Further clarifications about the success-index

Brief Communication

Further clarifications about the *Success*-index

Fiorenzo Franceschini[1], Maurizio Galetto, Domenico Maisano, Luca Mastrogiacomo

[1] *fiorenzo.franceschini@polito.it*
Politecnico di Torino, DISPEA (Department of Management and Production Engineering),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

Abstract

The aim of this brief communication is to reply to a letter by Kosmulski (Journal of Informetrics 6(3):368-369, 2012), which criticizes a recent indicator called "*success*-index". The most interesting features of this indicator, presented in [Franceschini et al., to appear in Scientometrics, DOI: 10.1007/s11192-011-0570-z], are: (i) allowing the selection of an "elite" subset from a set of publications and (ii) implementing the field-normalization at the level of an individual publication. We show that the Kosmulski's criticism is unfair and inappropriate, as it is the result of a misinterpretation of the indicator.

Keywords: *success*-index, citation propensity, field normalization, *h*-index.

Reconstruction of the dispute

With this brief communication we reply to a letter by Kosmulski (2012), who criticized the *success*-index, presented in a recent article [Franceschini et al. 2012a]. We anticipate that this criticism is unfair since it was based on a misinterpretation of the indicator. Let's now try to trace the genesis of this dispute.

1. **Kosmulski's indicator (*NSP*)**. In 2011, Kosmulski (2011) presented a novel bibliometric indicator, denominated *Number of Successful Papers* (hereafter abbreviated as *NSP*). Precisely, for a generic group of scientific publications examined—e.g., those associated to a scientist or a journal—the articles that have received more citations than those made are classified as "successful". In other words, a score is associated to each (*i*-th) of the (*P*) publications of interest:

$$\begin{aligned} score_i &= 1 \quad \text{when } c_i > r_i \\ score_i &= 0 \quad \text{otherwise} \end{aligned} \tag{1}$$

where c_i are the citations received and r_i the citations made by the *i*-th publication.

NSP is defined as:

$$NSP = \sum_{i=1}^{P} score_i \ . \tag{2}$$

It can be noted that *NSP*—being an indicator based on the citations accumulated over a non-fixed reference time-window—is time dependent.

According to the authors, *NSP* is very interesting for two reasons: (i) the indicator has a great simplicity and immediate meaning—almost equivalent to those of the *h*-index [Hirsch, 2005]; (ii) the indicator can be applied to groups of publications from different disciplines, as it (potentially) implements a field-normalization at the level of a single publication.

Considering *NSP* from a broader perspective, it can be seen that—given a generic set of publications—this indicator allows to select an "elite" subset. This selection can also be made by other indicators in the literature: e.g., let us consider the *h*-core approach [Hirsch, 2007], the selection by π-indicator [Vinkler, 2011], the *characteristic scores and scales* (CSS) method [Glänzel, 2011] or the ESI's Highly Cited Papers method [ISI Web of Knowledge, 2012]. We remark that, differently from *NSP*, the aforementioned methods require that the set of publications examined are necessarily within the same scientific discipline.

Unfortunately, *NSP* has the serious defect of estimating the citation propensity of a publication in a very fragile way, in terms of statistical significance. In addition, it is prone to manipulation. More details about these specific limitations, can be found in [Franceschini et al. 2012a].

2. **The *success*-index**. In [Franceschini et al. 2012a] we suggested a new indicator—i.e., the *success*-index—which is inspired by *NSP*, but aimed at reducing its limitations. Here is the definition of the *success*-index:

$$\begin{aligned} score_i &= 1 \qquad \text{when } c_i > CT_i \\ score_i &= 0 \qquad \text{otherwise} \end{aligned} \tag{3}$$

$$Success\text{-index} = \sum_{i=1}^{P} score_i \ . \tag{4}$$

It can be noticed that, in Eq. 3, the term r_i of Eq. 1 is replaced by CT_i, i.e., a generic comparison (or normalization) term associated with the *i*-th publication; in other words *CT_i is an estimate of the number of citations that a publication—in a certain scientific context and period of time—should potentially achieve. Note that we have not put any constraint on the definition of the new comparison term, provided that it must be based on a reasonably representative sample of publications, "close" to that one of interest* [Franceschini et al., 2012a, 9th page].

