Rapporto n. 224

From Risk Measures to Research Measures

Marco Fritelli, Ilaria Peri

Marzo 2012

Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali
Università degli Studi di Milano Bicocca
Via Bicocca degli Arcimboldi 8, - 20126 Milano - Italia
Tel +39/02/6448.3103 o 3139 Fax +39/2/64483105
Segreteria di redazione: Andrea Bertolini

From Risk Measures to Research Measures

Marco Frittelli* Ilaria Peri†

March 12, 2012

Abstract

In order to evaluate the quality of the scientific research, we introduce a new family of scientific performance measures, called Scientific Research Measures (SRM). Our proposal originates from the more recent developments in the theory of risk measures and is an attempt to resolve the many problems of the existing bibliometric indices.

The SRM that we introduce are: *flexible* to fit peculiarities of different areas and seniorities; *inclusive*, as they comprehend several popular indices; *coherent*, as they share the same structural properties; *calibrated* to the particular scientific community; *granular,* as they allow a more precise comparison between scientists and are based on the whole scientist's citation curve.

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

1 Introduction

In the recent years the evaluation of the scientist's performance has become increasingly important. In fact, most crucial decisions regarding faculty recruitment, accepting research projects, research time, academic positions, travel money, award of grants and promotions depend on great extent upon the scientific merits of the involved researchers.

The scope of the valuation of the scientific research is mainly twofold:

- Provide an updated picture of the existing research activity, in order to allocate financial resources in relation to the scientific quality and scientific production;
- Determine an increase in the quality of the scientific research (of the structures).

*Università degli Studi di Milano.
†Università degli Studi di Milano Bicocca.

The methodologies for the valuation can be divided into two categories:

- content valuation, based on:
  - internal judgments committee;
  - external reviews of peer panels.
- context valuation, based on:
  - bibliometrics (i.e. statistics derived from citation data);
  - characteristics of the Journals associated to the publications.

Economic considerations strongly depone of using the context method on a systematic (yearly) base, while peer review is more plausible on a multiple year base and should also be finalized to check, harmonize, and tune the outcomes based on bibliometric indices.

So in order to have a simple and cheaper assessment method and thank to the major availability of the online database (i.e. Google Scholar, ISI Web of Science, MathSciNet and Scopus) several different bibliometric measures have been introduced.

There are several critics, as those clearly underlined by the Citation Statistics Report of the International Mathematical Union (2008) [CIT], to the use of the citations as a key factor in the assessment of the quality of the research. However, many of these critics can be satisfactorily addressed and our proposal is one reasonable way to achieve this task. We emphasize that the output of the valuation is the classification of authors (and structures) into few classes of homogeneous research quality: it is not intended to provide a fine ranking.

Many indices were developed to quantify the production of researchers, e.g. the total number of published papers in a period of time; or the impact of their publications, e.g. the total number of citations, the average number of citations per paper, the number and percentage of significant papers (with more than a certain amount of citations).

In 2005 Hirsch [H05] proposed the *h-index,* that is now the most popular and used citation-based metric. A scientist has index h if h of his n papers have at least h citations each and the other $(n-h)$ papers have at most h citations each. The h-index is an attempt measures at the same time the productivity in terms of number of publications and the research quality in terms of citations per publication.

After its introduction, the h-index received wide attention from the scientific community and it has been extended by many authors who have proposed other indices (for an overview see Alonso et al., 2009 [ACHH]) in order to overcome some of the drawbacks of it (see Bornmann and Daniel, 2007 [BD07]).

Differently from any existing approach, our formulation is clearly germinated from the Theory of Risk Measures. The axiomatic approach developed in the seminal paper by Artzner et al.[ADEH99] turned out to be, in this last decade, very influential for the theory of risk measures: instead of focusing on some particular measurement of the risk carried by financial positions (the variance, the $V@R$, etc. etc.), [ADEH99] proposed a class of measures satisfying some reasonable properties (the "coherent" axioms). Ideally, each 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. We propose the same approach in order to determine a good class of scientific performance measures, that we call Scientific Research Measures (SRM).

The theory of coherent risk measures was later extended to the class of convex risk measures (Follmer and Schied [FS02], Frittelli and Rosazza [FR02]). The origin of our proposal can be traced in the more recent development of this theory, leading to the notion of quasi-convex risk measures introduced by Cerreia-Vioglio et al. [CMMM] and further developed in the dynamic frame work by Frittelli and Maggis [FM11]. Additional papers in this area include: Cherny and Madan [CM09], that introduced the concept of an Acceptability Index having the property of quasi-concavity; Drapeau and Kupper [DK10], where the correspondence between a quasi-convex risk measure and the associated family of acceptance sets - already present in [CM09] - is fully analyzed. The representation of quasi-concave monotone maps in terms of family of acceptance sets, as well as their dual formulations, are the key ideas underlying our definition of SRM.

We propose a family of SRMs that are:

- *flexible* in order to fit peculiarities of different areas and ages;
- *inclusive*, as they comprehends several popular indices;
- *calibrated* to the particular scientific community;
- *coherent*, as they share the same structural properties - based on an axiomatic approach;
- *granular,* as they allow a more precise comparison between scientists and are based on *the whole citation curve* of a scientist.

The definition of the SRM, the relative properties and some examples are given in Section 2. A new interesting approach to the whole area of bibliometric indices is provided by the dual representation of a SRM discussed in Section 3. We also show the method to compute a particular SRM, called ϕ-index, and we report some empirical results obtained by *calibrating* the performance curves to a specific data set (built using Google Scholar).

2 On a class of Scientific Research Measure

We represent each author by a vector X of citations, where the i-th component of X represents the number of citations of the i-th publication and the components of X are ranked in decreasing order. We consider the whole *citation curve* of an author as a decreasing bounded step functions X (see Fig.1) in the convex cone:

$$\mathcal{X}^+ = \left\{ \begin{array}{c} X : \mathbb{R} \to \mathbb{R}_+ \mid X \text{ is bounded, with only a finite numbers of values,} \\ \text{decreasing on } \mathbb{R}_+ \text{ and such that } X(x) = 0 \text{ for } x < 0. \end{array} \right\}$$

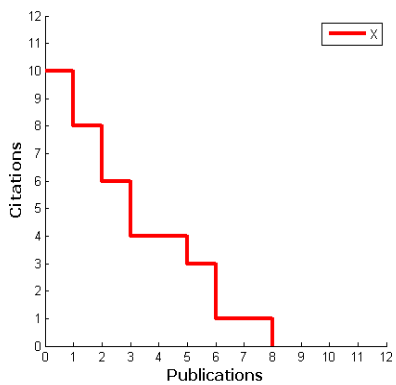


Fig. 1. Author's Citation Curve

We compare the citation curve X of an author with a theoretical citation curve f_q representing the desiderata citations at a fix performance level q. For this purpose we introduce the following class of curves. Let $\mathcal{I} \subseteq \mathbb{R}$ be the index set of the *performance level*. For any $q \in \mathcal{I}$ we define the theoretical *performance curve of level* q as a function $f_q : \mathbb{R} \to \mathbb{R}_+$ that associates to each publication $x \in \mathbb{R}$ the corresponding number of citations $f_q(x) \in \mathbb{R}_+$.

Definition 1 (Performance curves) *Given a index set* $\mathcal{I} \subseteq \mathbb{R}$ *of performance levels* $q \in \mathcal{I}$*, a class* $\mathbb{F} := \{f_q\}_{q \in \mathcal{I}}$ *of functions* $f_q : \mathbb{R} \to \mathbb{R}_+$ *is a* family of performance curves *if*

i) $\{f_q\}$ *is increasing in* q*, i.e. if* $q \geq p$ *then* $f_q(x) \geq f_p(x)$ *for all* x*;*

ii) for each q*,* $f_q(x)$ *is left continuous in* x*;*

iii) $f_q(x) = 0$ *for all* $x < 0$ *and all* q*.*

The main feature of these curves is that a higher performance level implies a higher number of citations. This family of curves is crucial for our objective to build a SRM able to comprehend many of the popular indices and calibrated to the scientific area and the seniority of the authors.

