curves: square type functions corresponding to the  $h$ -index and power law type functions corresponding to a specific SRM that will be used in our empirical application. Moving from the origin  $(0, 0)$  of the graph in a northeasterly direction, the research performance level increases, i.e., each performance curve  $f_q$  is associated to a higher and higher level  $q$ . The four performance curves in Figure 1 associated to the  $h$ -index correspond to an  $h$ -index of 1, 2, 3 and 4, respectively. Similarly, each power law type performance curve is associated to an increasing level of research performance.

Our novel class of SRMs is defined by the following  $\phi$ -index:

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

where  $:=$  means defined as. The  $\phi$ -index is obtained by comparing the whole citation curve  $X$  and the family of performance curves  $\{f_q\}_{q \in I}$ . Specifically, the  $\phi$ -index is the highest level  $q$  of the performance curve still below the author's citation curve. Consider again the power law type performance curves in Figure 1. The  $\phi$ -index is 24.75 and determined by the performance curve passing through the point (3, 5). Generally speaking, numerical values of the  $\phi$ -index have no interpretation and are simply used to rank scientists' research performances. However, the  $\phi$ -index can be normalized in order to interpret it.<sup>9</sup>

The family of performance curves  $\{f_q\}_{q \in I}$  plays a key role in the definition of SRMs. Performance curves create the common ground or "level the playing field" for the ranking of scientists' research performances. The logic behind our approach is the following. Citation curves of researchers (with similar seniority) in a given field, say mathematics, tend to be quite similar. Across fields, however, say comparing mathematics and finance, the citation curves of researchers can be very different. In mathematics, researchers typically have a large number of publications with a small number of citations per publication. In finance, exactly the opposite happens. Thus different fields are characterized by different publication and citation practices. In our view these features of the research field, and seniority, should be reflected in the research ranking process, in a theoretically coherent manner. We propose to achieve this by calibrating performance curves to observed citation curves. Through statistical estimation, calibrated performance curves would then reflect the publication and citation practices in a given field. These performance curves vary across fields (area of research, seniority of researchers).

A central feature of our method is that research performance is reflected by the whole citation curve, and not just part of it. We can illustrate this point with a simple example. Suppose two researchers, A and B, have a similar number of publications with similar citations, but researcher A has some additional, lightly cited publications compared to B. It appears that A should be ranked higher than B, but generally this ranking can be established only by considering the whole citation curves. In fact, if the additional, lightly cited publications of researcher A are disregarded, the two researchers would have very similar citation curves. Along similar lines, in many real cases it can be important to consider the whole citation curves in order to rank researchers correctly.

Notice that by taking into account the whole citation curve, our  $\phi$ -index does not necessarily penalize researchers with new or uncited publications. For example, if the eighth publication in Figure 1 had zero citations rather than one citation, the  $\phi$ -index would not change.

Under the premise that research performance is reflected by the whole citation curve, the issue is now to rank such curves. We tackle this by using our novel tool, namely the performance curves. Given that the citation curves of researchers of a given field and seniority are rather similar, it seems natural to us that performance curves should mimic the observed citation curves. These calibrated performance curves offer two advantages. First, as mentioned above, calibrated performance curves would reflect specific features of the field and seniority of researchers. Second, the calibration method is essentially an objective, data-driven procedure for the selection of performance curves, which is not significantly exposed to subjective or arbitrary judgments.

When performance curves are not calibrated to observed citation curves, at least two issues may arise. The first is the arbitrariness in the selection of performance curves. These curves may have many shapes and it is a priori unclear which curves should be used to rank research

performance. As shown in our empirical application, different shapes of performance curves lead to different authors' rankings, and the question remains as to which ranking should be used. The second is that arbitrarily chosen performance curves may not reflect important features of the research field and seniority of researchers.

As already mentioned, performance curves in (1) can be specified in a flexible way, reflecting features of the scientific field, research areas, and seniority. Unfortunately most existing bibliometric indices lack such flexibility and their performance curves do not mimic observed citation curves. Figure 1 suggests that power law type performance curves mimic the citation curve  $X$  much better than square type performance curves do. As shown in our empirical application, observed citation curves exhibit power law shapes, calling for power law performance curves.

While citation curves are always non-negative, the power law type performance curves in Figure 1 become negative for large  $x$ . This poses no problem to the  $\phi$ -index and allows the taking into account of the scientist's publications that have zero citations, i.e., those publications for which  $X(x) = 0$ .<sup>10</sup>

### *Monotonicity and Quasi-Concavity of SRMs*

The class of SRMs defined in (1) has two desirable properties: monotonicity and quasi-concavity. The first property simply indicates that better scientists, as reflected by their citation curves, have a higher  $\phi$ -index. Formally, 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, the monotonicity of the  $\phi$ -index is well justified.

If SRMs were characterized by monotonicity only, they would be of limited use. This is because the monotonicity property only allows the ranking of scientists whose citation curves satisfy a dominance order, like  $X_1 \leq X_2$ . Evaluators often need to rank scientists whose citation curves intersect each other. The quasi-concavity of the  $\phi$ -index enables scientists to be ranked 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  $\phi$ -index implies that, for all  $\lambda \in [0, 1]$ ,

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

Quasi-concavity indicates that a convex combination of two citation curves leads to a performance level which is at least as high as the research performance of the less well performing scientist. In other words, when the citation curve of a less well 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 well performing scientist. This seems to be a natural property of the SRM. Appendix A shows that the  $\phi$ -index in (1) satisfies (2), i.e., it is quasi-concave. Moreover, quasi-concavity ensures an internal consistency or coherency of scientists' rankings and Appendix A provides a numerical example to illustrate this point. The appendix is available online at <http://www.ssrn.com> or from the authors upon request.

### *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 appropriately 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  $1_{(0,q]}(x) = 1$  when  $x \in (0, q]$  and 0 otherwise. The performance curves of the real  $h$ -index are the same as those of the  $h$ -index but the  $q$  parameter varies in the real numbers. The maximum number of citations, as bibliometric index, is characterized by the performance curve  $f_q(x) = q1_{(0,1]}(x)$ . Similarly, the number of publications with at least one citation has performance curves given by  $f_q(x) = 1_{(0,q]}(x)$ . Table 1 shows the  $h^2$ -,  $h_q$ - and  $w$ -indices and their performance curves.

Insert Table 1 here

Not all bibliometric indices proposed in the literature are SRMs. The  $\phi$ -index in (1) has three features: it takes into account the whole citation curve, it is defined in terms of performance curves, and it is monotone and quasi-concave. Any bibliometric index that lacks any of these features is not an SRM. For example, the  $A$ -,  $R$ - and  $g$ -indices (defined in Table 1) are not SRMs since they do not take into account the scientist's whole citation curve.

Figure 2 shows six families of performance curves associated to six different bibliometric indices; see also Table 1. As citation curves have typically power law shapes, none of these performance curves provides an adequate description of actual citation curves.