Of course, determining the "(non-)success status" of an individual paper should not be intended as a comprehensive assessment of quality. Nevertheless, this does not mean that counting the number of papers from a set above/below some appropriate citation thresholds could not provide useful information. Also, this is the basic idea of the *highly cited publications* indicator, theorized by Waltman and Van Eck (2012).

We remark that, for any indicator implementing the field-normalization (not necessarily the *success*-index), it is essential to determine an appropriate procedure for constructing the normalization term (CT_i, in the case of the *success*-index). Three are the most critical issues in

doing this, as also described in [Franceschini et al., 2012b]:

- Defining the procedure for selecting the reference sample of publications. Possible approaches are: (i) the selection of papers published by the same journal, (ii) the use of superimposed classifications such as ISI subject categories, (iii) or the implementation of "adaptive" techniques in which the sample is determined considering the "neighbourhood" of the publication(s) of interest—typically consisting of the set of publications citing or being cited by them.
- Deciding whether to consider (i) the distribution of the number of references made or (ii) the distribution of the citations received by the publications of the reference sample.
- Identifying a suitable (central tendency) indicator for obtaining CT_i from the distribution of interest, e.g., mean, median, harmonic mean, percentiles, etc..

These three issues are valid for the construction of a generic field-normalized indicator, not necessarily the *success*-index. The first issue is particularly critical and currently much debated among bibliometricians; the reason is that the sample must be large enough to be statistically representative but, at the same time, should not be "polluted by outsider papers", such as papers from other (sub-)disciplines.

Franceschini et al. (2012a, 9th page) mention—for the mere purpose of example—some simplified procedures for calculating CT_i:

- *r_i, i.e., the number of citations made by the (i-th) publication concerned (case of the Kosmulski's NSP-index);*
- *$(\overline{r}_{JY})_i$ or $(\tilde{r}_{JY})_i$, i.e., the mean or median number of references made by the articles published in the same journal (J) and year (Y) of the (i-th) publication concerned;*
- *$(\overline{c}_{JY})_i$ or $(\tilde{c}_{JY})_i$, i.e., the mean or median number of citations received by the articles published in the same journal (J) and year (Y) of the (i-th) publication concerned;*
- *$(\overline{r}_N)_i$ or $(\tilde{r}_N)_i$, i.e., the mean or median number of references made by a sample of publications representing the "neighbourhood" of the (i-th) publication concerned;*
- *$(\overline{c}_N)_i$ or $(\tilde{c}_N)_i$, i.e., the mean or median number of citations received by a sample of publications representing the "neighbourhood" of the (i-th) publication concerned.*

We remark again that choosing the optimum procedure is still an open question, as reported in [Franceschini et al., 2012a, 9th page]: *the typical issues concerning (1) the sample selection and (2) the choice of a suitable indicator for denoting the propensity to cite remain still open.*

Also, we note that the first of the aforementioned alternative procedures is the one used for the *NSP*-index.

3. **The letter by Kosmulski**. Let's come now to the point. In a recent letter to the editor, Kosmulski (2012, page 368) begins as follows: *Franceschini, Galetto, Maisano, and Mastrogiacomo (in press) defined a new bibliometric index representing the scientific output of a scientist: success-index=$\sum score_i$, where the sum is taken over all publications of a scientist, and $score_i$ = 1 when the number of citations received by the publication i is greater than the median number of citations received by all articles published in the same journal and in the same year*—i.e., $\left(\tilde{c}_{JY}\right)_i$—*and $score_i$ = 0 otherwise* (please, compare this definition—especially the text underlined—with the original one [Franceschini et al., 2012a, 9th page] …any discrepancy?!).

   Then follows a detailed criticism to this specific definition. In a nutshell, Kosmulski explains that estimating the citation propensity of an article by $\left(\tilde{c}_{JY}\right)_i$ leads to penalize the articles published by prestigious journals—i.e., journals with articles of relatively high citation impact—while would favour modestly cited articles published by low impact journals. The concept can be interpreted through a metaphor that we introduce: for one star (*publication*) of moderate shine (*citation impact*) is much easier to stand out in a constellation (*journal*) embracing not very bright stars (*publications of low impact*), than in a constellation with many "blinding" stars (*publications of high impact*).