Definition 2 (Performance sets and SRM) *Given a family of performance curves* $\mathbb{F} = \{f_q\}_q$, *we define the* family of performance sets $\mathcal{A}_{\mathbb{F}} := \{\mathcal{A}_q\}_q$ *by*

$$\mathcal{A}_q := \left\{ X \in \mathcal{X}^+ \mid X(x) \geq f_q(x) \text{ for all } x \in \mathbb{R} \right\}.$$

The Scientific Research Measure (SRM) *is the map* $\phi_{\mathbb{F}} : \mathcal{X}^+ \to [0, \infty]$ *associated to* $\mathbb{F}$ *and* $\mathcal{A}_{\mathbb{F}}$ *given by*

$$\begin{aligned} \phi_{\mathbb{F}}(X) \quad : &= \sup \{ q \in \mathcal{I} \mid X \in \mathcal{A}_q \} \\ &= \sup \{ q \in \mathcal{I} \mid X(x) \geq f_q(x) \text{ for all } x \in \mathbb{R} \}. \end{aligned} \tag{1}$$

The SRM $\phi_{\mathbb{F}}$ is obtained by the comparison between the real citation curve of an author X (the red line in Fig.2) and the family $\mathbb{F}$ of performance curves (the blue line in Fig.2): the $\phi_{\mathbb{F}}(X)$ is the greatest level q of the performance curve f_q below the author's citation curve X.

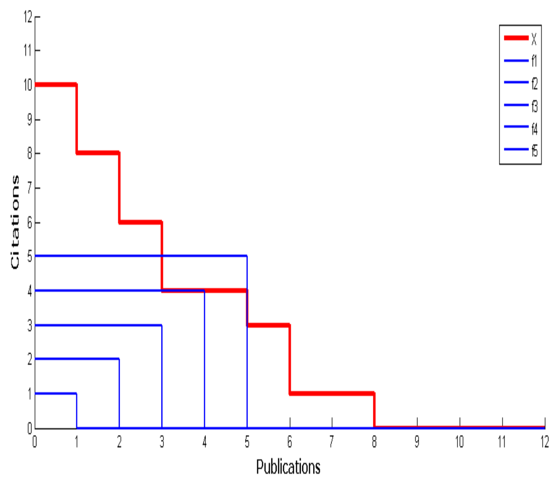


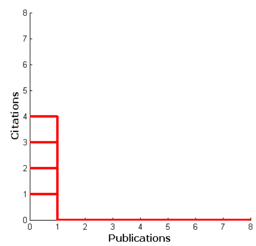
Fig. 2. Determination of a particular SRM, the h-index (that in this example is equal to 4).

2.1 Some examples of existing SRMs

The previous definition points out the importance of the family of theoretical performance curves for the determination of the SRM. It is clear that different choices of $\mathbb{F} := \{f_q\}_q$ lead to different SRM $\phi_{\mathbb{F}}$. The following examples show that some well known indices of scientific performance are particular cases of our SRM. In the following examples, if X has $p \geq 1$ publications that received at least one citation, we set: $X = \sum_{i=1}^{p} x_i 1_{(i-1,i]}$, with $x_i \geq x_{i+1}$ for all i.

Example 3 (max # of citations) *The* maximum number of citations *of the most cited author's paper is the SRM* $\phi_{\mathbb{F}_{c_{\max}}}$ *defined by (1), where the family* $\mathbb{F}_{c_{\max}}$ *of performance curves is:*

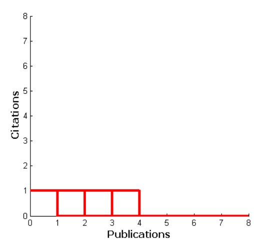
$$f_q(x) = \begin{cases} q & 0 < x \leq 1 \\ 0 & x > 1 \end{cases} \quad \textit{for all } x \in \mathbb{R}_+. \tag{2}$$



(3)

Example 4 (total number of publications) *The* total number of publications *with at least one citation is the SRM* $\phi_{\mathbb{F}_p}$ *defined by (1), where the family* $\mathbb{F}_p$ *of performance curves is:*

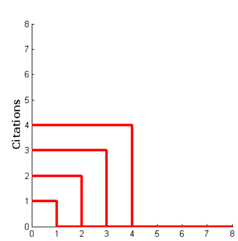
$$f_q(x) = \begin{cases} 1 & 0 < x \leq q \\ 0 & x > q \end{cases} \quad \textit{for all } x \in \mathbb{R}_+. \tag{4}$$



(5)

Example 5 (h-index) *According to the definition given by Hirsch, 2005 [H05]: "A scientist has index* h *if* h *of his or her* N_p *papers have at least* h *citations each and the other* $(N_p - h)$ *papers have* $\leq h$ *citations each". The* h-index *is the SRM* $\phi_{\mathbb{F}_h}$ *defined by (1), where the family* $\mathbb{F}_h$ *of performance curves is:*

$$f_q(x) = \begin{cases} q & 0 < x \leq q \\ 0 & x > q \end{cases} . \quad \textit{for } x \in \mathbb{R}_+. \tag{6}$$



(7)

Example 6 (h^2-index) *Kosmulski, 2006 [K06] defined a scientist's* h^2-index *"as the highest natural number such that his h^2 most cited papers received each at least $[h^2]^2$ citations". This index is the SRM $\phi_{\mathbb{F}_{h^2}}$ defined by (1), where the family $\mathbb{F}_{h^2}$ of performance curves is:*

$$f_q(x) = \begin{cases} q^2 & 0 < x \leq q \\ 0 & x > q \end{cases} \qquad \text{for } x \in \mathbb{R}_+.$$

Example 7 (h_α-index) *Eck and Waltman, 2008 [EW06] proposed the* h_α-index *as a generalization of the h-index so defined: "a scientist has* h_α*-index* h_α *if* h_α *of his n papers have at least* $\alpha \cdot h_\alpha$ *citations each and the other* $n - h_\alpha$ *papers have fewer than* $\leq \alpha \cdot h_\alpha$ *citations each". Hence,* h_α-index *is the SRM* $\phi_{\mathbb{F}_{h_\alpha}}$ *defined by (1),where the family* $\mathbb{F}_{h_\alpha}$ *of performance curves is:*

$$f_q(x) = \begin{cases} \alpha q & 0 < x \leq q \\ 0 & x > q \end{cases} \qquad \text{for } x \in \mathbb{R}_+ \text{ and } \alpha \in (0, \infty).$$

Example 8 (w-index) *Woeginger, 2008 [W0308] introduced the* w-index *defined as: "a* w-index *of at least k means that there are k distinct publications that have at least* 1, 2, 3, 4,..., *k citations, respectively". It is the SRM* $\phi_{\mathbb{F}_w}$ *defined by (1), where the family* $\mathbb{F}_w$ *of performance curves is:*

$$f_q(x) = \begin{cases} q - x + 1 & 0 < x \leq q \\ 0 & x > q \end{cases} \qquad \text{for all } x \in \mathbb{R}_+. \tag{8}$$

Example 9 (h_{rat}-index & h_r-index) *The rational and the real h-index,* h_{rat}-index *and* h_r-index, *introduced respectively by Ruane and Tol, 2008 [RT08] and Guns and Rousseau, 2009 [GR09] are SRMs, indeed they could be defined as the h-index but taking respectively* $q \in \mathcal{I} \subseteq \mathbb{Q}$ *and* $q \in \mathcal{I} \subseteq \mathbb{R}$.

Example 10 (h_m-index) *Schreiber, 2008 [S08] proposed a new index called* h_m-index *that keeps into account the influence of the number of co-authors for a researcher's publication, counting the papers fractionally according to the number of authors. The* h_m-index *is the SRM* $\phi_{\mathbb{F}_{h_m}}$ *defined by (1),where the family* $\mathbb{F}_{h_m}$ *of performance curves is:*

$$f_q(x) = \begin{cases} \sum_{j=1}^{q} \frac{1}{a(j)} & 0 < x \leq q \\ 0 & x > q \end{cases} \qquad \text{for } x \in \mathbb{R}_+,$$

where $a(j)$ is the number of authors for the paper j.

2.2 Key properties of the SRM

Now we point out some relevant properties of the family $\mathcal{A}_{\mathbb{F}} = \{\mathcal{A}_q\}_q$ of performance sets and of the SRM $\phi_{\mathbb{F}}$.