Insert Figure 2 here

## Dual Scientific Research Measures

We now introduce a further general class of research measures designated Dual SRMs. The goal is to reflect additional information in the ranking process in a theoretically consistent manner. We define a scientist's *citation record* as the number of citations of publications in each journal. Dual SRMs rank citation records.<sup>11</sup> Dual SRMs play the important role of allowing the aggregation of scientists' research outputs in a meaningful way by aggregating citation records, namely the citations of publications of different scientists in the same journal. This in turn enables research teams, departments and academic institutions to be ranked in a theoretically sound way. The ranking of institutions has 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.

### *Numerical Example of Dual SRMs*

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

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

Consider the following three journals: Management Science (MS), Mathematical Finance (MF), and Finance and Stochastics (FS). One possible way of assessing a journal's value is to use the two-year and five-year journal's impact factor.<sup>12</sup> For MS, MF, and FS, the two-year

impact factors are 1.73, 1.25, and 1.15, and the five-year impact factors are 3.30, 1.66, and 1.58, respectively.<sup>13</sup> The information expressed by each impact factor is collected in a probability measure,  $Q_2$  and  $Q_5$ , respectively, on the journal space  $\Omega = \{MS, MF, FS\}$ . For example,  $Q_2(MS) = 1.73/(1.73 + 1.25 + 1.15)$ , and similarly for the other journals and impact factors.

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

The second step is to assess the scientist's research performance relating to each criterion or probability measure  $Q_2$  and  $Q_5$ . To that end, we introduce the function  $\gamma(Q_2, m)$  that gives the smallest  $Q_2$ -average of citations to reach a performance level of  $m$ . Intuitively, if the evaluator deems the criterion  $Q_5$  as highly important, then  $\gamma(Q_5, m)$  will be high, and it will be a challenging task for each scientist to do well according to that criterion. If  $Q_5$  is more important than  $Q_2$ , then  $\gamma(Q_5, m) > \gamma(Q_2, m)$ .

Hence, under  $Q_2$  the research performance of a scientist with citation record  $X(\omega)$  is

$$\beta(Q_2, X) := \sup\{m \in \mathbb{R} / E_{Q_2}[X] \geq \gamma(Q_2, m)\}. \quad (3)$$

Using citation record and impact factor, the value of  $E_{Q_2}[X]$  can be easily computed. In our example  $E_{Q_2}[X] = 10 Q_2(MS) + 6 Q_2(MF) + 8 Q_2(FS) = 8.2$ .

Suppose the evaluator has no preference between  $Q_2$  and  $Q_5$ . Then  $\gamma(Q, m)$  can be set to a constant value, i.e.,  $\gamma(Q, m) := m$ , for any  $Q$ , and thus

$$\beta(Q_2, X) := \sup\{m \in \mathbb{R} / E_{Q_2}[X] \geq m\} = E_{Q_2}[X] = 8.2.$$

If  $Q_2$  were the only criterion to assess journal quality, then  $\beta(Q_2, X) = 8.2$  was the scientist's final research performance. Since  $Q_5$  is another criterion, the evaluator also computes the scientist's research performance under  $Q_5$ , obtaining a corresponding index of  $\beta(Q_5, X)$ . In our example  $E_{Q_5}[X] = 10 Q_5(MS) + 6 Q_5(MF) + 8 Q_5(FS) = 8.5$ , and as  $\gamma(Q_5, m) := m$ , then  $\beta(Q_5, X) = E_{Q_5}[X] = 8.5$ .

Finally, the evaluator takes the smaller value between  $\beta(Q_2, X)$  and  $\beta(Q_5, X)$ . The Dual SRM gives the scientist's final research performance

$$\Phi(X) := \min(\beta(Q_2, X), \beta(Q_5, X)) = \min(E_{Q_2}[X], E_{Q_5}[X]) = 8.2.$$

This example shows that the Dual SRM leads to a conservative or "worst-case scenario" assessment of the research performance. The evaluator assesses the research performance using the least favorable journals' evaluation criterion for the scientist. Generally, there will be no consensus on the journals' evaluation criteria. Opting for a conservative assessment of research performance seems to be a sensible approach. To assess the research performance of another scientist, the evaluator repeats the steps above using the same  $Q_2$  and  $Q_5$  criteria. Comparing the scientists'  $\Phi$ -indices gives the scientists' research rankings.

Table 2 extends the numerical example above by introducing an additional journal evaluation criterion and considering four evaluators who assign different weights to the three criteria. The additional criterion is the number of years since the journal's first edition expressed by the probability measure  $Q_0$ .<sup>14</sup> Under  $Q_0$ ,  $E_{Q_0}[X] = 10 Q_0(\text{MS}) + 6 Q_0(\text{MF}) + 8 Q_0(\text{FS}) = 8.8$ .

Insert Table 2 here

The evaluators' decision to assign different weights to the three criteria can be expressed by choosing  $\gamma(Q, m) := m \alpha(Q)$ , where  $\alpha(Q) \geq 0$  represents the weight that the evaluator assigns to criterion  $Q$ . The more important the criterion, the higher the  $\alpha(Q)$ . A natural range for the weights is the interval  $[0, 1]$  but weights larger than one are also admissible. If a particular criterion is controversial or deemed unimportant, it can be attributed a weight of zero.

As in the initial example, for each evaluation criterion  $i = 0, 1, 5$ ,

$$\beta(Q_i, X) := \sup\{m \in \mathbb{R} / E_{Q_i}[X] \geq m \alpha(Q_i)\} = E_{Q_i}[X] / \alpha(Q_i)$$

and the scientist's final research performance  $\Phi$ -index is

$$\begin{aligned} \Phi(X) &:= \min(\beta(Q_0, X), \beta(Q_2, X), \beta(Q_5, X)) \\ &= \min(E_{Q_0}[X] / \alpha(Q_0), E_{Q_2}[X] / \alpha(Q_1), E_{Q_5}[X] / \alpha(Q_5)). \end{aligned}$$

### *Theoretical Derivation of Dual SRMs*

We now present a theoretical derivation of the Dual SRM. Formally, the citation record is a random variable  $X(\omega)$  defined on the events  $\omega \in \Omega$ , where each event  $\omega$  corresponds to the journal in which the publication appeared, and  $\Omega$  is the set of journals. Thus,  $X(\omega)$  is the number of citations of the publications in journal  $\omega$ .<sup>15</sup>

We fix a family of probabilities  $\mathcal{P}$  defined on  $\Omega$ , where for each  $Q \in \mathcal{P}$ ,  $Q(\omega)$  represents the value of journal  $\omega \in \Omega$ . A priori there will be no consensus regarding how journals are evaluated and this is reflected in the set  $Q \in \mathcal{P}$  rather than just a single probability measure.

We then select a family of functions  $\{\gamma(Q, m)\}_{m \in \mathbb{R}}$ . Each function  $\gamma(\cdot, m) : \mathcal{P} \rightarrow \mathbb{R}$  associates to each measure  $Q$  the value  $\gamma(Q, m)$ , i.e., the smallest  $Q$ -average of citations to reach a research performance level  $m$  under  $Q$ . The Dual SRM is given by the  $\Phi$ -index

$$\Phi(X) := \inf_{Q \in \mathcal{P}} \beta(Q, X) \tag{4}$$

where  $\beta(Q, X) := \sup\{m \in \mathbb{R} / E_Q[X] \geq \gamma(Q, m)\}$  represents the research performance of a scientist with citation record  $X$ , under the fixed measure  $Q$ . The function  $\gamma(Q, m)$  does not depend on the scientist's citation record  $X$  and is non-decreasing in  $m$ .