   Next, using a mocking tone, Kosmulski renames the *success*-index as *modesty*-index, as *it primarily rewards publication of high-impact articles in low-impact journals* [Kosmulski, 2012, page 368].

   Kosmulski's criticism is reasonable because the median number of citations received by the articles of a scientific journal—specialized in a certain (sub-)discipline—provides a rather distorted estimate of the citation propensity, which does not necessarily reflect the citation propensity of the totality of the publications in that (sub-)discipline. It is well known that there is a certain "bias" in the sense that a few prestigious journals tend to include most of the articles of great impact, while less prestigious journals tend to publish articles (almost exclusively) of low impact [Garfield, 1979].

   It is worth recalling that the use of $CT_i = r_i$—as suggested by Kosmulski (2011)—also introduces a bias, although of a different nature.

Further considerations

The real misunderstanding of the Kosmulski's letter is not given by his reflections; instead, it is represented by the underhand alteration of the definition of the *success*-index [Kosmulski, 2012, page 368].

As seen before, the *success*-index is constructed by associating each publication with a comparison

term CT_i, which should represent the citation propensity of a statistically significant sample of homologous publications. Despite the ample variety of options for constructing CT_i (some of which mentioned before) it is stated that: *estimating the propensity to cite by a sample of publications that represent the neighbourhood* […] *seems to be a more "adaptive" and accurate method* [Franceschini et al, 2012a, 2nd page]. Also, we clarify that *the "neighbourhood" of the publication(s) of interest* has been defined as *the set of publications citing or being cited by them* [Franceschini et al, 2012a, 2nd page].

In a recent paper the authors propose a more structured technique for selecting the sample of homologous publications, referring to journal articles [Franceschini et al., 2012c]. From the perspective of an *i*-th paper of interest, this technique is based on the following steps (see Fig. 1):

(a) identification of the *i*-th paper of interest and the corresponding journal (*J*);

(b) identification of other articles published by *J* in the recent years (e.g., the last 5-10 years);

(c) definition of a reference sample consisting of the papers (from the whole scientific literature) that cite the papers identified at point (b).

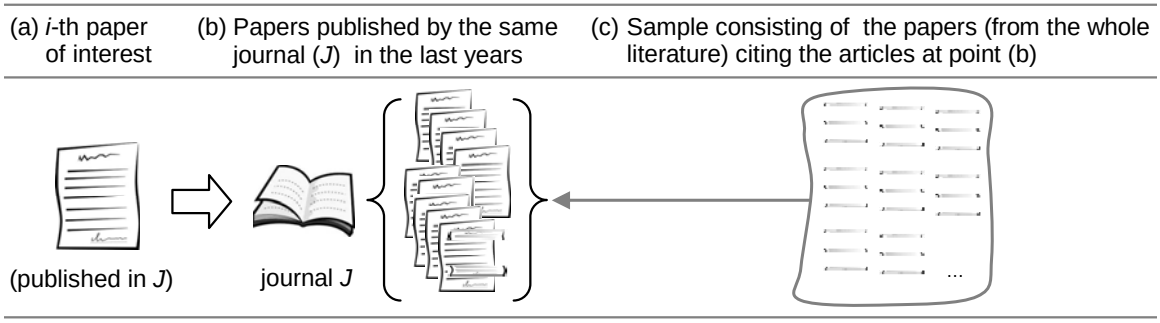

Fig. 1. Scheme of a possible technique for selecting a reference sample of publications, according to which to construct the CT_i related to an *i*-th paper of interest.

This technique is inspired by a procedure by Moed (2010) to determine the *Database Citation Potential* (*DCP*), which is used as a normalization term for the *Source Normalized Impact per Paper* (*SNIP*), i.e., an annual field-normalized indicator for ranking scientific journals.

There are two key assumptions underlying this technique: (1) articles issued by the same journal (*J*) roughly concern the same (sub-)discipline and (2) articles citing other articles issued by *J* are relatively similar as regards their citation propensity. Also, it is necessary to avoid inconsistencies among the different article types within the reference sample (e.g., research articles, reviews, brief communications, letters), due to the different propensity to cite. For instance, this can be done by limiting the analysis to research articles only.