Proposition 11 *Let $X_1, X_2 \in \mathcal{X}^+$.*

1. *If* $\mathcal{A}_{\mathbb{F}} = \{\mathcal{A}_q\}_q$ *is a family of* performance sets *then:*

   i) $\{\mathcal{A}_q\}$ *is decreasing monotone:* $\mathcal{A}_q \subseteq \mathcal{A}_p$ *for any level* $q \geq p$*;*

   ii) $\mathcal{A}_q$ *is monotone for any q:* $X_1 \in \mathcal{A}_q$ *and* $X_2 \geq X_1$ *implies* $X_2 \in \mathcal{A}_q$*;*

   iii) $\mathcal{A}_q$ *is convex for any q: if* $X_1, X_2 \in \mathcal{A}_q$ *then* $\lambda X_1 + (1-\lambda)X_2 \in \mathcal{A}_q$ *for* $\lambda \in [0,1]$*;*

2. *If* $\phi_{\mathbb{F}}$ *is a* SRM *then it is:*

   i) *increasing monotone: if* $X_1 \leq X_2 \Rightarrow \phi_{\mathbb{F}}(X_1) \leq \phi_{\mathbb{F}}(X_2)$*;*

   ii) *quasi-concave:* $\phi_{\mathbb{F}}(\lambda X_1 + (1-\lambda)X_2) \geq \min(\phi_{\mathbb{F}}(X_1),\ \phi_{\mathbb{F}}(X_2))$ *for all* $\lambda \in [0,1]$.

Proof.

1) The proof of the monotonicity and convexity of $\mathcal{A}_{\mathbb{F}}$ follows from the definition.

2.i) It is sufficient to show that

$$\{q \in \mathcal{I} \mid X_1 \geq f_q\} \subseteq \{q \in \mathcal{I} \mid X_2 \geq f_q\}.$$

As $X_1 \leq X_2$, $X_1 \geq f_{\overline{q}}$ implies $X_2 \geq f_{\overline{q}}$.

2.ii) Let $\phi_{\mathbb{F}}(X_1) \geq m$ and $\phi_{\mathbb{F}}(X_2) \geq m$. By definition of $\phi_{\mathbb{F}}$, $\forall \varepsilon > 0$ $\exists q_i$ s.t. $X_i \geq f_{q_i}$ and $q_i > \phi_{\mathbb{F}}(X_i) - \varepsilon \geq m - \varepsilon$. 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 $\phi_{\mathbb{F}}(\lambda X_1 + (1-\lambda)X_2) \geq m$ and $\phi_{\mathbb{F}}$ is quasi-concave.

■

It is obviously reasonable that a SRM should be *increasing:* if the citations of a researcher X_2 dominate the citations of another researcher X_1 publication by publication, then X_2 has a performance greater than X_1.

Now, we introduce a counterexample in order to show that a SRM is not in general quasi-convex, that is $\phi_{\mathbb{F}}(\lambda X_1 + (1-\lambda)X_2) \leq \max(\phi_{\mathbb{F}}(X_1),\ \phi_{\mathbb{F}}(X_2))$ for all $\lambda \in [0,1]$. We consider two researchers, $X_1 = [8 \quad 6 \quad 4 \quad 2]$ and $X_2 = [4 \quad 2 \quad 2 \quad 2 \quad 2]$, where X_2 has more publications than X_1 but less cited. If we compute for example the w-index we obtain that $\phi_{\mathbb{F}_w}(X_1) = 4$ and $\phi_{\mathbb{F}_w}(X_2) = 3$, while taking $\lambda = \frac{1}{2}$ the SRM $\phi_{\mathbb{F}_w}$ of the combined citation curve $X = \frac{1}{2}X_1 + \frac{1}{2}X_2 = [6 \quad 4 \quad 3 \quad 2 \quad 1]$ is $\phi_{\mathbb{F}_w}(X) = 5$.

2.3 Additional properties of SRMs

We have seen that all the SRMs $\phi_{\mathbb{F}}$ share the same structural properties of monotonicity and quasiconcavity. We start this section classifying the SRMs on the basis of the addition of citations to the old papers.

Definition 12 (Additional citation properties) *A SRM* $\phi_{\mathbb{F}} : \mathcal{X}^{+} \to [0, \infty]$ *is*

a) C-*superadditive if* $\phi_{\mathbb{F}}(X + m) \geq \phi_{\mathbb{F}}(X) + m$ *for all* $m \in \mathbb{R}_{+}$*;*

b) C-*subadditive if* $\phi_{\mathbb{F}}(X + m) \leq \phi_{\mathbb{F}}(X) + m$ *for all* $m \in \mathbb{R}_{+}$*;*

c) C-*additive if* $\phi_{\mathbb{F}}(X + m) = \phi_{\mathbb{F}}(X) + m$ *for all* $m \in \mathbb{R}_{+}$*.*

A SRM is C-superadditive (C-subadditive) if the additional citations to the old papers lead an increase of the measure more (less) than linear. In other terms, a C-superadditive SRM gives more weight than the C-subadditive SRM to the additional citations to the oldest papers.

We have seen that the SRM $\phi_{\mathbb{F}}$ depends on the family of performance curves $\mathbb{F} := \{f_q\}_q$ under consideration. The main feature of this family of curves is that is increasing monotone over q. We provide a characterization of this family in terms of the speed in the increase of the performance curves.

Definition 13 *Let* $\mathbb{F}$ *a family of performance curve. We say that:*

a) $\mathbb{F}$ *is* slowly increasing in q *if* $f_{q+m} - f_q \leq m$ *for all* $m \in \mathbb{R}_{+}$*;*

b) $\mathbb{F}$ *is* fast increasing in q *if* $f_{q+m} - f_q \geq m$ *for all* $m \in \mathbb{R}_{+}$*;*

c) $\mathbb{F}$ *is* linear increasing in q *if* $f_{q+m} - f_q = m$ *for all* $m \in \mathbb{R}_{+}$*.*

These properties of the family of performance curves can be express in terms of corresponding properties of the family $\mathcal{A}_{\mathbb{F}}$ of performance sets.

Lemma 14 *Let* $\mathbb{F}$ *a family of performance curve.*

1. $\mathbb{F}$ *is slowly increasing in* q*, if and only if*

$$\mathcal{A}_q + m \subseteq \mathcal{A}_{q+m} \tag{9}$$

for all $m \in \mathbb{R}_{+}$ *and* $q \in \mathcal{I}$;

2. $\mathbb{F}$ *is fast increasing in* q*, if and only if*

$$\mathcal{A}_{q+m} \subseteq \mathcal{A}_q + m \tag{10}$$

for all $m \in \mathbb{R}_{+}$ *and* $q \in \mathcal{I}$;

3. $\mathbb{F}$ *is linear increasing in* q*, if and only if*

$$\mathcal{A}_{q+m} = \mathcal{A}_q + m$$

for all $m \in \mathbb{R}_{+}$ *and* $q \in \mathcal{I}$;

Proof. (1) In order to show that $\mathcal{A}_q + m \subseteq \mathcal{A}_{q+m}$ we observe that:

$$\begin{aligned}\mathcal{A}_{q+m} &:= \left\{X \in \mathcal{X}^+ \mid X \geq f_{q+m}\right\} \\ \mathcal{A}_q + m &= \left\{X \in \mathcal{X}^+ \mid X \geq f_q\right\} + m \\ &= \{X \mid X \geq f_q + m\}\end{aligned}$$

As $f_q + m \geq f_{q+m}$, if $\overline{X}$ is such that $\overline{X} \geq f_q + m$ then $\overline{X} \geq f_{q+m}$. This means that $\overline{X} \in \mathcal{A}_q + m$ implies that $\overline{X} \in \mathcal{A}_{q+m}$.

Regarding the other implication, we know that if $X \in \mathcal{A}_q + m$ then $X \in \mathcal{A}_{q+m}$, that is $X \geq f_q + m$ implies $X \geq f_{q+m}$. This implies that $f_q + m \geq f_{q+m}$.

(2) By hypothesis we know that $f_{q+m} \geq f_q + m$. Hence, if $\overline{X}$ is such that $\overline{X} \geq f_{q+m}$ then $\overline{X} \geq f_q + m$. This means that $\overline{X} \in \mathcal{A}_{q+m}$ implies that $\overline{X} \in \mathcal{A}_q + m$.

