UNIVERSITÀ DEGLI STUDI DI SIENA

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA

Stefano Vannucci

Dominance Dimension: A Common Parametric Formulation for Authorial Scientific Impact Indexes

n. 561 – Aprile 2009

Abstract - We introduce the dominance dimension principle and the parameterized family of criteria for the assessment of publication/citation profiles it generates. We show that by a suitable choice of parameters dominance dimension may specialize to the most widely known and used of those impact scores for the scientific output of individual authors which disallow indirect reputation effects, including the Durfee- or h-number, the publication number and the citation number.

JEL Classification Numbers: D63, Z00, Z19

Stefano Vannucci, Dipartimento di Economia Politica, Università di Siena

1 Introduction

Authorial impact indexes (henceforth AIIs) are meant to assess the actual influence of the scientific output of individual authors by attaching numerical scores to publication/citation profiles as provided by the relevant citation databases. Number of publications, number of citations, and most recently, Durfee number of the citation-partition (under the label 'h-index': see Hirsch (2005), Anderson, Hankin, Killworth (2008), and Andrews,Eriksson (2004)) have been suggested, and are now the most widely used authorial impacts indexes. Other AIIs have been proposed as amendments of the former (e.g. the so called 'g-index' has been advanced as a putative improvement on the 'h-index': see Egghe (2006)). Apparently, the properties and underlying motivations of the foregoing AIIs are widely diverse. However, it turns out that they can be easily represented as specialized versions of a general principle we label 'dominance dimension' which amounts to a parameterized family of impact score functions. The dominance dimension principle relies on two sequences of integers, namely a suitable sequence of partial scores, and a sequence of corresponding benchmarks. *The dominance dimension of a certain sequence of partial scores with respect to a given sequence of benchmarks records the maximum size of an initial segment of the first sequence whose terms are invariably not smaller than the corresponding benchmarks.*

This note is devoted to a presentation of dominance dimension in the scientific impact assessment domain, and its specialization to current AIIs. In particular, it is shown that the most widely known indexes arise from a remarkably 'small' cluster of values of the relevant parameters for dominance dimension. As a by-product, the dominance dimension principle also provides a recipe for defining new possible AIIs.

2 Dominance Dimension and Authorial Impact Scores

Let $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}} \in \mathbb{Z}_+^{\mathbb{N}}$ two (either nonincreasing or nondecreasing) *monotonic* sequences of nonnegative integers. The *dominance dimension dom* $(x_n)_{n\in\mathbb{N}}|(y_n)_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ w.r.t. $(y_n)_{n\in\mathbb{N}}$ is the maximum $d \in \mathbb{Z}_+$ such that $x_n \geq y_n$ for any $n \leq d$ if $x_1 \geq y_1$ and $x_l < y_l$ for some $l \in \mathbb{N}$, ∞ if $x_n \geq y_n$ for each $n \in \mathbb{N}$, and 0 otherwise. Now, let $\mathbf{c} = (c_n)_{n=1,..,m} \in \mathbb{Z}_+^m$ denote the (passive) citation profile of a certain author (i.e. c_n records the number of citations of article n by other articles in the relevant database), and $\mathbf{c}^* = (c_n^*)_{n\in\mathbb{N}}$ its *nonincreasing ordered extension*, namely $(c_1^*, ..., c_m^*)$ is a permutation of $(c_1, ...c_m)$ such that $c_n^* \geq c_{n+1}^*$ for any $n = 1, ..., m-1$, and $c_n^* = 0$ for any $n \geq m+1$. Next, for any $n \in \mathbb{N}$ consider a (weakly) isotonic function $f_n : \mathbb{Z}_+^n \rightarrow \mathbb{Z}_+$ i.e. $f_n(\mathbf{x}) \geq f_n(\mathbf{y})$ whenever $\mathbf{x} \geq \mathbf{y}$. Finally, consider the power sequence $(n^k)_{n\in\mathbb{N}}$ for some integer $k \in \mathbb{Z}_+$. Then, the *dominance score* of $\mathbf{c}$ *with respect to* $((f_n)_{n\in\mathbb{N}}, k)$ - written $ds_{((f_n)_{n\in\mathbb{N}},k)}(\mathbf{c})$- is *dom* $(f_n(c_1^*, ..., c_n^*))_{n\in\mathbb{N}}|(n^k)_{n\in\mathbb{N}}$ namely the dominance dimension of $(f_n(c_1^*, ..., c_n^*)_{n\in\mathbb{N}})$ with respect to integer power sequence

$(n^k)_{n\in\mathbb{N}}$. Thus, for any $n \in \mathbb{N}$, f_n *works as a partial assessment function* and integer power n^k *as a benchmark*: $f_n(c_1^*, ..., c_n^*)$ provides an assessment of the n-th initial segment of the citation/publication profile's ordered extension to be compared with n^k. The *dominance dimension records the maximum n' such that $f_n(c_1^*, ..., c_n^*)$ meets the standard defined by n^k for any $n \leq n'$.*