Any Dual SRM in (4) is monotone increasing and quasi-concave, as proved in Proposition 2 in Appendix B. The appendix is available online at <http://www.ssrn.com> or from the authors upon request. Both properties have a clear motivation. The monotonicity indicates that better scientists have higher research performances. The quasi-concavity indicates that the

performance of a research team is at least equal to the scientific performance of the less well performing member. Notably, quasi-concavity leads to sound rankings of research teams and institutions. This is because the citation records of two (or more) scientists can be aggregated by simply taking a linear convex combination  $\lambda X_1 + (1 - \lambda)X_2$ , where  $\lambda \in [0, 1]$ .<sup>16</sup> The aggregation is meaningful since it is at the level of citation records, i.e., scientists' citations of publications in the same journal.

An important question is: Does any monotone increasing and quasi-concave map admit the representation in (4)? Essentially, the answer is positive and formalized in Theorem 1. To prove the theorem we assume the continuity from above of  $\Phi$  and introduce some new notations.

Let  $P$  be a probability measure defined on  $\Omega$ . Since the citation record of an author  $X$  is a bounded function, it seems natural to take  $X \in L_\infty$ , where  $L_\infty$  is the space of functions that are  $P$  almost surely bounded. We also denote with  $\Delta := \{Q \ll P\}$  the set of all probabilities that are absolutely continuous with respect to  $P$ .

**Theorem 1** *Any map  $\Phi : L_\infty \rightarrow \mathbb{R}$  that is quasi-concave, monotone increasing and continuous from above can be represented as*

$$\Phi(X) := \inf_{Q \in \Delta} \beta(Q, X), \quad \forall X \in L_\infty \quad (5)$$

where  $\beta : \Delta \times L_\infty \rightarrow \mathbb{R}$  and  $\gamma : \Delta \times \mathbb{R} \rightarrow \mathbb{R}$  are defined as

$$\beta(Q, X) := \sup \{m \in \mathbb{R} \mid E_Q[X] \geq \gamma(Q, m)\}$$

$$\gamma(Q, m) := \inf_{Y \in L_1} \{E_Q[Y] \mid \Phi(Y) \geq m\}. \quad (6)$$

The proof is provided in Appendix B.<sup>17</sup>

There are two technical differences between the  $\phi$ -index in (1) and the  $\Phi$ -index in (4). Remember that a law invariant function associates the same value to random variables with the same distribution law. The first difference is that the  $\phi$ -index is law invariant while the  $\Phi$ -index may not be law invariant. In other words, if two scientists have the same citation curve, they have the same  $\phi$ -index. If they publish in different journals, they generally have different  $\Phi$ -indices. Thus, the  $\Phi$ -index can rank scientists with the same citation curves, but who publish in different journals.<sup>18</sup> The second difference is that the  $\phi$ -index is not necessarily quasi-concave on the citation records (even if it is always quasi-concave on the citation curves). When the  $\phi$ -index is quasi-concave on the citation records, the  $\Phi$ -index includes the  $\phi$ -index as a special case. Finally, note that if we endow the space  $L_\infty$  with the weak topology  $\sigma(L_\infty, L_1)$ , its dual space is  $L_1$ .

## Empirical Analysis

This section provides an empirical application of our SRM, discusses its implementation, and compares the empirical findings to those of seven existing bibliometric indices.<sup>19</sup>

## Data

We consider 173 full professors affiliated to 11 major US business schools or finance departments. The set of researchers is quite large and homogenous in terms of seniority and research field, making comparison significant. A Python script is run to extract their citation curves from Google Scholar (<http://scholar.google.com/>). Our data set is created on September 6, 2012. While entries in Google Scholar may suffer from data errors like any other database, the American Scientist Open Access Forum (2008) remarked that “Google Scholar’s accuracy is growing daily, with growing content.”<sup>20</sup> Citation curves in Google Scholar are publicly and freely available. Our SRMs are not tailored to Google Scholar and can be applied to any other data set.

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 eight members at Berkeley and Cornell to 34 members at Columbia. Table 3 also reports summary statistics for the number of citations and publications, aggregated at 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 positively skewed. Princeton has the largest number of publications, and fifth largest faculty size.

Insert Table 3 here

## Estimates of the Scientific Research Measure

To accurately rank research performance, performance curves should mimic actual citation curves. As all citation curves in our data set display power law shapes, we fit the power law function  $f_q(x) = q/x^c - k$  to each author’s citation curve.<sup>21</sup> All estimated parameters are highly significant and adjusted  $R^2$ ’s are above 95%.<sup>22</sup> As an example, Figure 3 shows four citation curves and fitted power law functions.

Insert Figure 3 here

We use the following  $\phi$ -index to rank authors:

$$\phi(X) := \sup \{q \in \mathbb{R} / X(x) \geq q/x^c - k, \forall x \in [1, \infty)\} \quad (7)$$

where  $c = 0.59$  and  $k = 543.2$ , which are the weighted averages of all estimated  $c$  and  $k$  coefficients, respectively, with weights given by the number of data points used to estimate each coefficient.<sup>23</sup>

The following two-step algorithm can be used to compute the  $\phi$ -index in (7). For each author:

1. Compute  $q_x := (X(x) + k) x^c$ , for  $x = 1, 2, \dots, p$ , and  $q_* := k p^c$
2. Take  $\phi(X) := \min(q_1, q_2, \dots, q_p, q_*)$

where  $p$  is the number of publications with at least one citation. The performance curve associated to  $q_x$  passes through the point  $(x, X(x))$ . This holds true for each  $x = 1, 2, \dots, p$ .<sup>24</sup> The performance curve associated to  $q_*$  crosses the horizontal axis at  $(p, 0)$ , i.e.,  $f_{q_*}(p) = 0$ .

This performance curve is considered because the whole citation curve may lie above the performance curve  $f_{q^*}$  and in that case  $f_{q^*}$  determines the  $\phi$ -index in (7).<sup>25</sup>

To better understand the two-step algorithm, consider again Figure 1. This figure shows the  $p = 8$  performance curves associated to  $q_1, \dots, q_p$ . The performance curve associated to  $q^*$  is not shown but would pass through the point (8, 0). Taking the minimum of all the  $q_x$ 's and  $q^*$  (which is  $q_3 = 24.75$ ) 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 below the citation curve because it touches the citation curve at one point (which is (3, 5) in Figure 1).

### *Empirical Results*

For each of the 173 authors in our data set, we compute the  $\phi$ -index in (7) and the other seven bibliometric indices listed in Table 1. To provide an overview of the results, we aggregate estimates of the  $\phi$ -index and bibliometric indices at the university level. Table 4 shows the mean and median for each index and university. Averages are often significantly larger than medians (e.g., for Princeton and Stanford), indicating that the distribution of the indices is positively skewed. Aggregate results should thus be interpreted cautiously.

