Journal of Informetrics 12 (2018) 330–343

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Effect of publication month on citation impact

Paul Donner

German Centre for Higher Education Research and Science Studies, DZHW7 Schützenstraße 6a, 10117 Berlin, Germany

ARTICLE INFO

Article history:
Received 5 November 2017
Received in revised form 25 January 2018
Accepted 26 January 2018

Keywords:
Citation impact
Publication date
Citation analysis
Bias
Bibliometric indicators
Self-citations
Control variable
Systematic error
Validity

ABSTRACT

A standard procedure in citation analysis is that all papers published in one year are assessed at the same later point in time, implicitly treating all publications as if they were published at the exact same date. This leads to systematic bias in favor of early-months publications and against late-months publications. This contribution analyses the size of this distortion on a large body of publications from all disciplines over citation windows of up to 15 years. It is found that early-month publications enjoy a substantial citation advantage, which arises from citations received in the first three years after publication. While the advantage is stronger for author self-citations as opposed to citations from others, it cannot be eliminated by excluding self-citations. The bias decreases only slowly over longer citation windows due to the continuing influence of the earlier years' citations. Because of the substantial extent and long persistence of the distortions, it would be useful to remove or control for this bias in research and evaluation studies which use citation data. It is demonstrated that this can be achieved by using the newly introduced concept of month-based citation windows.

1. Introduction

Citation impact normalization is a central concept for the construction of advanced bibliometric indicators which eliminate the effects of different scientific discipline, type of document and date of publication (Waltman, 2016). By delineating sets of publications that are similar to each other in content and formal characteristics and by using these sets to compute reference values and by computing relative impact indicators based on these reference values, the heterogeneity in citation counts due to these factors is removed. The intention is to make possible fair comparisons, to compare like with like (Schubert & Braun, 1986). The basic formal characteristics, as opposed to the content (disciplinary area), that are taken into account are document type (such as research articles, review papers, letters, editorials etc.) and publication date. Further characteristics have also been demonstrated to co-vary with citation counts, for example differences between methodological, theoretical and empirical works (Peritz, 1983), clinical vs. basic research in medicine (Van Eck, Waltman, van Raan, Klautz, & Peul, 2013) or clinical study level (e.g. Bhandari et al., 2007).

One important component of normalization is controlling for publication date, as, ceteris paribus, the more time has passed since publication the more papers will be published whose authors had the opportunity to read and cite a given publication. The publication year is commonly used to operationalize publication date. This practice is based on the implicit assumption that, for the question of interest of a study, it makes no difference when exactly in a year a paper was published. The fact that, when citations are counted at some later date, documents published in January have eleven months more to be read and cited than works published in December of the same years raises the question if the above assumption is justified, and, if it is not, under which conditions and how a more precise publication date ought to be used in citation analysis.

E-mail address: donner@dzhw.eu

https://doi.org/10.1016/j.joi.2018.01.012
1751-1577/

The question of the influence of a more exact publication date is related to the problem of choosing adequate citation windows, the period in which citations to papers in a set of publications are counted. A citation window that is very short, say two years, would more obviously lead to bias against papers published towards the end of the investigation period compared to those published towards the beginning. Consider the following simple illustration. Citations are counted at the end of the year after publication (2-year citation window). Then papers from January had 24 months to be read and cited, assuming the case they were published on the first day of the month and citations counted after the last day of the citation window, while December papers had 11 months less, just 13 months, which is 54% of the time period of the January papers. This relative disadvantage becomes smaller as the citation window length is increased. In a five year citation window, for example, the December papers had 82% of the citation duration of the January papers. Citations do not accumulate uniformly over time and one is not only concerned with January and December papers, so this reckoning does not say much about the actual size of the distortion. But it might serve as a first order approximation model. Just how big this 'head-start' effect is in reality and at what point in time it vanishes is the topic investigated in this paper.

The article is organized as follows. In the next section, previous work on the topic is briefly reviewed and some knowledge gaps are pointed out which this study addresses. Next, the data on which the study is based are presented. The major part of this contribution is comprised of the analysis of the results regarding the month effect from several points of view, including its size as reflected in basic citation scores and in regression analysis, also taking into consideration the online publication date, the change of the effect size over longer citation windows and its presence and patterns across disciplines. Furthermore, we introduce a method to eliminate the month bias and use the resulting corrected citation counts to demonstrate the bias on a simulated academic impact assessment of institutions under realistic conditions similar to currently employed research evaluation procedures. We finish with a discussion of the results and their implications for the field.

2. Related work

There have been a number of prior studies that have noted and investigated the month effect. They will be briefly reviewed in the following and their results used as a point of departure for this study.

Haslam et al. (2008) used publication month as a control variable throughout their regression analyses of influence factors of citation impact in a psychology sub-discipline. Their criterion was the natural logarithm of articles' citations counted after ten years. In their results, the standardized regression coefficients for month are small, not statistically significant and of the expected negative direction in 3 out of 4 domain regressions and also negative in the final overall regression equation. The small number of publications studied (308) and long citation window might account for the inconclusive findings with regard to publication month influence.

Levitt and Thelwall (2011) proposed a new impact indicator based on a paper's early citation count and the impact factor of its journal and investigated the compound indicator's predictive validity. One of their research questions relates to the correlation between the proposed indicator and publication month and the correlation between raw citation counts and publication month. Their sample is comprised of UK articles in the Social Sciences Citation Index category 'Economics', published between 2001 and 2004 and they counted citations up to 2008. As an aside, the authors note that "the notional publication date of an article does not necessarily reflect the true publication date; frequently a journal issue appears many months before or after its notional publication date" (Levitt & Thelwall, 2011, p. 303). This applies to the present study as well and should be kept in mind as one of its limitations. The authors confirm the existence of the month bias of citation counts for their sample, finding advantages for early-month publications for up to 3 years after publication. The authors conclude that "it seems important to take into account month of publication when analyzing short citation windows" (Levitt & Thelwall, 2011, p. 307).

In a study of citation window lengths, Wang (2013) considered publication month as one of several factors possibly influencing citation ageing. Due to the design of the study, papers uncited after 31 years are excluded from the analysis. Wang also introduces a method for the estimation of publication month when this data is missing. The present study adopts this method, as will be detailed further below. The Wang (2013) study finds no discernible difference in citation ageing across months based on the first and last year after publication in which half or more of the publications are cited. While there is no month difference in ageing according to this measure, there are differences between research fields, document types and papers of different final citation count tiers. Another measure, 'Citation Speed', is also applied, which is based on the cumulative citation counts throughout the observation years. ANOVA analysis of this measure's results does indicate significant differences in citation ageing by publication month, though the effect appears to be small.

Gai, Liu, Zhang, and Liu (2015) study the month citation bias as such in one field, one specialty and one journal, respectively, for citation windows from 1 to 9 years. They exclude a few very highly cited papers. They group publication months into six two-month periods. For papers in physical geography they find significant group differences for citation counts for up to five years after publication. The results are similar for papers on the topic of diabetes. For the Journal of Biological Chemistry, group differences are significant up to the seventh year after publication.

The present study is intended to improve upon the outlined results in a number of ways. As previous studies have been limited to specific disciplines we investigate if the effect is general to all disciplines (as far as covered in Web of Science). Neither uncited nor very highly cited publications are excluded from analysis. A range of disciplines with very different citation dynamics are investigated to provide insight into field differences. Field differences in publication month effect on citation counts can be hypothesized to mirror fields' size in terms of number of papers, their citation density (average