Regarding the other side, we know that if $X \in \mathcal{A}_{q+m}$ then $X \in \mathcal{A}_q + m$, that is $X \geq f_{q+m}$ implies $X \geq f_q + m$. This implies that $f_{q+m} \geq f_q + m$.

(3) The proof of this point follows directly from the previous ones, observing that $\mathbb{F}$ is linear increasing in q if and only if $\mathbb{F}$ is slowly and fast increasing in q and $\mathcal{A}_{q+m} = \mathcal{A}_q + m$ if both of the inclusions (9) and (10) hold. ■

The following lemma shows that the additional citation properties of the SRM $\phi_{\mathbb{F}}$ can be *built in* from the corresponding properties of the family $\mathbb{F}$ of the performance curves or $\mathcal{A}_{\mathbb{F}}$ of performance sets.

Lemma 15 *Let $\mathbb{F}$ a family of performance curves.*

1. *If $\mathbb{F}$ is slowly increasing in q, then $\phi_{\mathbb{F}}$ is* C-*superadditive;*
2. *If $\mathbb{F}$ is fast increasing in q, then $\phi_{\mathbb{F}}$ is* C-*subadditive;*
3. *If $\mathbb{F}$ is linear increasing in q, then $\phi_{\mathbb{F}}$ is* C-*additive.*

Proof. (1) In order to show that $\phi_{\mathbb{F}}(X + m) - m \geq \phi_{\mathbb{F}}(X)$ for all $m \in \mathbb{R}_+$ we use the definition in (1) and we observe that

$$\begin{aligned}\phi_{\mathbb{F}}(X + m) - m &= \sup\{q \in \mathcal{I} \mid X + m \geq f_q\} - m \\ &= \sup\{q - m \in \mathcal{I} \mid X \geq f_q - m\} \\ &= \sup\{q \in \mathcal{I} \mid X \geq f_{q+m} - m\} \qquad (11)\end{aligned}$$

Hence it's sufficient to show that $\{q \mid \ X \geq f_q\} \subseteq \{q \mid X \geq f_{q+m} - m\}$ and this is true since $f_q \geq f_{q+m} - m$;

(2) In order to show that $\phi_{\mathbb{F}}(X + m) - m \leq \phi_{\mathbb{F}}(X)$ for all $m \in \mathbb{R}^+$ we use the definition in (1) and the relation (11). Hence it's sufficient to show that $\{q \mid \ X \geq f_{q+m} - m\} \subseteq \{q \mid X \geq f_q\}$ and this is true since $f_{q+m} - m \geq f_q$;

(3) It follows directly from the previous points observing that $\mathbb{F}$ is linear increasing in q if and only if it is slowly and fast increasing in q and that $\phi_{\mathbb{F}}$ is C-additive if and only if it is C-superadditive and C-subadditive. ■

Now we give some examples using some popular SRMs.

Example 16 *The h-index in the example (5) is a* C*-subadditive SRM, but the associated family* $\mathbb{F}$ *of performance curves defined in (6) is fast increasing in* q*. Indeed the property is true only on* $[0, q+m]$ *for any* $m \in \mathbb{R}_+$ *since the performance curves are equal to zero outside. Hence, the performance curves of the h-index are fast increasing only in the Hirsch core.*

The same considerations hold for the h^2*- and* h_α*- index (see examples (6) and (7)).*

Example 17 *The family* $\mathbb{F}$ *defined in 8 of the* w*-index (see example 8) is slowly increasing in* q*. This condition is sufficient to say that the* w*-index is a* C*-superadditive SRM.*

Example 18 *The maximum number of citations of an article (see example 3) is a* C*-additive SRM, even if the family* $\mathbb{F}$ *of performance curves defined in 2 is not linear increasing in* q*. This property holds only on* $[0, 1]$*, since the performance curves are equal to zero outside.*

Example 19 *The total number of publications (see example 4) is a* C*-superadditive SRM since the family* $\mathbb{F}$ *of performance curves defined in 4 is slowly increasing in* q*.*

We now define further properties linked to the addition of a single publication to the author's citation curve.

Definition 20 (Additional paper properties) *Let* $p := \max\{x : X(x) > 0\}$ *the maximum number of publications with at least one citation of the author* X*. A SRM* $\phi_{\mathbb{F}} : \mathcal{X}^+ \to \mathbb{R}_+$ *is*

a) P*-superadditive if* $\phi_{\mathbb{F}}(X + 1_{\{p+1\}}) \geq \phi_{\mathbb{F}}(X) + 1$*;*

b) P*-subadditive if* $\phi_{\mathbb{F}}(X + 1_{\{p+1\}}) \leq \phi_{\mathbb{F}}(X) + 1$*;*

c) P*-additive if* $\phi_{\mathbb{F}}(X + 1_{\{p+1\}}) = \phi_{\mathbb{F}}(X) + 1$*;*

c) P*-invariance if* $\phi_{\mathbb{F}}(X + 1_{\{p+1\}}) = \phi_{\mathbb{F}}(X)$*.*

A SRM is P-superadditive if the addition of one citation to a new publication leads to an increase of the measure more than linear. Someway if we use a P-superadditive SRM in our evaluation we are giving more weight to the additional publication than in case of P-subadditive SRM. Many known SRMs are P-invariance (i.e. the $c_{\max}$, h-, h^2- and h_α-index in the examples (3) (5), (6) and (7)) as the addition of one citation to a new publication leaves the SRM invariant. The w-index (in the example (8)) is P-subadditive as the addition of one citation to a new publication makes it greater at most of 1 unit. While the total number of publications p with at least one citation (in the example (4)) is clearly P-additive.

3 On the Dual Representation of the SRM

The goal of this section is to provide a dual representation of the SRM. To this scope, we need some topological structure. Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ be a probability space, where $\mathcal{B}(\mathbb{R})$ is the σ-algebra of the Borel sets, μ is a probability measure on $\mathcal{B}(\mathbb{R})$. Since the citation curve of an author X is a bounded function, it appears natural to take $X \in L^\infty(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, where $L^\infty(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ is the space of $\mathcal{B}(\mathbb{R})$-measurable functions that are μ almost surely bounded. 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.

We have seen in the Section 1 that the SRM is a quasi-concave and monotone map. Under appropriate continuity assumptions, the dual representation of these type of maps can be found in [PV90],[Vo98], [CMMM].

Definition 21 *A map $\phi : L^\infty(\mathbb{R}) \to \mathbb{R}$ is $\sigma(L^\infty, L^1)-$upper-semicontinuous if the upper level sets*

$$\{X \in L^\infty(\mathbb{R}) \mid \phi(X) \geq q\}$$

are $\sigma(L^\infty, L^1)-$closed for all $q \in \mathbb{R}$.

Lemma 22 *If $\mathcal{A}_\mathbb{F} = \{\mathcal{A}_q\}_q$ is a family of performance sets then $\mathcal{A}_q$ is $\sigma(L^\infty, L^1)$-closed for any q.*

Proof. To prove that $\mathcal{A}_q$ is $\sigma(L^\infty, L^1)$-closed let $Y_n \in \mathcal{A}_q := \{X \in L^\infty \mid X \geq f_q\}$ satisfy $Y_n \overset{\sigma(L^\infty,L^1)}{\to} Y$. By contradiction, suppose that $\mu(A) > 0$ where $A := \{x \in \mathbb{R} \mid Y(x) < f_q(x)\} \in \mathcal{B}(\mathbb{R})$. Taking as a continuous linear functional $Z = 1_A \in L^1$, from $Y_n \overset{\sigma(L^\infty,L^1)}{\to} Y$ we deduce: $E[1_A f_q] \leq E[1_A Y_n] \to E[1_A Y] < E[1_A f_q]$. ■

The following lemma shows the relation between the continuity property of the family $\mathbb{F}$ of performance curves, those of the family $\mathcal{A}_\mathbb{F}$ of performance sets and those of the SRM $\phi_\mathbb{F}$.