The main message from Table 4 is that rankings of universities based on our  $\phi$ -index and the other bibliometric indices are generally different. Under the premise that research performance is reflected by the whole citation curve rather than only part of it, our  $\phi$ -index provides a more accurate ranking of research performances than the other bibliometric indices. Differences in rankings are more concentrated on “average” performing universities, rather than best and least well performing universities, in relative terms. For example, according to the average  $\phi$ -index, Chicago ranks first and Princeton second. According to the average  $h$ -index, Princeton ranks first and Chicago second. Berkeley ranks tenth according to both averages. For the remaining universities, the rankings provided by the  $\phi$ - and  $h$ -indices, as well as the other indices, are quite different.

Insert Table 4 here

Empirical results at the level of individual professors parallel the empirical results at university level. Owing to space constraints, we cannot report the estimates of all indices for all 173 full professors. To give some indication of the empirical findings, Figure 4 shows the scatter plots of the  $\phi$ -index versus the  $h$ -index and versus the  $g$ -index for each scholar in our sample.<sup>26</sup> The relationship between the indices is clearly positive. Best performing scholars tend to have the largest  $\phi$ -,  $h$ -, and  $g$ -indices. Scholars in the top right-hand corner of the scatter plots are typically Nobel laureates in economics. This obviously results in a positive correlation between the indices. However, the relationship is far from monotone, meaning that different indices lead to different rankings.

A closer look at the scatter plots reveals that various scholars have the same (or almost the same)  $h$ -index, while  $\phi$ -indices are always different. Thus, the  $h$ -index cannot be used to rank them, while the  $\phi$ -index provides a more granular ranking. This situation tends to occur when two authors have roughly the same number of publications but one author has more citations per publication, especially in the Hirsch core, and therefore a higher  $\phi$ -index. In our data set, nine authors have an  $h$ -index of 21, nine authors of 25, and eight authors of 28. Thus, using

the  $h$ -index, 26 authors, i.e., 15% of all authors in our data set, would fall into three broad ranks, effectively not ranking them. All authors have a different  $\phi$ -index.

To summarize, this empirical application shows that our  $\phi$ -index can be easily implemented and leads to different rankings from those of other bibliometric indices.

Insert Figure 4 here

## **Conclusion**

The funding of research projects, faculty recruitment, and other key decisions for universities and research institutions depend to a large extent upon the scientific merits of the researchers involved. To rank scientists' research performances, this paper introduces the novel class of Scientific Research Measures. The main feature of SRMs is that they take into account scientists' whole citation curves as well as specific features of the scientific field, research areas, and seniority of scientists. This is achieved using flexible performance curves, whereas most existing bibliometric indices lack such flexibility.

We also introduce the further general class of Dual SRMs to obtain more informed research rankings. Dual SRMs permit the aggregation of scientists' research outputs and therefore theoretically result in well-founded rankings of research teams, departments and universities, something that has constituted a central theme in the scientific community over the last decades.

Finally, we provide an empirical application using a novel data set of 173 finance scholars' citation curves. Empirical results show that power law performance curves fit actual citation curves well. We find that authors' rankings based on the SRM characterized by power law performance curves, and those produced by seven other bibliometric indices, are quite different. Under the premise that research performance is reflected by the whole citation curve, rather than only part of it, the SRM provides a more judicious ranking of research performances.

## Notes

<sup>1</sup> For a review of these proposals, see e.g., Alonso, Cabrerizo, Herrera-Viedma and Herrera (2009), Panaretos and Malesios (2009), and Schreiber (2010).

<sup>2</sup> The citation curve is the number of citations of each publication, in decreasing order of citations.

<sup>3</sup> 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. Rather than focusing on a specific measurement of risk due to financial positions (e.g., variance of asset returns), Artzner et al. (1999) proposed a class of measures satisfying coherent axioms. Ideally, each financial institution could select its own risk measure, provided it obeyed the coherent axioms. This approach made the selection of the risk measure more flexible and, at the same time, established a unified framework. In the same spirit, we introduce Dual SRMs to rank research performance, rather than financial risks. Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002) extended the theory of (coherent) risk measures to the class of convex risk measures.

<sup>4</sup> Other studies in this area include Cherny and Madan (2009) who introduce the concept of an acceptability index having the property of quasi-concavity, and Drapeau and Kupper (2013) who analyze the relationship between quasi-convex risk measures and the associated family of acceptance sets, already present in Cherny and Madan (2009).

<sup>5</sup> The Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) is currently undertaking a large-scale evaluation process requiring 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 approximately 15,000 research items to evaluate in less than one year. All panels, except those in humanistic areas, decided to use bibliometric indices.

<sup>6</sup> See <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>.

<sup>7</sup> Like any 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 reduce 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.

<sup>8</sup> As the number of publications and citations are positive integers, the domain and range for  $X$  appear to be the positive integers. However, in the research evaluation process, the domain and range of  $X$  need not be the positive integers. For example, if a publication is co-authored by two or more scientists, the evaluator may choose to count that publication as half or  $1/(\text{number of co-authors})$ , when ranking the research performance of one of the co-authors. The evaluator may also choose to attribute more weight to citations of research papers, rather than review papers or books, making the range of  $X$  positive reals, rather than positive integers.

<sup>9</sup> 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  $\alpha$  can be used, for example, to normalize to 100 the  $\phi$ -index of the scientist with the largest number of publications or citations, or any other scientist. This normalization is achieved by replacing  $q$  by  $\alpha q$  in the two-step algorithm developed in the section “*Estimates of the Scientific Research Measure*” and determining  $\alpha$  using the citation curve of the scientist with the largest number of publications or citations.

<sup>10</sup> We thank Michael Schreiber for providing this comment as well as other insightful comments throughout the paper.

<sup>11</sup> To implement Dual SRMs a richer data set is needed including citations, in which journals publications appeared, and quality of the journals. Retrieving such data can be challenging but necessary if this information is to be reflected in research rankings.

<sup>12</sup> Other criteria are obviously conceivable to assess journal quality, such as overall citations of the journal or acceptance rate of journal submissions. Such additional criteria can be easily taken in account in the Dual SRM. For illustration purposes, we consider only two impact factors for each journal in this example.

<sup>13</sup> The impact factors were obtained from the website of each journal in 2012. The impact factor was devised in the mid-1950s by Eugene Garfield. See Garfield (2005) for a recent account of the impact factor.

<sup>14</sup> The first editions of MS, MF and FS appeared in 1954, 1991 and 1996, respectively, and in 2012 this implied  $Q_0(\text{MS}) = (2012 - 1954)/95 = 0.61$ ,  $Q_0(\text{MF}) = (2012 - 1991)/95 = 0.22$  and  $Q_0(\text{FS}) = (2012 - 1996)/95 = 0.17$ .

<sup>15</sup> Over the years any scientist may have published more than one paper in the same journal. A simple approach to take this case into account would be to consider the total number of citations of all publications in a 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 statistic when assessing the scientist’s research performance as  $EQ_i[X]$  is a linear operator. To simplify the discussion of Dual SRMs we implicitly adopted this approach in the current section. If however 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.

<sup>16</sup> 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  $\sum_i \lambda_i X_i$ , where  $\sum_i \lambda_i = 1$  and  $\lambda_i \geq 0$  for all  $i$ .

