

Measuring the Citation Levels of Subfields as Delineated by Keywords: Investigating Economics Articles in Scopus

Jonathan M. Levitt¹ and Mike Thelwall²

¹*j.levitt@lboro.ac.uk*

Department of Information Science, Loughborough University, Loughborough, LE11 3TU, UK and Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.

²*m.thelwall@wlv.ac.uk*

Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.

Introduction

The intuition underlying this research is that scientific article keywords might be useful to help assess the growth of research topics over time, to investigate the emergence of new research topics, and to delineate and measure the citation levels of subfields. Keywords have been used to assess the growth of research topics over time (Alvarez & Pulgarin, 1996) and to delineate fields and investigate the emergence of new research topics (Hellsten, Lambiotte, Scharnhorst & Ausloos, 2007), although it seems that no word-based approach can be fully effective for such tasks (e.g., Leydesdorff, 1997). The recently formed Scopus database facilitates the investigation of keywords, in that it allows search results to be refined according to keyword. The research-in-progress described here measures the citation levels of keyword-delineated ‘subfields’ of economics.

Methods

This study investigates the growth of the most frequently occurring keywords in articles in the Scopus subject category of ‘economics’ published between 1995 and 2004. For every even year in that period a list of the 160 most frequently occurring keywords was obtained and the 63 keywords present in each of these lists were identified. This study focuses on the 31 keywords we regard to be directly related to economics. The citation level of a keyword is defined to be the Normalised h-index (Levitt & Thelwall, 2007 and 2008) of all articles to which the keyword has been assigned. The Normalised h-index (h_{norm}) is used instead of the h-index, as it adjusts for sample size and is proportional to the square of the h-index (h); for a set of n documents:

$$h_{norm} = 100h^2/n$$

h_{norm} was used rather than average citation, as the latter data is not readily available on Scopus and Levitt and Thelwall (2008) found a .96 Pearson correlation between h_{norm} and average citation

(indicating that h_{norm} is a close indicator of average citation).

Findings

For each of the 31 economics-related keywords, the Normalised h-index for 1999-2000 was expressed as a multiple of the median Normalised h-index for 1999-2000. The five Normalised h-indexes with the highest ratio to this median, representing apparently the most ‘cited’ keywords, are: Socioeconomics, 2.37; Socioeconomic Factors, 2.00; Health care cost, 1.68; Poverty, 1.60 and Income, 1.48. The five lowest multiples are: Transitional economy, .54; Inflation, .47; Exchange rate, .47; Trade policy, .44; and National economy, .44.

For each economics-related keyword the Normalised h-index was calculated for the periods 1995–1996, 1999–2000, 2003–2004 and the correlations between the Normalised h-indices are presented in Table 1. The number of articles associated with the keywords ranges from 34 to 259 (median 73) for 1995–1996, from 81 to 274 (median 132) for 1999–2000, and from 95 to 396 (median 181) for 2003–2004.

Table 1: Spearman correlations between the Normalised h-indices of three two-year periods, for the 31 economics-related keywords ($p < .01$ throughout)

Two-year period	1995–96	1999–2000	2003–04
1995–96	1	.69	.58
1999–00	.69	1	.73
2003–04	.58	.73	1

Although the Normalised h-indices correlate strongly with each other over time (Table 1), the behaviour of the keywords is not uniform. In order to demonstrate this, for every keyword the ratio of h_{norm} for 2003–2004 to h_{norm} for 1995–1996 was calculated and expressed as a multiple of the median ratio. The five keywords with the highest multiples (i.e., the largest relative increase in h_{norm})

are: Socioeconomics, 2.30; Socioeconomic Factors, 2.26; Economic development, 2.14; Industrial economics, 2.09 and Exchange rate, 1.56. The five keywords with the lowest multiples are; National economy, .69; Competition, .68; Productivity, .67; Cointegration, .62 and Human capital, .50. These findings may reflect changes in the impact of subfields.

The 32 keywords regarded as not being related to economics fall into four categories: (1) keywords relating to methodology (e.g., Major clinical study, Mathematical models), (2) keywords relating to country or region (e.g., European Union, United States), (3) keywords relating to age or gender (e.g., Child, Female), and other miscellaneous keywords (e.g., Decision making, Education). As for the economics-related keywords, there were also large variations in citation levels for this set. For example: (a) Compared with the median, the Normalised h-index for 'Mathematical models' in 2003–2004 was a multiple of 2.52 times that of 1995–1996, whereas for 'Theoretical study' the multiple was only .77, and (b) Compared with the median, the Normalised h-index for 'Canada' in 2003–2004 was a multiple of 1.75 times that of 1995–1996, whereas for 'United States' the multiple was only .48.

Conclusions and Limitations

The findings indicate substantial variations in the citation levels of subfields, as delineated by keywords. The strong correlations between the citation levels of keywords for the periods 1995–1996, 1999–2000 and 2003–2004 indicates considerable stability in the overall citation levels of keywords. However, despite this stability, the citation level of some keywords relative to the median was substantially higher in 2003–2004 than in 1995–1996, whereas for other keywords it was substantially lower. The large differences in the change in citation level between the 'Mathematical models' and 'Theoretical study' and between 'Canada' and 'United States' indicate that the method could produce interesting findings on keywords associated with general topics, such as methodologies and countries.

The results also indicate that Scopus' option of refining search results according to keyword can

provide interesting findings on the citation levels of subfields, as delineated by keywords. More specifically, this facility may potentially be used to assess the growth of research topics over time, and to investigate the emergence of new research topics, although we have not assessed the face validity of the results for these tasks. In particular, the delineation of subfields by keyword is an oversimplification. Another limitation is that some keywords have a high percentage of articles in a single journal; for four of the 31 economics-related keywords (Socioeconomic Factors, Socioeconomics, Costs, and Health care cost) more than a third of the articles published in 1995–2004 associated with the keyword were published in a single journal.

Acknowledgements

This research and its presentation are supported by the Economic and Social Research Council [grant number PTA-026-27-2228].

References

- Alvarez, P. & Pulgarin, A. (1996). The Rasch model. Measuring information from keywords: The diabetes field. *Journal of the American Society for Information Science and Technology*, 47(6), 468–476.
- Hellsten, I., Lambiotte, R., Scharnhorst, A. & Ausloos, M. (2007). Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics*, 72(3), 469–486.
- Levitt J.M. & Thelwall M. (2008). Is multidisciplinary research more highly cited? A macro-level study. *Journal of the American Society for Information Science and Technology*, 59(2), 1973–1984.
- Levitt, J.M. & Thelwall, M. (2007). Two new indicators derived from the h-index for comparing citation impact: Hirsch frequencies and the Normalised Hirsch Index. *11th International Conference of the International Society for Scientometrics and Informetrics, Proceedings*, 876–877.
- Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science*, 48(5), 418–427.