Lemma 23 *Let $\mathbb{F}$ be a family of performance curves. If $\mathbb{F}$ is* left continuous in *q, that is*

$$f_{q-\varepsilon}(x) \uparrow f_q(x) \text{ for } \varepsilon \downarrow 0, \text{ for all } x,$$

then:

1. *$\mathcal{A}_\mathbb{F}$ is left-continuous in q, that is*

$$\mathcal{A}_q = \bigcap_{\epsilon > 0} \mathcal{A}_{q-\varepsilon},$$

2.

$$\mathcal{A}_q = \{X \in L^\infty \mid \phi_{\mathbb{F}}(X) \geq q\}, \text{ for all } q \in \mathcal{I}. \tag{12}$$

3. $\phi_{\mathbb{F}}$ *is* $\sigma(L^\infty, L^1)$*-upper-semicontinuous.*

Proof.

1. By assumption we have that $f_{q-\varepsilon}(x) \uparrow f_q(x)$ for $\varepsilon \to 0$, for all $x \in \mathbb{R}$. We have proved in Proposition (11) that $\{\mathcal{A}_q\}$ is decreasing monotone, hence we know that $\mathcal{A}_q \subseteq \bigcap_{\varepsilon>0} \mathcal{A}_{q-\varepsilon}$. By contradiction we suppose that

$$\bigcap_{\varepsilon>0} \mathcal{A}_{q-\varepsilon} \supsetneqq \mathcal{A}_q,$$

so that there will exist $X \in L^\infty$ such that $X \geq f_{q-\varepsilon}$ for every $\varepsilon > 0$ but $X(A) < f_q(A)$ for some $A \in \mathcal{B}(\mathbb{R})$ such that $\mu(A) > 0$. Then there exists $\delta > 0$ such that $f_q(x) - X(x) \geq \delta$ for any x in $B \subseteq A$ such that $\mu(B) > 0$. Then $f_q(x) - X(x) > \frac{\delta}{2}$ for any $x \in B$. Since $f_{q-\varepsilon} \uparrow f_q$ we may find $\varepsilon > 0$ such that $f_q(x) - f_{q-\varepsilon}(x) < \frac{\delta}{2}$ for $x \in B$. Thus $X(x) \geq f_{q-\varepsilon}(x) > f_q(x) - \frac{\delta}{2}$ for $x \in B$ and this is a contradiction

2. Now let

$$B_q := \{X \in L^\infty \mid \phi_{\mathbb{F}}(X) \geq q\}.$$

$\mathcal{A}_q \subseteq B_q$ follows directly from the definition of $\phi_{\mathbb{F}}$. We have to show that $B_q \subseteq \mathcal{A}_q$. Let $\overline{X} \in B_q$. Hence $\phi_{\mathbb{F}}(\overline{X}) \geq q$ and for all $\varepsilon > 0$ there exists $\overline{q}$ such that $\overline{q} + \varepsilon \geq q$ and $\overline{X}(x) \geq f_{\overline{q}}(x)$ for all $x \in \mathbb{R}$. Since f_q are increasing in q we have that $\overline{X}(x) \geq f_{q-\varepsilon}(x)$ for all $x \in \mathbb{R}$ and $\varepsilon > 0$, therefore $\overline{X} \in \mathcal{A}_{q-\varepsilon}$. By the left continuity in q of the family $\mathbb{F}$ we have know that $\{\mathcal{A}_q\}$ is left-continuous in q for the previous item and so: $\overline{X} \in \bigcap_{\epsilon>0} \mathcal{A}_{q-\varepsilon} = \mathcal{A}_q$.

3. By Lemma (22) we know that $\mathcal{A}_q$ is $\sigma(L^\infty, L^1)-$closed for any q and therefore the upper level sets $B_q = \mathcal{A}_q$ are $\sigma(L^\infty, L^1)-$closed and $\phi_{\mathbb{F}}$ is $\sigma(L^\infty, L^1)$ upper semicontinuous.

■

Notice that $\sigma(L^\infty, L^1)$-upper semicontinuity is equal to the continuity from above of a SRM. This fact can be proved in a way similar to the convex case (see for example [FS04]).

Lemma 24 *Let* $\phi_{\mathbb{F}} : L^\infty \to \mathbb{R}_+$ *be a SRM. Then the following are equivalent:*
$\phi_{\mathbb{F}}$ *is* $\sigma(L^\infty, L^1)$*-upper semicontinuous;*
$\phi_{\mathbb{F}}$ *is continuous from above:* $X_n, X \in L^\infty$ *and* $X_n \downarrow X$ *imply* $\phi_{\mathbb{F}}(X_n) \downarrow \phi_{\mathbb{F}}(X)$

Proof. Let $\phi_{\mathbb{F}}$ be $\sigma(L^\infty, L^1)$-upper semicontinuous and suppose that $X_n \downarrow X$. As the elements in L^1 are order continuous, we also have: $X_n \overset{\sigma(L^\infty,L^1)}{\longrightarrow} X$.

The monotonicity of $\phi_{\mathbb{F}}$ implies $\phi_{\mathbb{F}}(X_n) \downarrow$ and $q := \lim_n \phi_{\mathbb{F}}(X_n) \geq \phi_{\mathbb{F}}(X)$. Hence $\phi_{\mathbb{F}}(X_n) \geq q$ and $X_n \in B_q := \{Y \in L^\infty \mid \phi_{\mathbb{F}}(Y) \geq q\}$ which is closed by assumption. Hence $X \in B_q$, which implies that $\phi_{\mathbb{F}}(X) = q$ and that $\phi_{\mathbb{F}}$ is continuous from above.

Conversely, suppose that $\phi_{\mathbb{F}}$ 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 $C := B_q \cap \{X \in L^\infty \mid \ \| X \|_\infty < r\}$ is $\sigma(L^\infty, L^1)$-closed for any fixed $r > 0$ and $q \in \mathbb{R}$. 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 $\phi_{\mathbb{F}}(Y_n) \geq q$ 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 $Z_m \geq Y_m$, $\phi_{\mathbb{F}}(Z_m) \geq \phi_{\mathbb{F}}(Y_m)$ and $Z_m \downarrow X$. From the continuity from above we conclude: $\phi_{\mathbb{F}}(X) = \lim_m \phi_{\mathbb{F}}(Z_m) \geq \limsup_m \phi_{\mathbb{F}}(Y_m) \geq q$. Thus $X \in B_q$ and consequently $X \in C$. ■

When the family of performance curves $\mathbb{F}$ is left continuous, Lemma (23) shows that the SRM is $\sigma(L^\infty, L^1)$-upper semicontinuous. Hence we can provide a dual representation for the SRM in the same spirit of [Vo98] and [DK10].

Denote

$$\mathcal{P} := \{Q \ll P\} \text{ and } \mathcal{Z} := \left\{ Z = \frac{dQ}{dP} \mid Q \in \mathcal{P} \right\} = \left\{ Z \in L^1_+ \mid E[Z] = 1 \right\}$$

Theorem 25 *Suppose that the family of performance curves $\mathbb{F}$ is left continuous. Each SRM $\phi_{\mathbb{F}} : L^\infty(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu) \to \mathbb{R}$ defined in (1) can be represented as*

$$\begin{aligned} \phi_{\mathbb{F}}(X) &= \inf_{Z \in \mathcal{Z}} H(Z, E[ZX]) = \inf_{Z \in \mathcal{Z}} H^+(Z, E[ZX]) \\ &= \inf_{Q \in \mathcal{P}} H^+(Q, E_Q[X]) \quad \textit{for all } X \in L^\infty \end{aligned} \tag{13}$$

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

$$H(Z, t) := \sup_{\xi \in L^\infty} \{\phi_{\mathbb{F}}(\xi) \mid E[\xi Z] \leq t\},$$

$H^+(Z, \cdot)$ is its right continuous version:

$$H^+(Z, t) := \inf_{s > t} H(Z, s) = \sup\{q \in \mathbb{R} \mid t \geq \gamma(Z, q)\}, \tag{14}$$

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

$$\gamma(Z, q) := \inf_{X \in L^\infty} \{E[ZX] \mid \phi_{\mathbb{F}}(X) \geq q\}. \tag{15}$$

Proof. Step 1: $\phi_{\mathbb{F}}(X) = \inf_{Z\in\mathcal{Z}} H(Z, E[ZX])$.