<sup>17</sup> In (5) the infimum is taken with respect to all probabilities in  $\Delta$ , and not in a prescribed subset  $P \subseteq \Delta$ . This must be the case, since in the theorem the map  $\Phi$  is given a priori and only the objects that can be determined from  $\Phi$  can appear in the representation (5).

<sup>18</sup> Frittelli, Maggis and Peri (2014) provide a theoretical analysis of quasi-convex maps defined on distributions.

<sup>19</sup> As discussed in the previous section, an empirical application of Dual SRMs requires a richer data set, including in which journal each author's publications appeared and the values of those journals. We defer such an application to future work.

<sup>20</sup> <http://openaccess.eprints.org/index.php?archives/417-Citation-Statistics-International-Mathematical-Union-Report.html>.

<sup>21</sup> To estimate the power law function we use nonlinear least squares and minimize  $\sum_x (X(x) - f_q(x))^2$  with respect to the parameters  $q$ ,  $c$  and  $k$ , where  $p$  is the author's number of publications with at least one citation. Starting values for the nonlinear minimization are obtained as follows. The function  $f_{\sim q}(x) = q/x^c$  is a restricted version of  $f_q(x)$ , imposing  $k = 0$ . As in the log-log space  $f_{\sim q}(x)$  is linear in  $\log(x)$ , i.e.,  $\log(f_{\sim q}(x)) = \log(q) - c \log(x)$ , the two parameter estimates  $\log^{\wedge}(q)$  and  $c^{\wedge}$  can be easily obtained by the least square regression of  $\log(X(x))$  on a constant and  $\log(x)$ , where  $x = 1, 2, \dots, p$ . Starting values for  $q$ ,  $c$  and  $k$  are  $\exp(\log^{\wedge}(q))$ ,  $c^{\wedge}$  and 0, respectively.

<sup>22</sup> Estimation results are available from the authors upon request.

<sup>23</sup> We also used other values of  $c$  and  $k$ , such as the simple average of all estimated  $c$  and  $k$  coefficients. The empirical results remained largely unchanged and are available from the authors upon request.

<sup>24</sup> 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  $f_{q_p}(p) = q_p/p^c - k = (X(p) + k) p^c/p^c - k = X(p)$ , confirming that this performance curve passes through the point  $(p, X(p))$ .

<sup>25</sup> This situation may occur when the citation curve is roughly linear and steep. Although this is unlikely, the performance curve associated to  $q_*$  needs to be considered when computing the  $\phi$ -index in (7). More specifically, as  $p$  is the number of publications with at least one citation, the citation curve at  $x = p$  is such that  $X(p) \geq 1 > 0 = f_{q_*}(p)$ . Thus the citation curve  $X(x)$  may lie above  $f_{q_*}(x)$  for all  $x$ . Given the shape of the performance curves in Figure 1, this situation can occur when the citation curve is steep and roughly linear. 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  $q > q_*$ . All these performance curves lie above  $f_{q_*}$ , and thus above  $X(x) = 0$  at least for all  $x > p$ , and do not determine the  $\phi$ -index.

<sup>26</sup> Scatter plots of the  $\phi$ -index versus the other bibliometric indices are not reported, but share similar patterns.

## References

- Adler, R., Ewing, J., & Taylor, P. (2009). Citation statistics. *Statistical Science*, 24, 1–14.
- Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., & Herrera, F. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3, 273–289.
- Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., & Herrera, F. (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.R., Hankin, R.K.S., & Killworth, P.D. (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics*, 76, 577–588.
- Artzner, P., Delbaen, F., Eber, J.M., & Heath, D. (1999). Coherent measure of risk. *Mathematical Finance*, 9, 203–228.
- Bornmann, L., & Daniel, H.D. (2007). What do we know about the h-index?. *Journal of the American Society for Information Science and Technology*, 58, 1381–1385.
- Bornmann, L., Mutz, R., & Daniel, H.D. (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.
- Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2014). Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers: A visualization of results from multi-level models. *Online Information Review*, 38, 43–58.
- Borokhovich, K.A., Bricker, R.J., Brunarski, K.R., & Simkins, B.J. (1995). Finance research productivity and influence. *Journal of Finance*, 50, 1691–1717.
- Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., & Herrera, F. (2010). q2-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., Maccheroni, F., Marinacci, M., & Montrucchio, L. (2011). Risk measures: Rationality and diversification. *Mathematical Finance*, 21, 743–774.
- Cherny, A., & Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, 22, 2571–2606.
- Chung, K.H., & Cox, R.A.K. (1990). Patterns of productivity in the finance literature: A study of the bibliometric distributions. *Journal of Finance*, 45, 301–309.

- Deineko, V.G., & Woeginger, G.J. (2009). A new family of scientific impact measures: The generalized Kosmulski-indices. *Scientometrics*, 80, 819–826.
- Drapeau, S., & Kupper, M. (2013). Risk preferences and their robust representation. *Mathematics of Operations Research*, 38, 28–62.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69, 131–152.
- Egghe, L., & Rousseau, R. (2008). An h-index weighted by citation impact. *Information Processing and Management*, 44, 770–780.
- Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6, 429–447.
- Frittelli, M., & Maggis, M. (2011). Dual representation of quasiconvex conditional maps. *SIAM Journal on Financial Mathematics*, 2, 357–382.
- Frittelli, M., Maggis, M., & Peri, I. (2014). Risk measures on  $P(R)$  and value at risk with probability/loss function. *Mathematical Finance*, 24, 442–463.
- Frittelli, M., & Rosazza Gianin, E. (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., & Rousseau, R. (2009). Real and rational variants of the h-index and the g-index. *Journal of Informetrics*, 3, 64–71.
- Hauser, J.R. (1998). Research, development, and engineering metrics. *Management Science*, 44, 1670–1689.
- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569–16572.
- Jin, B. (2006). h-index: An evaluation indicator proposed by scientist. *Science Focus*, 1, 8–9.
- Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52, 855–863.
- Kalaitzidakis, P., Mamuneas, T.P., & Stengos, T. (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.G., & Johnson, S.D. (2008). Assessment of research performance in biology: How well do peer review and bibliometry correlate?. *BioScience*, 58, 160–164.
- May, R. (1997). The scientific wealth of nations. *Science*, 275, 793–796.

- Marchant, T. (2009). Score-based bibliometric rankings of authors. *Journal of the American Society for Information Science and Technology*, 60, 1132–1137.
- Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, 81, 635–670.
- Ravallion, M., & Wagstaff, A. (2011). On measuring scholarly influence by citations. *Scientometrics*, 88, 321–337.
- Rousseau, R. (2006). Simple models and the corresponding h- and g-index. preprint.
- Rousseau, R. (2014). A note on the interpolated or real-valued h-index with a generalization for fractional counting. *Aslib Journal of Information Management*, 66, 2–12.
- Ruane, F., & Tol, R.S.J. (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.J., & Waltman, L. (2008). Generalizing the h- and g-indices. *Journal of Informetrics*, 2, 263–271.
- Vinkler, P. (2009). The  $\pi$ -index: A new indicator for assessing scientific impact. *Journal of Information Science*, 35, 602–612.
- Wiberley, S.E. (2003). A methodological approach to developing bibliometric models of types of humanities scholarship. *Library Quarterly*, 73, 121–159.
- Woeginger, G.J. (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.