The following parameters of citation profile $\mathbf{c} = (c_n)_{n=1,..,m} \in \mathbb{Z}_+^m$ are most typically used (or have been recently proposed) as alternative impact scores:

number of publications $p(\mathbf{c}) = m$

number of cited publications $p^c(\mathbf{c}) = |\{n : c_n \neq 0\}|$

number of citations $c(\mathbf{c}) = \sum_{n=1}^{m} c_n$

Durfee number (or h-score) $h(\mathbf{c}) = |\{n : c_n^* \geq n\}|$

g-score $g(\mathbf{c}) = \max\{n : \sum_{l=1}^{n} c_l^* \geq n^2\}$.

It is easily checked that the following proposition holds true.

Proposition 1 *Let* $\mathbf{c} = (c_n)_{n=1,..,m} \in \mathbb{Z}_+^m$. *Then*

i) $p(\mathbf{c}) = dom\ ((f_n(c_1^*, ..., c_n^*))_{n\in\mathbb{N}}|(n^0)_{n\in\mathbb{N}}$ *with* $f_n = succ(proj_n^n)$ *if* $n \leq m$ *and* $f_n = \mathbf{O}$ *(the zero constant function) if* $n > m$;

ii) $p^c(\mathbf{c}) = dom\ ((proj_n^n(c_1^*, ..., c_n^*))_{n\in\mathbb{N}}|(n^0)_{n\in\mathbb{N}}$;

iii) $c(\mathbf{c}) = dom\ (\sum_{l=1}^{n} c_l^*)_{n\in\mathbb{N}}|(n^1)_{n\in\mathbb{N}}$;

iv) $h(\mathbf{c}) = dom\ (proj_n^n(c_1^*, ..., c_n^*))_{n\in\mathbb{N}}|(n^1)_{n\in\mathbb{N}}$

v) $g(\mathbf{c}) = dom\ (\sum_{l=1}^{n} c_l^*)_{n\in\mathbb{N}}|(n^2)_{n\in\mathbb{N}}$

(where succ denotes the successor function and $proj_i^n$ the i-th n-variable projection or identity function i.e. $succ(z) = z + 1$ *and* $proj_i^n(z_1, ..., z_n) = z_i$ *for any* $z, z_1, ..., z_n \in \mathbb{Z}_+$ *and* $i \in \{1, ..., n\}$*).*

Corollary 2 *Let* $\mathbf{c} = (c_n)_{n=1,..,m} \in \mathbb{Z}_+^m$ *and* $s(.) \in \{p(.), p^c(.), c(.), h(.), g(.)\}$. *Then there exist* $k \in \{0, 1, 2\}$ *and a sequence* $(f_n)_{n\in\mathbb{N}}$ *of* primitive recursive functions *such that* $s(\mathbf{c}) = ds_{((f_n)_{n\in\mathbb{N}},k)}(\mathbf{c})$.

Proof. *Just recall that by definition the class of primitive recursive functions is the smallest class that contains the 'initial functions' (namely the zero constant function, the successor function and the projections) and closed under composition and primitive recursion (see e.g. Odifreddi(1999)). Then, observe that (finite) sums may be easily obtained by (finitely repeated use of) primitive recursion from the successor function and suitably chosen projections.* ■

Remark 3 *Notice that, indeed, the definition of* $p(.), p^c(.), h(.)$ *only involves 'initial' primitive recursive functions.*

Moreover, by positing $(f_n = proj_n^n(c_1^*, ..., c_n^*))_{n\in\mathbb{N}}$ *and* $k = 2$ *the very same dominance dimension principle as defined above induces a somewhat flattened version of the* h-index, *while combining* $(f_n = \sum_{l=1}^{n} c_l^*)_{n\in\mathbb{N}}$ *and* $k = 0$ *induces a two-valued score function awarding value* ∞ *to authors of at least one cited article and value* 0 *to any other author.*

3 Concluding remarks

As mentioned above, the dominance dimension principle may contribute to impact analysis in at least two respects. First, it provides a unified parametric framework for (authorial scientific) impact indexes under which the most commonly used impact scores are generated by choices of parameter values within a remarkably restricted cluster in parameter space. Moreover, it also suggests a general recipe for defining new possible impact indexes e.g. by choosing other sequences of primitive recursive (or recursive) partial assessment functions and/or larger powers as benchmarks.

References

[1] 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.

[2] Andrews, G.E., Eriksson, K. (2004): *Integer Partitions.* Cambridge University Press, Cambridge UK.

[3] Egghe, L. (2006): Theory and practise of the g-index, *Scientometrics,* 69:131-152.

[4] 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: 16569-16572.

[5] Odifreddi, P.G. (1999): *Classical Recursion Theory, vol.1* (Second Edition). Elsevier, Amsterdam.