Fix $X \in L^\infty$. As $X \in \{\xi \in L^\infty \mid E[Z\xi] \le E[ZX]\}$, by the definition of $H(Z,t)$ we deduce that, for all $Z \in L^1$,

$$H(Z, E[ZX]) \ge \phi_{\mathbb{F}}(X)$$

hence

$$\inf_{Z\in L^1} H(Z, E[ZX]) \ge \phi_{\mathbb{F}}(X).$$

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

$$C_\varepsilon := \{\xi \in L^\infty \mid \phi_{\mathbb{F}}(\xi) \ge \phi_{\mathbb{F}}(X) + \varepsilon\}$$

As $\phi_{\mathbb{F}}$ is quasi-concave and $\sigma(L^\infty, L^1)$-upper semicontinuous, C is convex and $\sigma(L^\infty, L^1)-$closed. Since $X \notin C_\varepsilon$, the Hahn Banach theorem implies the existence of a continuous linear functional that strongly separates X and C_ε, that is there exist $k \in \mathbb{R}$ and $Z_\varepsilon \in L^1$ such that

$$E[\xi Z_\varepsilon] > k > E[XZ_\varepsilon] \text{ for all } \xi \in C_\varepsilon.$$

Hence

$$\{\xi \in L^\infty \mid E[\xi Z_\varepsilon] \le E[XZ_\varepsilon]\} \subseteq C_\varepsilon^c := \{\xi \in L^\infty \mid \phi_{\mathbb{F}}(\xi) < \phi_{\mathbb{F}}(X) + \varepsilon\}$$

and

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

Therefore, $\phi_{\mathbb{F}}(X) = \inf_{Z\in L^1} H(Z, E[ZX])$. 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 $\phi_{\mathbb{F}}$ is monotone increasing, $\xi + nY \in C_\varepsilon$ for every $n \in \mathbb{N}$ and we have:

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

As this holds for any $Y \in L^\infty_+$ we deduce that $Z_\varepsilon \subseteq L^1_+$. Therefore, $\phi_{\mathbb{F}}(X) = \inf_{Z\in L^1_+} H(Z, E[ZX])$.

By definition of $H(Z,t)$,

$$H(Z, E[ZX]) = H(\lambda Z, E[X(\lambda Z)]) \quad \forall Z \in L^1 \text{ and } Z \ne 0.$$

Hence we deduce

$$\phi_{\mathbb{F}}(X) = \inf_{Z\in L^1_+(\mathbb{R})} H(Z, E[ZX]) = \inf_{Z\in\mathcal{Z}} H(Z, E[ZX]) = \inf_{Q\in\mathcal{P}} H(Q, E_Q[X]).$$

Step 2: $\phi_{\mathbb{F}}(X) = \inf_{Z\in\mathcal{Z}} H^{+}(Z, E[ZX])$.

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

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

$$\phi_{\mathbb{F}}(X) = \inf_{Z\in L^1_+} H(Z, E[ZX]) \le \inf_{Z\in L^1_+} H^{+}(Z, E[ZX]) \le \inf_{Z\in L^1_+} \lim_{X_m\downarrow X} H(Z, E[X_m Z])$$

$$= \lim_{X_m\downarrow X} \inf_{Z\in L^1_+} H(Z, E[X_m Z]) = \lim_{X_m\downarrow X} \phi_{\mathbb{F}}(X_m) \overset{(CFA)}{=} \phi_{\mathbb{F}}(X).$$

Step 3: $H^{+}(Z,t) := \inf_{s>t} H(Z,s) = \sup\{q \in \mathbb{R} \mid t \ge \gamma(Z,q)\}$.

Now let the RHS of equation (14) be denoted by

$$S(Z,t) := \sup\{q \in \mathbb{R} \mid \gamma(Z,q) \le t\}, \ (Z,t) \in L^1 \times \mathbb{R}, \tag{16}$$

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

$$H(Z,p) \le S(Z,p), \tag{17}$$

Indeed, if (17) is true

$$H^{+}(Z,t) = \inf_{p>t} H(Z,p) \le \inf_{p>t} S(Z,p) = S(Z,t),$$

as both H^{+} and S are right continuous in the second argument.
Writing explicitly the inequality (17)

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

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

Suppose now that $\infty > \phi_{\mathbb{F}}(\xi) > -\infty$ and define $q := \phi_{\mathbb{F}}(\xi)$. As $E[\xi Z] \le p$ we have:

$$\gamma(Z,q) := \inf\{E[\xi Z] \mid \phi_{\mathbb{F}}(\xi) \ge q\} \le p.$$

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

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

$$\sup_{\xi\in L^\infty} \{\phi_{\mathbb{F}}(\xi) \mid E[\xi Z] \le p\} \ge \sup\{q \in \mathbb{R} \mid \gamma(Z,q) \le t\}. \tag{18}$$

Fix any $p > t$ and consider any $q \in \mathbb{R}$ such that $\gamma(Z,q) \le t$. By the definition of γ, for all $\varepsilon > 0$ there exists $\xi_\varepsilon \in L^\infty$ such that $\phi_{\mathbb{F}}(\xi_\varepsilon) \ge q$ and $E[\xi_\varepsilon Z] \le t + \varepsilon$. Take ε such that $0 < \varepsilon < p - t$. Then $E[\xi_\varepsilon Z] \le p$ and $\phi_{\mathbb{F}}(\xi_\varepsilon) \ge q$ and (18) follows. ∎

Remark 26 *This dual representation provides an interesting interpretation of the SRM. Let* Q *be the 'weight' that we can assign to the author's publications (for example, the impact factor of the Journal where the article is published). For a fixed* Q*, the term* $\gamma(Q,q) := \inf\{E_Q[\xi] \mid \phi_{\mathbb{F}}(\xi) \geq q\}$ *represents the smallest* Q*-average of citations that a generic author needs in order to have the SRM at least of* q*. We observe that this term is independent from the citations of the author* X*.*

On the light of these considerations we can interpret the term $H^+(Q, E_Q[X]) := \sup\{q \in \mathbb{R} \mid E_Q[X] \geq \gamma(Q,q)\}$ *as the greatest performance level that the author* X *can reach, in the case that we attribute the weight* Q *to the publications. Namely, we compare the* Q*-average of the author* X*,* $E_Q[X]$*, with the minimum* Q*-average necessary to reach each level* q*, that is* $\gamma(Q,q)$*.*

Finally, the SRM of the author X*,* $\phi_{\mathbb{F}}(X) = \inf_{Q\in\mathcal{P}} H^+(Q, E_Q[X])$*, corresponds to the smallest performance level obtained changing the weight attributed to the journals.*

The theorem exhibits the relationship between the performance curve approach and this average approach.

In the following examples we find *the dual representation of some existing indices*. In all these examples the family $\mathbb{F}$ of performance curves is left continuous hence, by Lemma (23), the associated SRM $\phi_{\mathbb{F}}$ is $\sigma(L^\infty, L^1)$-upper semicontinuous and X satisfies: $\phi_{\mathbb{F}}(X) \geq q$ iff $X \in \mathcal{A}_q$ iff $X \geq f_q$. Therefore, we find the dual representation computing γ, H^+ and $\phi_{\mathbb{F}}$ applying the formulas: (15),(14) and (13). Recall that $X = \sum_{i=1}^p x_i 1_{(i-1,i]}$, with $x_i \geq x_{i+1}$ for all i.