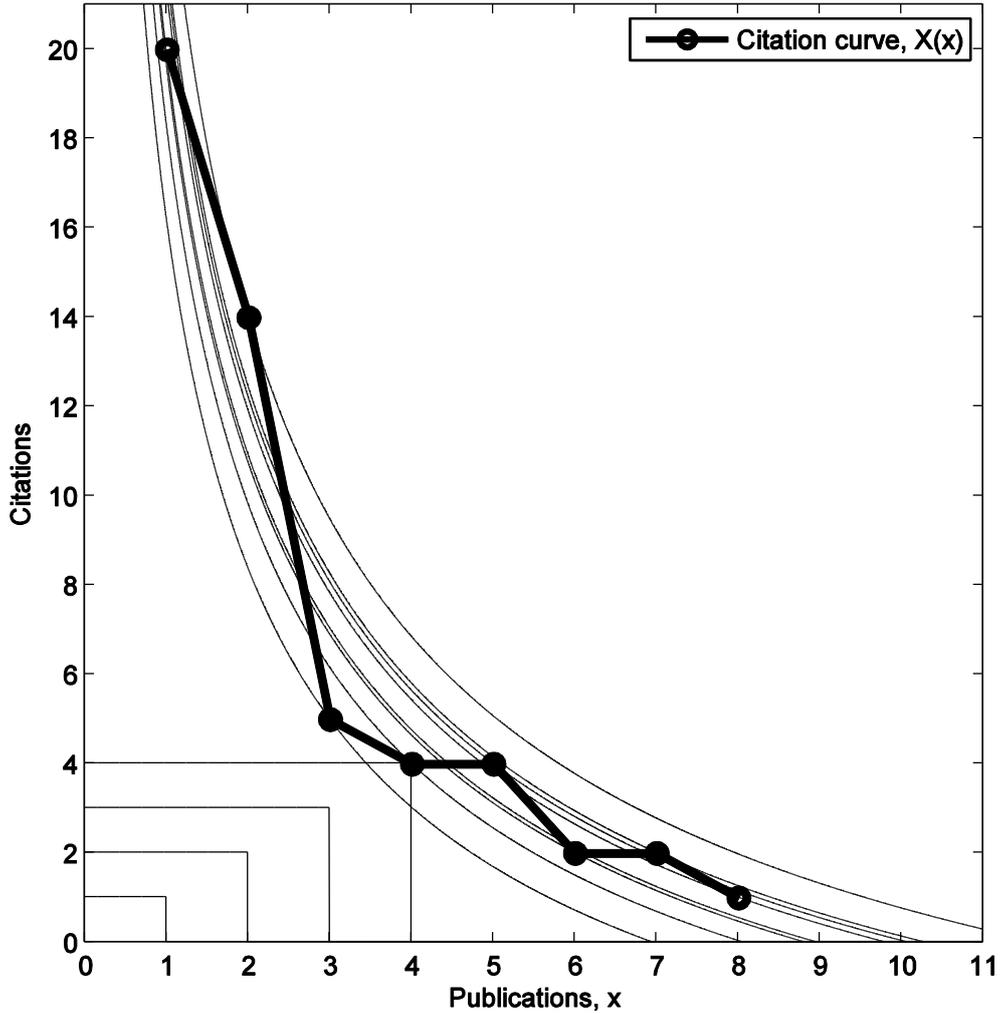


Figure 1: **Citation curve.** The graph shows the citation curve of a hypothetical scientist with eight publications, as well as performance curves based on the  $h$ -index and  $\phi$ -index. The citation curve is obtained by sorting publications in decreasing order of citations. Performance curves of the  $h$ -index are given by  $f_q(x) = q \cdot 1_{(0,q]}(x)$ , with  $q = 1, 2, 3, 4$ . Performance curves of the  $\phi$ -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  $x = 1, 2, \dots, 8$ . The hypothetical scientist has an  $h$ -index of four and  $\phi$ -index of 24.75. The  $\phi$ -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  $24.75/3^{0.55} - 8.49 = 5$ .

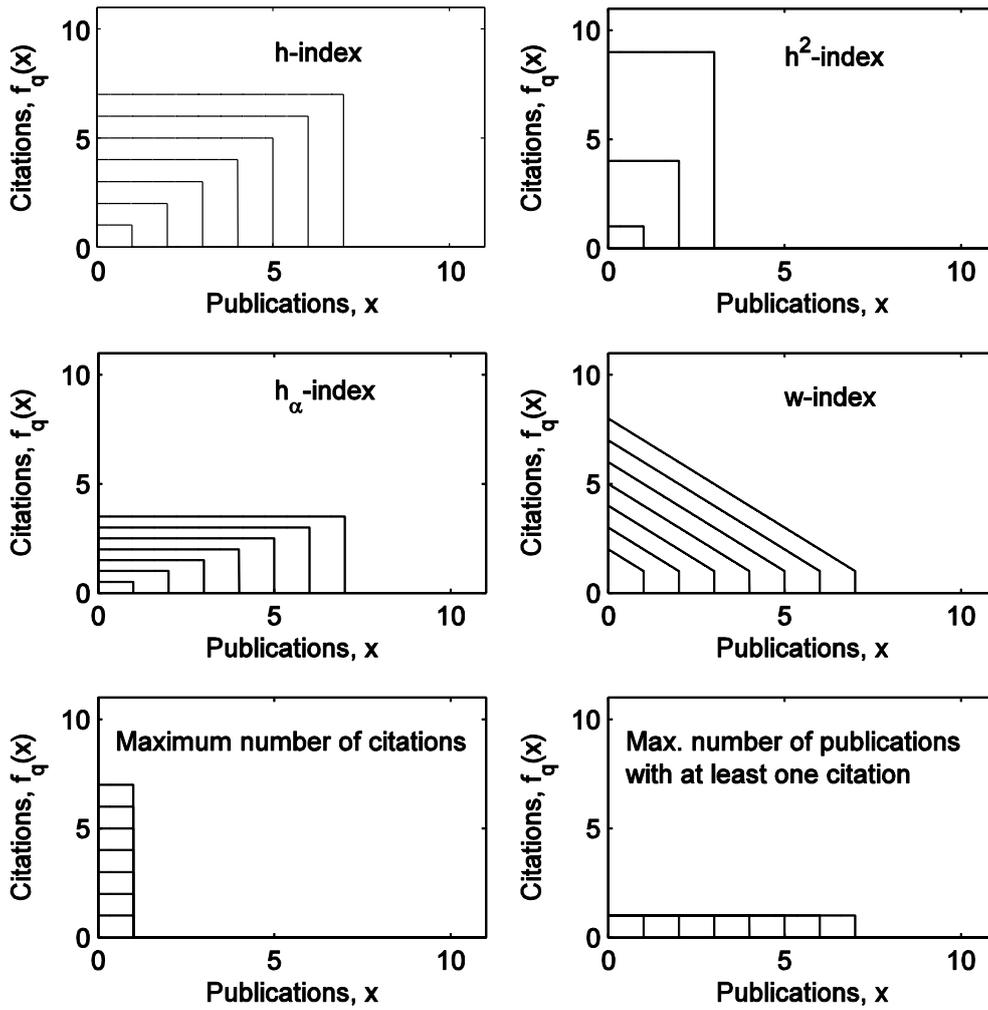


Figure 2: **Performance curves of existing bibliometric indices.** The graph shows six families of performance curves corresponding to six different bibliometric indices:  $h$ -index with performance curve  $f_q(x) = q1_{(0,q]}(x)$  and  $q = 1, 2, \dots, 7$ ;  $h^2$ -index with performance curve  $f_q(x)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