For the purpose of example, Fig. 2 reports a structured comparison among three possible standards for constructing CT_i, i.e., $CT_i^{(1)} = r_i$, $CT_i^{(2)} = \left(\tilde{r}_{JY}\right)_i$ and $CT_i^{(3)} = \left(\tilde{r}_{cJ}\right)_i$—the last standard is founded on the sample selection technique illustrated in Fig. 1. This example should not be intended as an empirical proof of the superiority of the third or the second standard with respect to the first one,

We note that the use of different standards may entail considerable differences in the resulting CT_i values.

(a) Comparison among three standards for calculating CT_i

CT_i	$CT_i^{(1)} = r_i$	$CT_i^{(2)} = \left(\tilde{r}_{JY}\right)_i$	$CT_i^{(3)} = \left(\tilde{r}_{cJ}\right)_i$
Description	Number of references made by the (*i*-th) paper of interest.	Median number of references made by the articles issued in the same journal (*J*) and year (*Y*) of the (*i*-th) publication concerned. Selection should be limited to articles of the same type.	Median number of references made by the articles citing other articles issued (in the last years) by the same journal (*J*) of the (*i*-th) publication concerned (see Fig. 1). Subscript "*cJ*" stands for "citing articles from *J*". Selection should be limited to articles of the same type.
Which citing propensity is estimated?	That of the very single article of interest.	That of the articles issued by journal *J*, in the year *Y*.	That of the articles (from the whole literature) citing other articles issued by *J*.
Computational load of database queries	Low: just count the number of references made by the paper concerned.	Medium: the papers issued by *J* have to be examined.	High: apart from analysing the papers issued by *J*, it is necessary to analyse the papers citing them.
(Potential) drawbacks	Statistically fragile and somehow prone to manipulation by authors.	It may reflect some potential particularities of a journal, instead of (sub-)field characteristics.	-

(b) Example of calculation of CT_i values and the *success*-index, for an anonymous scientist

i-th article of interest	Journ. (*J*)	c_i	$CT_i^{(1)} = r_i$		$CT_i^{(2)} = \left(\tilde{r}_{JY}\right)_i$		$CT_i^{(3)} = \left(\tilde{r}_{cJ}\right)_i$	
1	J1	117	11	✓	27.0	✓	20.0	✓
2	J2	52	9	✓	22.0	✓	16.4	✓
3	J3	21	21	×	29.0	×	24.2	×
4	J4	15	16	×	26.0	×	23.2	×
5	J4	11	10	✓	26.0	×	23.2	×
6	J5	4	6	×	15.0	×	18.2	×
7	J6	1	26	×	25.0	×	22.8	×
8	J7	1	71	×	22.0	×	22.4	×
9	J1	0	3	×	27.0	×	20.0	×
			success$^{(1)}$=*NSP*=3		*success*$^{(2)}$=2		*success*$^{(3)}$=2	

Fig. 2. Comparison among three possible standards to construct the CT_i related to an *i*-th paper of interest: $CT_i^{(1)} = r_i$, $CT_i^{(2)} = \left(\tilde{r}_{JY}\right)_i$, $CT_i^{(3)} = \left(\tilde{r}_{cJ}\right)_i$. Precisely, (a) reports a description of the major peculiarities, while (b) an example of calculation of CT_i values (and next the *success*-index) according to the three standards. The example refers to a portion of the scientific production of an anonymous scientist; ✓ and × respectively denote papers included and not included in the so-called *success*-core [Franceschini et al., 2012a].

It is worth remarking that the technique illustrated in Fig. 1 can be rather complex as regards the amount of database queries. Using the words of one of the referees, the construction of the *success*-index entails a "herculean" effort, much greater than that required for the construction of *NSP*. However, this is the price to pay for estimating CT_i properly. Of course, we are convinced that a practical prerequisite of the procedure is its automation. For this reason, we are currently developing an application able to automatically querying bibliometric databases (e.g., WoS or Scopus).