Example 27 (max # of citations) *Consider the example (3). For* $Z \in L^1_+$*, we compute* $\gamma(Z,q)$

$$\gamma(Z,q) := \inf_{\phi_{\mathbb{F}_{c_{\max}}}(X)\geq q} E[ZX] = \inf_{X(x)\geq q1_{(0,1]}(x)} E[ZX] = qE[1_{(0,1]}Z]$$

where the first equality is due to (12). We obtain

$$H^+(Z, E[ZX]) := \sup\left\{q \in \mathbb{R} \mid E[ZX] \geq qE[1_{(0,1]}Z]\right\} = \frac{E[ZX]}{E[1_{(0,1]}Z]}.$$

In our application, any non zero citation vector X *always satisfies* $X \geq x_1 1_{(0,1]}$ *and, since* $E[X1_{(0,1]}] = x_1 E[1_{(0,1]}]$*, we also have:* $\frac{1_{(0,1]}}{E[1_{(0,1]}]} \leq \frac{X}{E[X1_{(0,1]}]}$*. Therefore,*

$$E\left[Z\frac{1_{(0,1]}}{E[1_{(0,1]}]}\right] \leq E\left[Z\frac{X}{E[X1_{(0,1]}]}\right] \quad \forall Z \in L^1_+(\mathbb{R})$$

and

$$\frac{E\left[ZX\right]}{E\left[Z1_{(0,1]}\right]} \geq \frac{E[1_{(0,1]}X]}{E[1_{(0,1]}]} \quad \forall Z \in L^1_+(\mathbb{R}).$$

Hence:

$$\phi_{\mathbb{F}_{c_{\max}}}(X) = \inf_{Z\in L^1_+(\mathbb{R})} H^+(Z, E[ZX]) = \inf_{Z\in L^1_+(\mathbb{R})} \frac{E[ZX]}{E[Z1_{(0,1]}]}$$
$$= \frac{E[1_{(0,1]}X]}{E[1_{(0,1]}1_{(0,1]}]},$$

i.e. the infimum is attained at $Z = 1_{(0,1]} \in L^1_+$*, which is of course natural as this SRM weights only to the first publication.*

Example 28 (total # of publications) *Consider the example (4). For* $Z \in L^1_+$*, we compute* $\gamma(Z, q)$ *as in the previous example:*

$$\gamma(Z, q) = \inf_{X\geq 1_{(0,q]}} E[ZX] = E[1_{(0,q]}Z]$$

We obtain

$$H^+(Z, E[ZX]) := \sup \left\{q \in \mathbb{R} \mid E[ZX] \geq E[1_{(0,q]}Z]\right\}$$

Hence the dual representation of the total number of publications p *with at least one citation is*

$$\phi_{\mathbb{F}_p}(X) = \inf_{Z\in L^1_+(\mathbb{R})} \sup_{E[ZX]\geq E[1_{[0,q]}Z]} q$$

We show indeed that $\phi_{\mathbb{F}_p}(X) = p$*, where* p *is such that* $X = X1_{(0,p]} \in L^\infty_+$*. First we check that* $\phi_{\mathbb{F}_p}(X) \geq p$*. For all* $Z \in L^1_+$*, and* $q \leq p$ *we have*

$$E[ZX] = E[ZX1_{(0,p]}] \geq E[1_{(0,q]}Z]$$

and therefore

$$\sup_{E[ZX]\geq E[1_{(0,q]}Z]} q \geq p \qquad \forall Z \in L^1_+,$$

and $\phi_{\mathbb{F}_p}(X) \geq p$*. Regarding the* $\leq$ *inequality, it is enough to take* $Z = 1_{(p,p+\delta]}$*, with* $\delta > 0$*, for* $X = X1_{(0,p]}$*. In this case, the condition* $E[ZX] \geq E[1_{(0,q]}Z]$ *becomes*

$$0 = E[1_{(p,p+\delta]}X] \geq E[1_{(0,q]}1_{(p,p+\delta]}]$$

that holds only for $q \leq p$*, hence*

$$\sup_{E[X1_{(p,p+\delta]}]\geq E[1_{(0,q]}1_{(p,p+\delta]}]} q = p$$

and $\phi_{\mathbb{F}_p}(X) \leq p$*.*

Example 29 (h-index) *Consider the example (5). For* $Z \in L^1_+$*,*

$$\gamma(Z, q) = \inf_{X(x)\geq q1_{(0,q]}(x)} E[ZX] = E[Zq1_{(0,q]}]$$

We obtain

$$H^+(Z, E[ZX]) := \sup\left\{q \in \mathbb{R} \mid E[ZX] \geq E[Zq1_{(0,q]}]\right\}$$

Hence the dual representation of the h-index *is*

$$\phi_{\mathbb{F}_h}(X) = \inf_{Z \in L^1_+(\mathbb{R}^+)} \sup_{E[ZX] \geq E[Zq1_{(0,q]}]} q$$

We indeed show that $\phi_{\mathbb{F}_h}(X) = h$*, where* h *is such that* $X1_{(0,h]} \geq h1_{(0,h]}$ *and* $X1_{(h,+\infty)} \leq h1_{(h,+\infty)}$*. First we check that* $\phi_{\mathbb{F}_h}(X) \geq h$*. For all* $Z \in L^1_+$*, and* $q \leq h$ *we have*

$$E[ZX] \geq E[ZX1_{(0,h]}] \geq E[Zq1_{(0,q]}],$$

hence

$$\sup_{E[ZX] \geq E[q1_{(0,q]}Z]} q \geq h \qquad \forall Z \in L^1_+$$

and $\phi_{\mathbb{F}_h}(X) \geq h$*.*

Regarding the $\leq$ *side, take* $Z = 1_{(h,h+\delta]}$ *with* $\delta > 0$*. For any* $q \leq h$ *the condition*

$$E[X1_{(h,h+\delta]}] \geq E[q1_{(0,q]}1_{(h,h+\delta]}] = 0$$

holds. Instead, $\forall q > h$ *there exists* $\delta > 0$ *such that* $h + \delta < q$ *and then*

$$E[X1_{(h,h+\delta]}] \leq E[h1_{(h,h+\delta]}] < E[q1_{(0,q]}1_{(h,h+\delta]}]$$

hence

$$\sup_{E[X1_{(h,h+\delta]}] \geq E[q1_{(0,q]}1_{(h,h+\delta]}]} q \leq h$$

and $\phi_{\mathbb{F}_h}(X) \leq h$*.*

3.1 On an alternative approach to SRMs

The dual representation suggests us another approach for the definition of a generic class of SRMs. This approach is based on the assumption that we can represent the author's citation as a function $X(w)$ defined on the events $w \in \Omega$, where each event now corresponds to the journal in which the paper appeared.

We start fixing a plausible family $\mathcal{P} \subseteq \{Q \ll P\}$ where each $Q(w)$ represents the *'value'* attributed to the journal $w \in \Omega$. It is clear that the valuation criterion for journals (i.e. the selection of the family $\mathcal{P}$) has to be determined a priori and could be based on the *'impact factor'* or other criterion. A specific Q could attribute more importance to the journals with a large number of citations (a large impact factor); another particular Q to the journals having a high quality.

As suggested from the dual representation results and in particular from the equations (13) and (14) we consider, independently to the particular scientist X, a family $\left\{\gamma_\beta\right\}_{\beta \in \mathbb{R}}$ of functions $\gamma_\beta : \mathcal{P} \to \overline{\mathbb{R}}$ that associate to each Q the value $\gamma_\beta(Q)$, that should represent the smallest Q-average of citations in order to reach a quality index at least of β.

So given a particular value $Q(w_i)$ for each i^{th}-journal and the average citations $\gamma_\beta(Q)$ necessary to have an index level greater than β, we build the SRM in the following way. We define the function $H^+ : \mathcal{P} \times \mathbb{R} \to \overline{\mathbb{R}}$ that associates to each pair $(Q, E_Q(X))$ the number

$$H^+(Q, E_Q(X)) := \sup \left\{ \beta \in \mathbb{R} \mid E_Q(X) \geq \gamma_\beta(Q) \right\},$$

which represents the greatest quality index that the author X can reach when Q is fixed, and we build the SRM as follows:

$$\phi(X) := \inf_{Q \in \mathcal{P}} H^+(Q, E_Q(X))$$

which represents a prudential and robust approach with respect to $\mathcal{P}$, the plausible different selections of the evaluation of the Journals. This SRM is by construction *quasi-concave* and *monotone increasing*.

4 Empirical results

Since the SRM introduced in Section 2 depend on the particular family $\mathbb{F}$ of performance curves, in this section we provide a procedure to *calibrate* the family $\mathbb{F}$ from the historic data available for one particular scientific area and seniority. In this way, each SRM will fit appropriately the characteristics of the research field and seniority under consideration. The SRM should be used only in *relative* terms (to compare the author quality with respect to the other researchers in the same area) in order to classify the authors (and structures) into few classes of homogeneous research quality.

4.1 Sample setting

The first step consists in the selection of a representative sample of M authors in the same scientific area and with the same seniority.

If p is the total number of the author's publications with at least one citation, then $X = \sum_{i=1}^{p} x_i 1_{(i-1,i]}$, with $x_i \geq x_{i+1}$ for all i, where the first component x_1 corresponds to the number of citations received by the most cited article and similarly for $x_1 \geq x_2 \geq ... \geq x_p$.