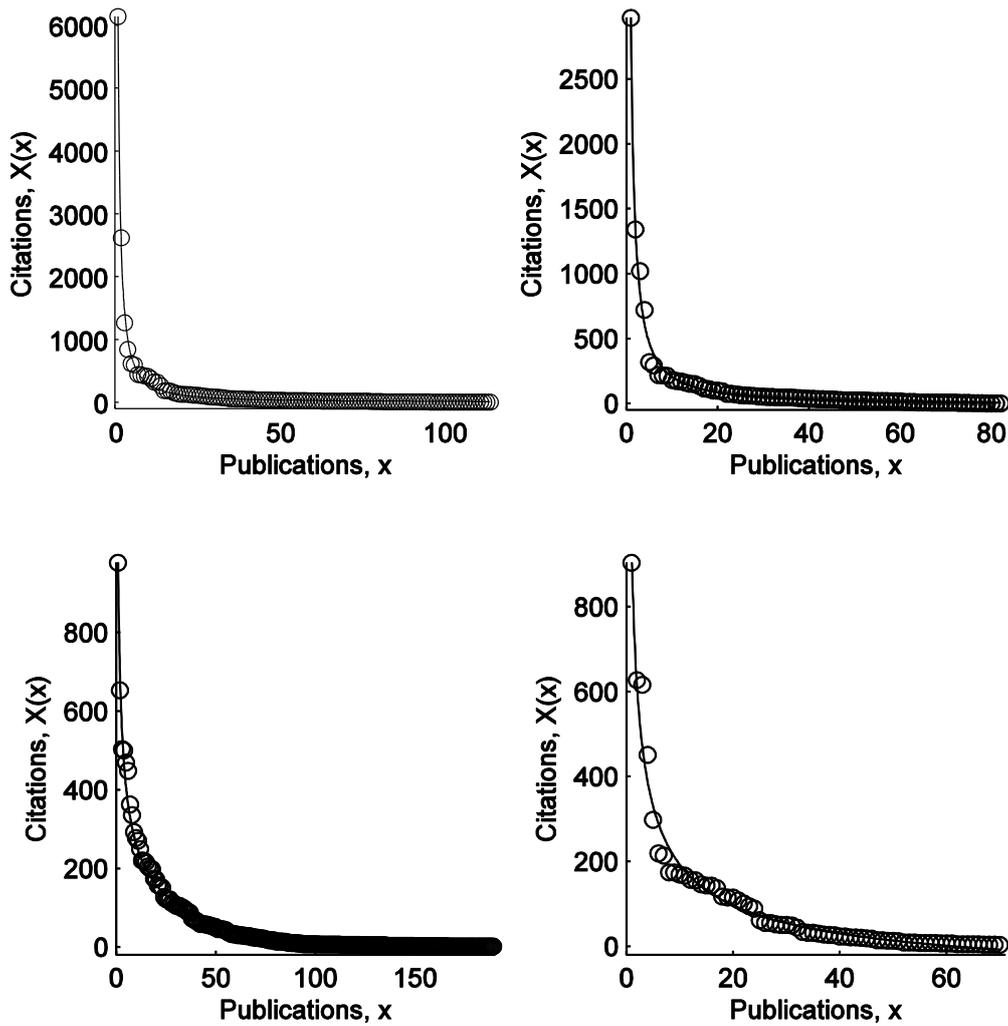


Figure 3: **Typical fit of citation curves.** Each graph shows one citation curve  $X(x)$  (circles), i.e., the author's citations in decreasing order of citations, where  $x = 1, \dots, p$  and  $p$  is the number of publications with at least one citation. Superimposed is the fitted power law curve  $f_q(x) = q/x^c - k$ . For each author, parameter estimates of the power law curve are obtained using nonlinear least squares, i.e., by minimizing  $\sum_x (X(x) - f_q(x))^2$  with respect to  $q, c$  and  $k$ . All adjusted  $R^2$ 's are above 97.4%.