In the three standards of Fig. 2, it may be also seen that CT_i is constructed based on the citation

made by a reference sample of publications (which is unitary in the first case). Of course, other possible estimates of CT_i can be based on the citations received; for instance, one could use the mean/median number of citations received by the articles of the reference sample. About this, in [Franceschini et al., 2012a, 6th page] the authors state that the indicators based on the distribution of citations made—rather than those received—have several advantages: *(1) the number of citations made (related to a reference sample of publications) is fixed over time, while the number of citations received tends to increase and requires a physiological accumulation period to stabilize—typically, around 3–5 years depending on the disciplines. For this reason, indicators based on the number of references look more stable and robust, especially for relatively recent samples of publications. (2) This stability is also derived by the fact that the number of references is likely to be, on average, less variable than the number of citations received. The estimation will therefore be less subject to fluctuations. (3) Certainly, the citations that a present publication will receive will come from future publications. Therefore, it is somehow questionable to estimate the future propensity to cite by the present one. However, since changes in the propensity to cite generally require a large number of years (hardly less than 10–15 years, the result of this approximation is not very distorted.*

We believe that the information given here is sufficient to demonstrate how inappropriate and unjustified the Kosmulski's criticism is, since it is based on a "far-fetched interpretation" of the *success*-index.

But there's more! In our article two simplified examples of application of the *NSP*- and the *success*-index are presented; the first concerning four scientific journals and the second concerning two scientists from different disciplines [Franceschini et al., 2012a, Subsection *Empirical application examples of the success-index*]. Before presenting the data, the following statement on the calculation of CT_i is made [Franceschini et al., 2012a, 9th page]: *Despite the claimed "freedom" in the construction of CT_i, for the purpose of simplicity and practicality, it will be hereafter calculated as* $\left(\tilde{r}_{JY}\right)_i$. These examples show that the results obtained using the *NSP*- and the *success*-index may be significantly different.

Having said that, the second example reported in the original paper [Franceschini et al., 2012a] is recalled by Kosmulski (2012, Table 1) inappropriately, in order to support his criticism.

The authors thank Kosmulski for his "funny irony(!)" and his contribution to the criticism of one of the possible alternative procedures for the construction of CT_i—i.e., that one based on $\left(\tilde{c}_{JY}\right)_i$. While Franceschini et al. (2012a) focused on the vulnerability of the estimation by $CT_i = r_i$, as proposed by Kosmulski (2011), Kosmulski (2012) focused on the fragility of the estimation by $CT_i = \left(\tilde{c}_{JY}\right)_i$.

According to the authors, the intense debate on the *success*-index's potential and the best strategy for constructing CT_i can lead to interesting ideas for tackling the problem of field-normalization in general.

References

Franceschini, F., Galetto, M., Maisano, D., Mastrogiacomo, L. (2012a). The success-index: an alternative approach to the h-index for evaluating an individual's research output. To appear in Scientometrics. DOI: 10.1007/s11192-011-0570-z.

Franceschini, F., Galetto, M., Maisano, D., Mastrogiacomo, L. (2012b). An informetric model for the *success*-index. Submitted to Journal of Informetrics.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2012c). Evaluating research institutions: the potential of the *success*-index. Submitted to Scientometrics.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? Scientometrics, 1(4), 359–375.

Glänzel, W. (2011). The application of characteristic scores and scales to the evaluation and ranking of scientific journals. Journal of Information Science, 37(1), 40–48.

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

Hirsch, J.E. (2007). Does the *h* index have predictive power? PNAS, 104(49), 19193–19198.

Kosmulski, M. (2011). Successful papers: A new idea in evaluation of scientific output. Journal of Informetrics, 5(3), 481–485.

Kosmulski, M. (2012). Modesty-index. Journal of Informetrics, 6(3), 368–369.

ISI Web of Knowledge (2012). Essential Science Indicators. http://thomsonreuters.com [11 July 2012]

Moed, H.F. (2010). Measuring contextual citation impact of scientific journals. Journal of Informetrics, 4(3), 265–277.

Waltman, L., Van Eck, N.J. (2012). The inconsistency of the *h*-index. Journal of the American Society for Information Science and Technology, 63(2), 406–415.

Vinkler. P. (2011). Application of the distribution of citations among publications in scientometric evaluations. Journal of the American Society for Information Science and Technology, 62(10), 1963–1978.