The citation data of each author are downloaded from Google Scholar by a procedure implemented in Pyton. This procedure performs a filter on the name of the author and on the scientific area we are analyzing.

4.2 Determination of the family $\{f_q\}_q$ and of the SRM

First of all we need to determine the family of curves $\{f_q\}_q$ that better represents the citation curve of the sample of the selected scientists. By the analysis of the data we found that the theoretical model is the following hyperbole-type equation:

$$y = f_q(x) = \frac{q}{x^{\beta}} \tag{19}$$

with $q, \beta \in \mathbb{R}_+$. Setting $\ln y = Y$, $\ln(q) = \hat{q}$, $\ln x = X$, $\beta = \hat{\beta}$ we obtain the linearized model

$$Y = \hat{q} - \hat{\beta} X. \tag{20}$$

For each i-th author of the sample we determine $\hat{\beta}_i$ that minimizes the sum of the square distances of the points from the line (20). Fixing the parameter $\bar{\beta}$, we obtain the ϕ-index of each author X as

$$\phi(X) = \sup \left\{ q \in \mathbb{R} \mid X(x) \geq \frac{q}{x^{\bar{\beta}}} \qquad \forall x \right\} \tag{21}$$

4.3 Our Results

We have chosen a group of 20 well established researchers in the mathematical finance area. The analysis of the citation vectors of each author (see Fig.4.3) brings out that the theoretical model is the in the formula (19). We have computed the $\hat{\beta}_i$ for each author and we have found that $\bar{\beta} = 1,62$.

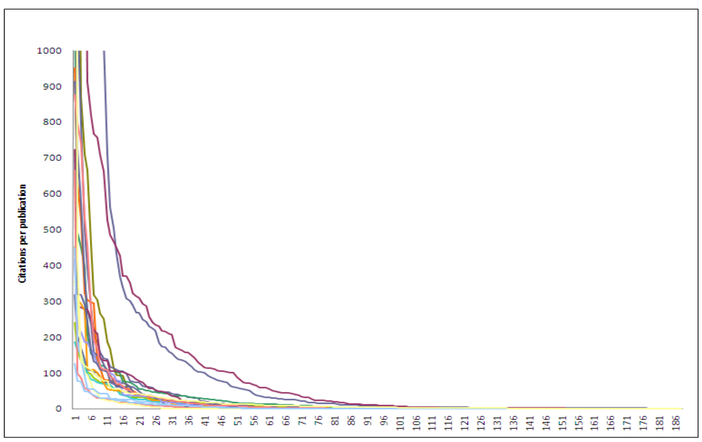


Fig.4.3. Citation curves of 20 senior authors in Math Finance area.

In the following table (Fig.4.3.a) we report the results and the respective ranking obtained calculating the ϕ-index as in (21) and the h-index for each author. Fig.4.3.b shows that the hyperbole-type curve (red line) corresponding to the author's ϕ-index is always below his citation curve (blue line).

Author	ø-index	rank ø-index	h-index	rank h-index
Author A	4423	1	53	2
Author B	2985	2	60	1
Author D	1235	3	35	3
Author E	1136	4	35	4
Author F	950	5	25	14
Author C	908	6	28	7
Author R	875	7	28	8
Author T	800	8	29	6
Author P	780	9	28	9
Author H	723	10	33	5
Author G	511	11	26	11
Author L	451	12	24	15
Author Q	449	13	20	17
Author M	417	14	27	10
Author J	318	15	26	12
Author N	304	16	17	19
Author I	240	17	26	13
Author O	221	18	15	20
Author K	186	19	23	16
Author S	127	20	18	18

Fig. 4.3.a. Comparison of the values and the ranking obtained with the ϕ- and h-index.

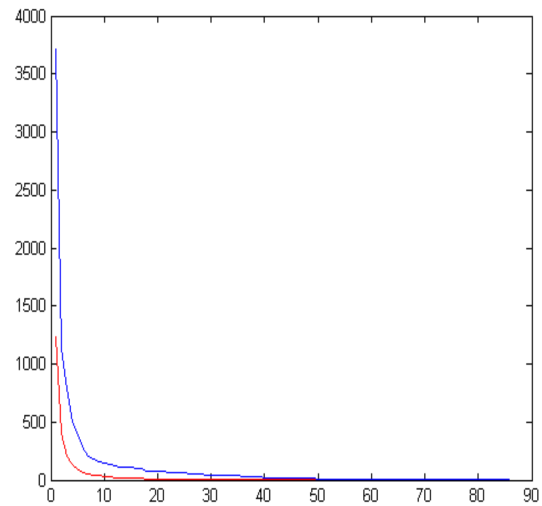


Fig. 4.3.b

We note that the ϕ-index is more granular, allowing a more precise comparison between scientists. For example, the author F increases his index, moving from the position 14 of h-index to 5 of ϕ-index. If we compare this author with the author I, we note that they have almost the same h-index but the F's ϕ-index is definitely greater than the I's ϕ-index. Analyzing their citation curves we observed that they have the same number of publications, but F has in general many more citations for any publication than I, especially those in the Hirsh-core. The same reasons can be provide for the comparison between the author H and the author D, in this case we noticed also that D has also more publications.

References

[ACHH] 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

[ADEH99] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999) "Coherent Measure of Risk", *Mathematical Finance*, 9, 203-228.

[BD07] Bornmann, L., Daniel, H. (2007) "What do we know about the h-index?", *Journal of the American Society for Information Science and Technology*, 58(9), 1381–1385.

[CIT] Adler, R., Ewing, J., Taylor, P. (2008) "Citation Statistics", http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf.

[CM09] Cherny, A. and Madan, D. (2009) "New measures for performance evaluation", *Review of Financial Studies*, 22, 2571-2606.

[CMMM] Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011a) "Risk measures: rationality and diversification", *Mathematical Finance*, 21, 743-774.

[DK10] Drapeau, S. and Kupper, M. (2010) "Risk preferences and their robust representation", *preprint*.

[EW06] Eck, N. V., Waltman, L. (2008) "Generalizing the h- and g-indices" *Journal of Informetrics*, 2(4), 263–271.

[FS02] Follmer, H., and Schied, A. (2002) "Convex measures of risk and trading constraints", *Finance and Stochastics*, 6, 429-447.

[FS04] Follmer and Schied (2004) "Stochastic Finance. An Introduction in Discrete Time", 2nd ed., *de Gruyter Studies in Mathematics,* 27.

[FM11] Frittelli, M. and Maggis, M. (2011) "Dual representation of quasiconvex conditional maps", *SIAM J. Financial Math*, Vol 2, pp 357-382

[FR02] Frittelli, M., and Rosazza Gianin, E. (2002) "Putting order in risk measures", *Journal of Banking and Finance*, 26, 1473–1486.

[GR09] Guns, R., Rousseau, R. (2009) "Real and rational variants of the h-index and the g-index", *Journal of Informetrics*, 3, 64–71.

[H05] Hirsch, J. E. (2005) "An index to quantify an individual's scientific research output", *Proceedings of the National Academy of Sciences of the United States of America*, 102 (46), 16569–16572.

[K06] Kosmulski, M. (2006) "A new Hirsch-type index saves time and works equally well as the original h-index", *ISSI Newsletter*, 2(3), 4–6.

[PV90] Penot, J.P. and Volle, M. (1990) "On Quasi-Convex Duality", *Mathematics of Operations Research*, Vol.**15**(4), 597-625 .

[RT08] Ruane, F., and Tol, R. (2008) "Rational (successive) h-indices: An application to economics in the Republic of Ireland", *Scientometrics*, 75(2), 395–405.

[S08] Schreiber, M. (2008), "To share the fame in a fair way, hm for multi-authored manuscripts", *New Journal of Physics*, 10(040201), 1–9.

[Vo98] Volle, M. (1998) “Duality for the level sum of quasiconvex functions and applications”, *Control, Optimisation and Calculus of variations*, Vol.**3**, 329-343.

[W0308] Woeginger, G. J. (Mar 2008) "An axiomatic characterization of the Hirsch-index", *Mathematical Social Sciences,* 56, 224–232.