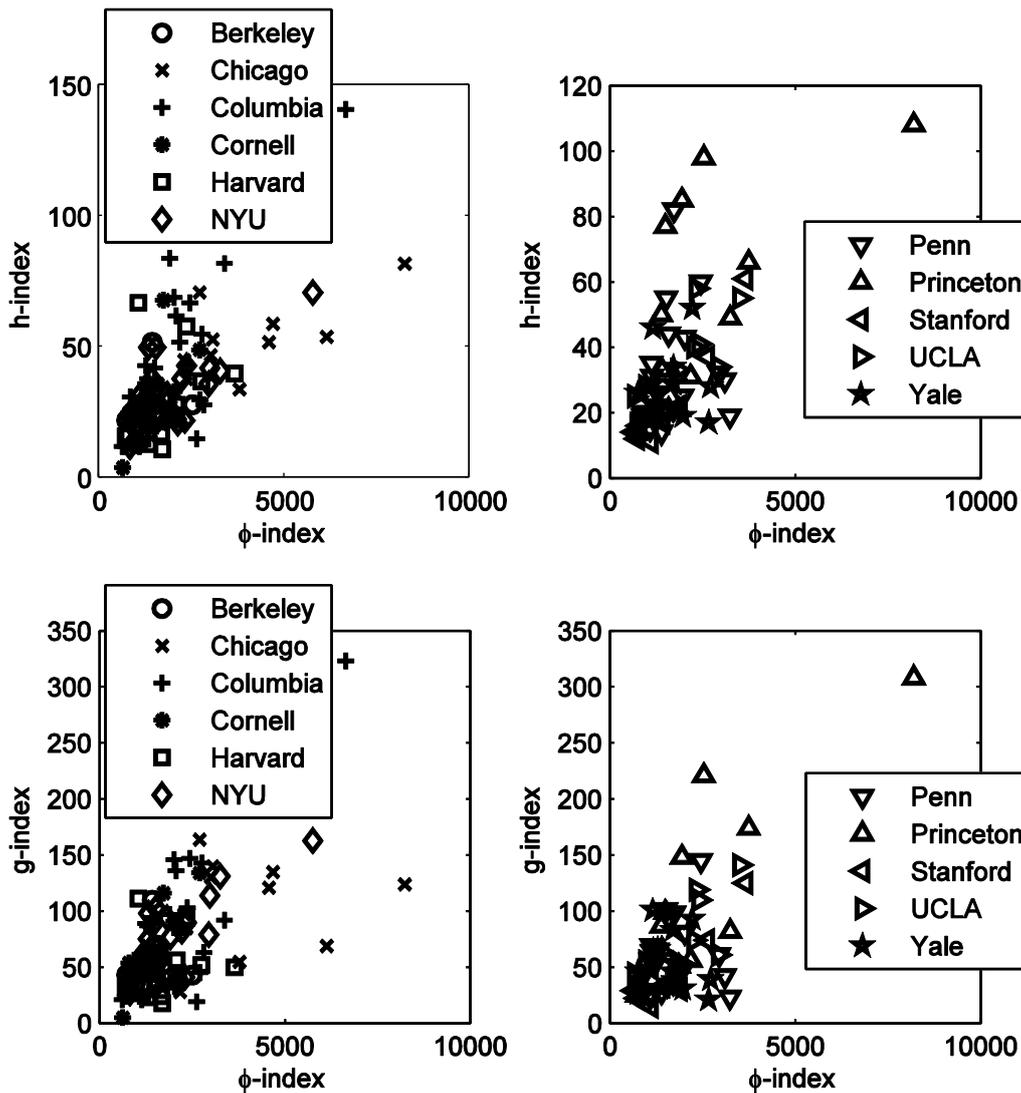


Figure 4: **Scatter plots of  $\phi$ -index versus  $h$ - and  $g$ -indices.** Upper graphs: scatter plot of  $\phi$ -index (horizontal axis) versus  $h$ -index (vertical axis). Lower graphs: scatter plot of  $\phi$ -index (horizontal axis) versus  $g$ -index (vertical axis). These indexes are computed for each one of the 173 scholars in our data set and displayed according to the university to which the scholar belongs. The relationships between  $\phi$ -index and  $h$ -index (respectively  $g$ -index) are not monotone. Thus rankings of scholars given by the  $\phi$ -index and  $h$ -index (respectively  $g$ -index) are different. The data set of citations covers 173 full professors affiliated to business schools or finance departments of 11 main US universities. The data set was extracted from Google Scholar on September 6, 2012.

Table 1

Index	Description	Author(s)	$f_q(x)$
$h$	A scientist has index $h$ if $h$ of their papers have at least $h$ citations	Hirsch (2005)	$q1_{(0,q]}(x)$
$h^2$	A scientist has index $h^2$ if $h$ of their papers have at least $h^2$ citations	Kosmulski (2006)	$q^2 1_{(0,q]}(x)$
$h_\alpha$	A scientist has index $h_\alpha$ if $h$ of their papers have at least $\alpha h$ citations	van Eck and Waltman (2008)	$\alpha q 1_{(0,q]}(x), \alpha > 0$
$w$	A scientist has index $w$ if $w$ of their papers have at least $w, w-1, \dots, 1$ citations	Woeginger (2008)	$(-x + q + 1)1_{(0,q]}(x)$
$A$	A scientist has index $A = \sum_x X(x)/h$ , where $h$ is the $h$ -index and $X(x)$ the citation curve	Jin (2006)	not SRM
$R$	A scientist has index $R = \sqrt{\sum_x X(x)}$ , 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 have at least $g^2$ citations	Egghe (2006)	not SRM

Table 1: **Bibliometric indices.** This table lists popular bibliometric indices, a short description of each, and the author(s) who introduced the index. The  $h$ -,  $h^2$ -,  $h_\alpha$ -, and  $w$ -indices 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  $g$ -indices are not SRMs.

Table 2

	Evaluator 1		Evaluator 2		Evaluator 3		Evaluator 4	
	$\alpha(Q)$	$\beta(Q, X)$						
$Q_0$	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  $\Omega$  consists of three journals,  $\Omega = \{MS, MF, FS\}$ . The scientist has three publications, one in each journal, with 10, six, and eight citations, respectively, i.e., their citation record  $X(\omega)$  is given by  $X(MS) = 10$ ,  $X(MF) = 6$ , and  $X(FS) = 8$ . To assess their research performance, each evaluator uses three criteria expressed by the probability measures  $Q_0$ ,  $Q_2$ , and  $Q_5$ . Criterion  $Q_0$  expresses the number of years since the journal's first edition (the journal's first edition for MS, MF, and FS is 1954, 1991, and 1996, respectively). Criterion  $Q_2$  expresses the two-year impact factor of the journal (the two-year impact factor for MS, MF, and FS is 1.73, 1.25, and 1.15, respectively). Criterion  $Q_5$  expresses the five-year impact factor of the journal (the five-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  $Q$ ,  $\beta(Q, X) := \sup\{m \in \mathbb{R} \mid E_Q[X] \geq \gamma(Q, m)\}$ , where  $\gamma(Q, m) := m \alpha(Q)$ . The Dual SRM gives the final research performance as  $\Phi(X) = \min(\beta(Q_0, X), \beta(Q_2, X), \beta(Q_5, X))$ .

Table 3

Full Name		Faculty	Citations			Publications		
			Avg	Med	Std	Avg	Med	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: **Data set.** The data set includes 173 full professors affiliated to business schools or finance departments of 11 main US universities. Faculty is the number of full professors for each university in our data set. For each author, the data set 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 data set of citations and publications is from Google Scholar and was extracted on September 6, 2012.

Table 4 – Part 1

	$\phi$		$h$		$h^2$		$h_a$	
	Avg	Med	Avg	Med	Avg	Med	Avg	Med
Berkeley	1,504	1,489	27	25	10	10	35	32
Chicago	3,007	2,338	41	38	16	15	48	42
Columbia	1,778	1,415	38	30	12	11	46	36
Cornell	1,495	1,446	32	29	10	11	43	36
Harvard	1,705	1,626	28	23	11	10	33	26
NYU	1,768	1,428	33	31	12	12	40	38
Penn.	1,653	1,516	31	25	11	10	39	32
Princeton	2,404	1,853	51	49	14	13	67	62
Stanford	1,536	1,109	23	19	11	9	28	20
UCLA	1,962	2,267	37	34	13	15	45	40
Yale	1,726	1,706	30	29	12	11	36	33

Table 4 – Part 2

	$w$		$A$		$R$		$g$	
	Avg	Med	Avg	Med	Avg	Med	Avg	Med
Berkeley	47	41	185	160	69	65	56	44
Chicago	66	60	479	331	135	117	83	85
Columbia	63	50	229	186	88	79	73	57
Cornell	57	51	171	146	73	71	69	56
Harvard	45	34	247	239	79	68	48	44
NYU	57	52	238	186	85	81	71	67
Penn.	52	42	235	174	79	64	58	52
Princeton	91	82	296	278	119	94	106	82
Stanford	37	28	219	218	70	61	43	31
UCLA	63	53	286	320	100	113	76	61
Yale	49	44	274	244	85	84	59	54

Table 4: **The Scientific Research Measure and bibliometric indices.** For each scholar in our data set,  $\phi$ -,  $h$ -,  $h^2$ -,  $h_a$  (with  $\alpha = 0.5$ ),  $w$ -,  $A$ -,  $R$ -, and  $g$ -index are computed. The  $\phi$ -index is based on power law performance curves. 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 data set of citations covers 173 full professors affiliated to business schools or finance departments of 11 main US universities. The data set was extracted from Google Scholar on September 6, 2012.