Journal of Informetrics 11 (2017) 324–335

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

An investigation on the skewness patterns and fractal nature of research productivity distributions at field and discipline level

Giovanni Abramo[a,*], Ciriaco Andrea D'Angelo[b,c], Anastasiia Soldatenkova[b]

[a] Laboratory for Studies in Research Evaluation, Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy, Via dei Taurini 19, 00185 Rome, Italy
[b] Department of Engineering and Management, University of Rome 'Tor Vergata', Via del Politecnico 1, 00133 Rome, Italy
[c] Laboratory for Studies in Research Evaluation, Institute for System Analysis and Computer Science (IASI-CNR), Via del Politecnico 1, 00133 Rome, Italy

ARTICLE INFO

Article history:
Received 19 December 2016
Received in revised form 1 February 2017
Accepted 1 February 2017
Available online 20 February 2017

Keywords:
Bibliometrics
Research evaluation
FSS
Italy
CSS

ABSTRACT

The paper provides an empirical examination of how research productivity distributions differ across scientific fields and disciplines. Productivity is measured using the FSS indicator, which embeds both quantity and impact of output. The population studied consists of over 31,000 scientists in 180 fields (10 aggregate disciplines) of a national research system. The Characteristic Scores and Scale technique is used to investigate the distribution patterns for the different fields and disciplines. Research productivity distributions are found to be asymmetrical at the field level, although the degree of skewness varies substantially among the fields within the aggregate disciplines. We also examine whether the field productivity distributions show a fractal nature, which reveals an exception more than a rule. Differently, for the disciplines, the partitions of the distributions show skewed patterns that are highly similar.

1. Introduction

A number of social phenomena do not show the common normal distributions. Classic examples include the cases of income, wealth and prices, for which most observations are concentrated towards the lower limit, and where distributions show strong skewness with long tail on the right, implying inequality.

Scientific activity is another social phenomenon whose main indicators are widely considered to be unequal in distribution. The literature provides empirical evidence on the subject, particularly through observation of two standard measures of researcher performance: numbers of publications produced and citations to the publications. Studies of skewness in the distribution of citations originate with Seglen (1992), and demonstrate that inequality in impact appears in various disciplines and fields and at different levels of aggregation (among recent works: Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011; Bornmann & Leydesdorff, 2017 in press; Chatterjee, Ghosh, & Chakrabarti, 2016; Franceschet, 2011; Ruiz-Castillo, 2012).

* Corresponding author.
E-mail addresses: giovanni.abramo@uniroma2.it (G. Abramo), dangelo@dii.uniroma2.it (C.A. D'Angelo), anastasiia.soldatenkova@uniroma2.it (A. Soldatenkova).

http://dx.doi.org/10.1016/j.joi.2017.02.001
1751-1577/

Lotka (1926) originally wrote on the frequency distribution of number of publications; since then this metric has generally been considered to show research productivity. Although his study did not emphasize the concept of skewness, "Lotka's Law" has come to imply that most researchers have a small number of published papers. Later research on productivity distribution asymmetry has concentrated on verifying the law in different fields, using data on publication counts. In our view, the most comprehensive investigation into skewness of performance distribution across fields is the one by Ruiz-Castillo and Costas (2014). These authors studied the shape of productivity distributions as measured by number of articles and mean citation per publication. Their field of observation consisted of 17.2 million disambiguated world authors, whose Web of Science (WoS) indexed publications in the period 2003–2011 were classified into 30 broad scientific fields. The main finding is that the distributions are highly skewed and have similar patterns. The analyses for the population as a whole, and for the part above the first mean value, also revealed the fractal nature of the distributions – an issue which we will return to later in the paper. Ruiz-Castillo and Costas took the only approach possible when examining performance distributions at the world level, which is to begin from the WoS indexed publications grouped by field, and from these identify and disambiguate the authors. However, in this paper we exploit a distinctive feature of the Italian university system, which is that every professor is classified into one and only one research field. This allows us to start from the researchers rather than their publications. Consequently, we are able to examine classes of researchers, rather than the examining those who at a given time publish in the different fields. As we next explain, the implications are significant.

Our approach is to begin from the 370 fields (called "Scientific Disciplinary Sectors", SDSs) of the Italian research system, which in turn group the researchers under 14 disciplines ("University Disciplinary Areas", UDAs). Using a disambiguation algorithm developed by D'Angelo et al. (2011), we then associate each professor with his/her WoS publications for the period under study. The approach offers immediate advantages. First, we can spot the unproductive researchers working in a particular field. Second, given that authors can publish in different fields, we are able to measure their real productivity, independent of how they diversify output among fields. To exemplify, in our approach, if a statistician publishes five works on statistical modelling and five on epidemiology, her performance by number of publications is 10. Differently, using any approach based on field classification of output, her performance would only be five as a statistician, while she would also show a performance of five among physicians (which she is not). Furthermore, we use "Fractional Scientific Strength" (FSS) as the indicator of productivity. This indicator embeds both the number of publications and their relative impact (Abramo & D'Angelo, 2014), thus addressing the weaknesses of performance indicators that rely on number of publications alone, or on mean citations per publication. We have examined the problems of such indicators in two specific works, published in this same journal (Abramo & D'Angelo, 2016a, 2016b).

The literature provides very broad evidence of skewness in research productivity, whether measured by quantity or impact. Given that FSS embeds both, we expect to find distributions of the same manner. We analyze the frequency distributions for productivity at the field and discipline levels, using the dataset of all Italian professors in the period 2009–2013. The aim of the paper is twofold. First, we intend to provide national and global readers with benchmarks of the yearly average productivity distribution in each field. Next, and more immediately interesting, we wish to investigate the between-field variation of skewness of productivity distributions and their fractal nature. More specifically, we try to answer the following questions:

- Is productivity distribution highly skewed in every field?
- Do the different fields within a discipline maintain similar patterns in productivity distribution?
- Are the distributions of a fractal nature, with the same shape in upper tails?
- Do productivity distributions at the discipline level preserve the shape characteristics of the fields? Are the different disciplines similar?

Throughout the paper we account for the fact that data collection and calculation of the FSS indicator can be difficult for some. For this, we also provide field distributions by number of publications alone (found in the Supplementary Material), and repeat several steps of the analysis using these.

In the next section of the paper we describe our data sources, indicators and the methodology used for the analyses. Sections 3 and 4 present the results and our conclusions.

2. Data and method

Data on Italian academics and their SDS classifications are extracted from the database of the Ministry of Education, Universities and Research (MIUR).[1]

[1] http://cercauniversita.cineca.it/php5/docenti/cerca.php, last accessed on January 23, 2017.