Examining national citation impact by comparing developments in a fixed and a dynamic journal set

Jesper W. Schneider[1], Thed van Leeuwen[2], Martijn Visser[2] & Kaare Aagaard[1]

[1]Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Denmark

[2]Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands

Corresponding author: Jesper W. Schneider, jws@ps.au.dk

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

Abstract

In order to examine potential effects of methodological choices influencing developments in relative citation scores for countries, a fixed journal set comprising of 3232 journals continuously indexed in the Web of Science from 1981 to 2014 is constructed. From this restricted set, a citation database depicting the citing relations between the journal publications is formed and relative citation scores based on full and fractional counting are calculated for the whole period. Previous longitudinal studies of citation impact show stable rankings between countries. To examine such findings coming from a dynamic set of journals for potential "database effects", we compare them to our fixed set. We find that relative developments in impact scores, country profiles and rankings are both very stable and very similar within and between the two journal sets as well as counting methods. We do see a small "inflation factor" as citation scores generally are somewhat lower for high-performing countries in the fixed set compared to the dynamic set. Consequently, using an ever-decreasing set of journals compared to the dynamic set, we are still able to reproduce accurately the developments in impact scores and the rankings between the countries found in the dynamic set. Hence, potential effects of methodological choices seem to be of limited importance compared to the stability of citation networks.

Introduction

The measurement and monitoring of national research performance is important, but has been a methodological challenge in the field of scientometrics since the inception of the citation indices in the 1960s. Numerous studies have presented claims of rises or declines in national research performances but the underlying data and methods used have been characterised by instability and vast "degrees of freedom" for investigators leading to disputable and contradictory results (e.g., Martin et al., 1987; Leydesdorff, 1988; Leydesdorff & Wagner, 2009; Leydesdorff, 2017; Martin, 2017). It is claimed that analyses, where changes over time for systems of interest, are examined require, at least in principle, a time invariant "basket of things measured" (Mirowski, 2011, p. 269). If this is unattainable, variations should as a minimum be accounted for in order for measurements to be interpretable.

Establishing a historical baseline seems therefore to be a basic condition for time series analyses in order to make comparisons conceivable; but this is a rare condition in scientometric analyses. The two commercial citation databases most often used for analysing developments in the scientific literature, *Web of Science* (WoS) and *Scopus*, do not provide such relatively stable reference sets. Both databases are characterised as being dynamic in their inclusion and exclusion of journals to their perceived "core" set. Initially, Garfield and Sher at the *Institute for Scientific Information* (ISI), developed the Journal Impact Factor (JIF) to guide their selection of journals to a relatively stable core set of roughly 3500 journals in the *Science Citation Index* (SCI) (Garfield, 1955; Garfield & Sher, 1963). Later, when ISI was acquired by *Thomson Reuters* in the early 1990s, more opaque decision criteria for journal in- and exclusions were introduced (Mirowski, 2010). Eventually after the appearance of Elsevier's *Scopus* database, being a competitor to WoS, in the mid-2000s, commercial interests seem to have been the main criteria for journal inclusions for both citation databases resulting in prodigious extensions (Testa, 2011).

The use of a historical baseline for longitudinal comparisons of publication activity has been discussed from time to time in the literature (e.g., Narin, 1976; Studer & Chubin, 1980). Narin and his company *Computer Horizons Inc.* (CHI) introduced an *ex ante* fixed journal set from SCI and used it for longitudinal comparisons in the early *Science and Engineering Indicator* reports commissioned by the *US National Science Board* arguing that "working with a changing data set would require much

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

additional statistical manipulation" (1976, p. 138). When Narin proposed to work with such fixed journal sets, the SCI was a relatively stable database and he estimated that for the period 1965 to 1973 the fixed journal set actually covered 89.5% of the publications compared to the dynamic journal set constituting the whole SCI database (Narin, 1976, p. 138). Hence, at the time the discrepancy in coverage between the two journal sets was small.

