Available online at www.sciencedirect.com
ScienceDirect
Journal of Informetrics 1 (2007) 185–192
Journal of INFORMETRICS
An International Journal
www.elsevier.com/locate/joi

Scoring research output using statistical quantile plotting

Jan Beirlant [a], Wolfgang Glänzel [b], An Carbonez [c,*], Herlinde Leemans [d]

[a] *Department of Mathematics and University Center of Statistics, K.U. Leuven, Belgium*
[b] *Faculty of Economics and Steunpunt O&O Statistieken, K.U. Leuven, Belgium*
[c] *University Center of Statistics, K.U. Leuven, Belgium*
[d] *Dienst Onderzoekscoordinatie, K.U. Leuven, Belgium*

Received 21 November 2006; received in revised form 20 April 2007; accepted 23 April 2007

Abstract

In this paper, we propose two methods for scoring scientific output based on statistical quantile plotting. First, a rescaling of journal impact factors for scoring scientific output on a macro level is proposed. It is based on normal quantile plotting which allows to transform impact data over several subject categories to a standardized distribution. This can be used in comparing scientific output of larger entities such as departments working in quite different areas of research. Next, as an alternative to the Hirsch index [Hirsch, J.E. (2005). An index to quantify an individuals scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572], the extreme value index is proposed as an indicator for assessment of the research performance of individual scientists. In case of Lotkaian–Zipf–Pareto behaviour of citation counts of an individual, the extreme value index can be interpreted as the slope in a Pareto–Zipf quantile plot. This index, in contrast to the Hirsch index, is not influenced by the number of publications but stresses the decay of the statistical tail of citation counts. It appears to be much less sensitive to the science field than the Hirsch index.

Keywords: Quantile plots; Standardizing; Normal quantile plot; Pareto quantile plot; Extreme value index

1. Introduction

Rating and comparing research groups on a macro level based on some kind of quality indicator concerning the journals in which they publish is regularly used in university policy decision-making. Such kinds of methods are often criticized, especially because the indicators are strongly dependent on the field of research. This is especially the case when using impact factors. Limitations of impact factors are, for instance, discussed in Moed (2002), Glänzel and Moed (2002) and Podlubny (2005). A simple rating measure of research groups such as on departmental level, could be the count of papers in those journals which are situated within the top 10% of a (S)SCI subject category. Such a scoring method could be a part of a management action to stimulate publication in top journals rather than stimulating a higher number of publications. This top 10% scoring is in many ways unsatisfactory due to the volatility of a journal impact factor over the years together with the problems linked to the composition of a subject category. Here we present a scoring technique, to be used only in macro level comparisons, characterized by standardization between different

* Corresponding author. Tel.: +32 16 32 22 42; fax: +32 16 32 28 31.
E-mail address: An.Carbonez@ucs.kuleuven.be (A. Carbonez).

1751-1577/$ – see front matter
doi:10.1016/j.joi.2007.04.002

subject categories, and where possible negative effects due to the composition of a subject category and the volatility of impact factors are somehow relaxed due to the continuity in the score (in contrast to a top 10% rule where 0–1 scores are attributed to journals). Of course this method would also benefit from its use on homogeneous composition of subject categories on which it is based.

Next, we also present a scoring method for individual researchers. Here recently the Hirsch index constitutes a recent proposal (Hirsch, 2005). However, it turns out to be quite dependent on the field of research and the number of papers a researcher has published, rather than scoring the intrinsic interest of the research community in the contents of the papers. In case of Lotkaian–Zipf–Pareto behaviour of citation counts of an individual, the Pareto tail index can be interpreted as the slope in a Pareto–Zipf quantile plot. More generally, one could use the extreme value index of the statistical distribution of the citations of an individual. This concept generalizes the Pareto tail index in cases of non-Lotkaian (or non-Zipf/Pareto) behaviour.

We start by considering a set of impact factors from a list of journals as a realization of a random variable. The same can be done for the number of citations of the different papers of an individual at a given moment. Both solutions are based on the concept of quantile plots. The use of quantile plots in informetrics investigations is illustrated in Huber (1998). In a quantile–quantile plot (QQ plot), the quantiles of the data are plotted against the quantiles of a specific distribution (assuming that the data follow the specific distribution). The basic idea is that, if the data follow that specific distribution, then the graph will essentially be a straight line. See Gilchrist (2000) for a comprehensive reference on QQ plots. Many statistical packages (including Excel) have quantile functions available for many standard distributions.

The statistical distribution of impact factors of a set of journals can be well approximated by a lognormal distribution. This assertion can be verified using a normal QQ plot or a goodness-of-fit test. Let us remind that a quantile function taken at a value $p \in (0,1)$ yields the $100p$-percentile of the corresponding distribution. The ith smallest observation $X_{i,n}$ ($1 \le i \le n$) from a sample of size n can be considered as the $100(i/n)$-percentile of the data distribution. Given a sample of size n (say impact factors of n scientific journals), the normal QQ plot is then defined by

$$\left(\Phi^{-1}\left(\frac{i}{n+1}\right), X_{i,n}\right), \quad i = 1, \ldots, n \tag{1}$$

where Φ^{-1} denotes the standard normal quantile function. Here one typically applies a continuity correction $i/(n+1)$ to the fraction i/n. Another possible choice is $(i-0.5)/n$. Because any normal random variable X with mean μ and variance σ^2 possesses the same distribution as $\mu + \sigma Z$ where Z is standard normally distributed, one has that a normal QQ plot is approximately linear in case the data constitute a normal random sample. More importantly, this plot defines a transformation to standard normality. Indeed if the normal QQ plot for instance is of exponential type then a logarithmic transformation of the data leads to a linear QQ plot, so that then the original data are lognormally distributed.

This can approximately be observed in case of the *JCR Science Edition 2005* impact factors for the subject categories *Statistics and probability* (with $n = 81$ journals), *Biochemistry and molecular biology* (with $n = 356$ journals), and *Medicine, research and experimental* (with $n = 89$ journals). In Fig. 1 the normal QQ plots of the original impact factors and log-transformed impact factors of these two subject categories are presented, together with a histogram representation of the data. Remark the differences between these two distributions: the maximum impact factors, respectively, are 8.4, 74.4, and 40.2, while the averages, respectively, are 1.5, 6.2 and 4.5.

In Section 2, we will present a scoring method transforming the different distributions towards the same lognormal distribution.

Section 3 deals with the problem of scoring the publication output of individual researchers. Here the distributional family that could serve as a reference is the Lotka–Zipf–Pareto distribution. Here we can refer to Glänzel (2006) and Egghe and Rousseau (2006). The strict Pareto distribution is defined by

$$P(X > x) = \left(\frac{x}{t}\right)^{-\alpha}, \quad x \ge t, \tag{2}$$

with $\alpha > 0$ and $t > 0$. As is shown in Beirlant, Goegebeur, Segers and Teugels (2004), this distribution is characterized by the fact that a log-transformed Pareto random variable Y is exponentially distributed with survival function $P(Y > y) = \mathrm{e}^{-\alpha(y - \log t)}$ ($y > \log t$) and quantile function $Q(p) = \log t - (1/\alpha)\log(1-p)$ ($p \in (0,1)$), such that a Pareto QQ