



ELSEVIER

Contents lists available at ScienceDirect

Journal of Informetrics

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

Correspondence

Quantity matters, but how does it work?

A comment on Lindahl



1. Introduction

In a recent paper (Sandström & van den Besselaar, 2016), we argued that publishing many papers is important as there is a significant positive relation between quantity (the number of papers: P) and quality (the number of top cited papers: $P_{10\%}$). Jonas Lindahl (2018) takes the analysis a step further by including an additional variable: the number of papers in journals with a high impact factor (P_{hif}). He also uses the number of co-authors (CA). He then does two analyses on a set of young researchers in mathematics:

- Using data over the first eight years of their career, do P , P_{hif} and CA explain excellence? (defined as $P_{10\%} \geq 2$)
- Using the same variables P , P_{hif} and CA but now over the first four years, can we *predict* excellence in the second four years' period? (now defined as $P_{10\%} \geq 1$ as the period is shorter)

Using a method called dominance analysis (Azen & Budescu, 2003), Lindahl then argues that not productivity (P) is the best predictor of excellence, but publications in top journals (P_{hif}). In other words, our finding on quantity and quality is challenged, and Lindahl argues that publishing many papers is much less important than publishing papers in top journals. We argue that his conclusions are too strong due to various methodological concerns. We redo the analysis, using Lindahl's data but with a different method. We show that the effect of high impact journals is partly a mediation effect, and that the R^2 of the relations is far too low to do predictions.

2. Some methodological and theoretical issues

2.1. Dominance analysis?

Lindahl mentions rather casually that without high productivity, one cannot have many papers in top journals, but unfortunately he does not elaborate on the consequences. It in fact points at an underlying path-analytic model; but dominance analysis (DA) *cannot* be used in those cases: "If one is interested in the effect of each predictor as it is added to the previously entered predictors, it is possible to perform a sequence of constrained DAs, but DA is not designed to generate or address one specific hierarchical order. DA was also not designed to address questions regarding path-analytic models; for example, DA would not be of help in studying a specific causal model in which X_1 predicts X_2 and X_2 predicts Y , nor would DA be appropriate for other path-analytic models such as mediation or indirect effect models" (Azen & Budescu, 2003). Mediation analysis is what we will do below.

2.2. Explaining and predicting

In his paper, Lindahl suggests two approaches, a synchronic one where the variables are measured over an eight years period, and a predictive model where the independent variables are measured over a four years period and used to predict the dependent variable (excellence) in the next four years period. Lindahl did so as in his understanding the second analysis may lead to indicators that can be used to predict (and select), and to support decision making. This, of course, depends on how strong the prediction is. As the correlations are not extremely high, the indicator when used for selection will lead to large numbers of false positives and false negatives.

<https://doi.org/10.1016/j.joi.2018.08.007>

1751-1577/© 2018 Elsevier Ltd. All rights reserved.

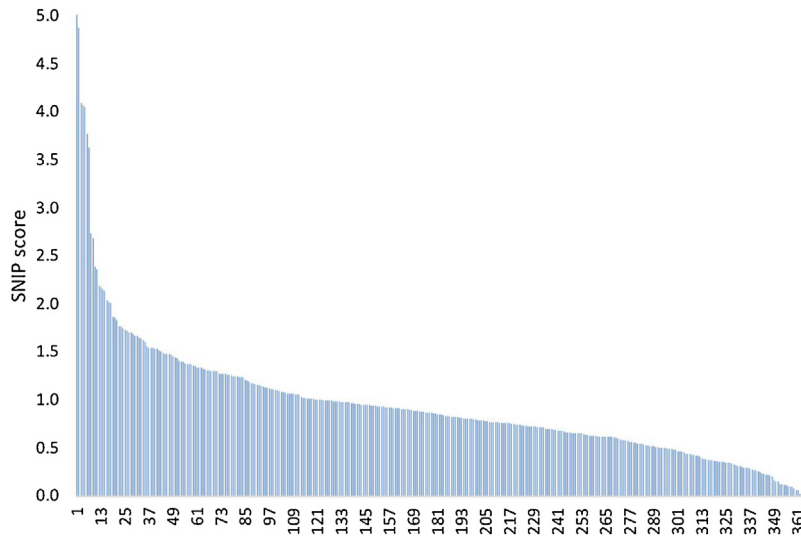


Fig. 1. Mathematics journals ordered by impact (SNIP).

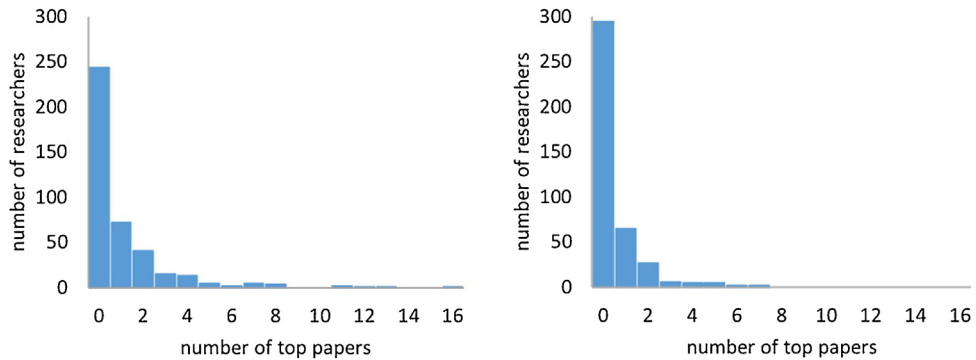


Fig. 2. Frequency distribution top papers synchronic model (left) and predictive model (right).

2.3. Full or fractional counts?

Normally, one would use fractional counting of output, as authors with many coauthors can *ceteris paribus* be more productive. Lindahl's publication-based variables are all integers, which shows that full counts are used, and consequently the analysis should be considered highly problematic. Using the co-author variable may solve this to some extent.

However, another issue comes in here too: the sample consists of young researchers, and the citation scores and the entrance to highly cited 'top' journals may depend on the productivity level of the *co-authors*. In that case, one does not predict excellence of young researchers, but the level of the more senior co-authors, and the quality of the young researcher's network.

2.4. What are top journals?

Lindahl takes the top 25% journals with the highest normalized impact factor (SNIP) for this, but there is no good reason to do so. The top becomes very large (as these journals contain more than 25% of the papers) and there is no statistical argument given the shape of the SNIP curve (Fig. 1). If there is a point to dichotomize, this is around journal number 25, and not at 25% (which is around journal number 90).

2.5. What is excellence?

Lindahl uses the number of top cited papers as indicator for excellence, which we would agree on. However, he then dichotomizes the variable $P_{10\%}$ in the following way. In the synchronous analysis, excellence is defined as $P_{10\%} \geq 2$; not excellent is defined as $P_{10\%} = 0$ or 1. In the 'predictive analysis', excellence is defined as $P_{10\%} \geq 1$ or more; not excellent is

Download English Version:

<https://daneshyari.com/en/article/10139627>

Download Persian Version:

<https://daneshyari.com/article/10139627>

[Daneshyari.com](https://daneshyari.com)