A claim by Martin and colleagues (1987) that British science was in decline generated a prolonged and heated methodological debate in the scientometric community in the late 1980s and early 1990s (e.g., Leydesdorff, 1988; Narin et al., 1988; Braun, Glänzel, & Schubert, 1989, 1991; Leydesdorff, 1991; Martin, 1991, 1994)[1]. Central to the debate was Martin and colleagues' use of the CHI fixed journal set to measure and monitor the publication output of the United Kingdom (UK). According to critics, the fixed set approach was untenable because analyses done on the corresponding dynamic journal set did not verify the general declining trends for British publication output, leading the critics to claim that Martin and colleagues' findings were an artefact of their chosen methodology (Leydesdorff, 1988). The choice of a fixed set was in fact only one of several debatable parameter choices discussed. Braun et al. (1989) identified 28 such parameters, where the most conspicuous besides choice of database was Martin and colleagues' use of fractional counting at the country level to depict relative shares of output (Leydesdorff, 1988). Leydesdorff (1988) argued that we can assume that authors in what he call "advanced" countries such as the UK will have more prolific publication activity in new journals associated with newly developed fields of science. According to Leydesdorff (1988, 2002), such presumed dynamic and innovative publication activities are penalized when using a fixed set approach. Recent findings from Cimini, Gabrielli and Sylos Labini (2014) seem to support the claim that "advanced" countries have considerably broader research profiles compared to "less-advanced" countries, albeit such claims are always conditional on the underlying publication classification system used. Obviously, Leydesdorff's argument addresses the fundamental dilemma with fixed journal sets: in order to establish a baseline to compare with, we impose a restriction whereby the universe and its evolution to a large extent becomes frozen at a particular point in time (Leydesdorff, 1989).

As is common with debates in the social sciences, this one also never really settled, it just died out. The proper choice of counting unit is still debated with arguments for and against fractional counting both in publication, collaboration and impact analyses (e.g., Leydesdorff, 1988; Katz & Martin, 1997; Gauffriau & Larsen, 2005; Gauffriau et al., 2007; Gauffriau et al., 2008; Hagen, 2009; Mirowski, 2010; Aksnes, Schneider, & Gunnarsson, 2012; Waltman & van Eck, 2015). In essence, the problem relates to the extensive unsolved theoretical and operational measurement challenges in scientometrics. Author and affiliation "credits" remain such a challenge. They are easily countable, albeit the construct, content and criterion validities of such counts and the claims based upon them are often questionable and problematic (e.g., Katz & Martin, 1997).

As for the use of fixed journal sets, several studies have pragmatically used such sets as comparable baselines in various publication analyses (e.g., van Leeuwen & Tijssen, 2000; Glanzel, Danell, & Persson, 2003; Tijssen & Winnink, 2016). Alternatives to using a fixed journal set approach have also been applied, for example by comparing relative national growth rates to the relative growth rate of the dynamic database (Schneider, 2010), or by accounting for changes in the database's journal coverage by comparing annual growth rates for journals indexed in two consecutive years (Moed, 2008). Irrespective of these different approaches, what is important to emphasise in a science policy context, is the fact that developments in publication output, whether in fixed or dynamic journal sets, is still only a reflection on the visibility in a particular citation database (determined by inconsistent

[1] see also Martin (2017) and Leydesdorff (2017) for recent reflexions on the debate.

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

indexing policies) and not an exhaustive outline of total publication output (Larsen & von Ins, 2010; Michels & Schmoch, 2012).

The previous uses of fixed journal sets have primarily focused on measuring developments in publication activity. Issues of citation impact have been almost absent in these discussions. In the present study, we return to the issue of using fixed journal sets as baselines in longitudinal scientometric studies. Contrary to previous studies, our focus is upon developments in and comparisons of relative citation impact for countries over time. Our aim is to substantiate recent findings of long-term stability between national impact rankings and at the same time, apparently continuous diminishing gaps in impact values between many Western countries, as well as a seemingly incessant rise in impact for a select number of "advanced" countries (Schneider & Aagaard, 2015). It is important to examine to what extent factors such as dynamic journal sets or counting methods, may influence such findings.

What perhaps is seldom recognized, especially not in a science policy context, is the fact that a field normalized citation reference value, such as the database average of one, for example used in the MNCS indicator (Waltman et al., 2012), is not a time invariant constant in a dynamic journal database[2]. The reference value is constructed annually based on the eligible publications in the database at that particular time. Hence, it is strictly dependent on the journal coverage in that year. Put simply, at any particular time calculating the database reference value is a zero-sum game where citations go in the numerator and publications in the denominator. Over time as the journal coverage and volume of a database changes, and especially when numerous new journals with lesser citation activity are included, *ceteris paribus*, the relation between numerator and denominator changes as well. In principle, this can create an imperceptible "inflation effect" (e.g., Wilson, 2007; Althouse et al., 2009; Neff & Olden, 2010; Kiesslich, Weineck, & Koelblinger, 2016; Stahlschmidt & Hinze, 2016). Journal changes in a database, especially inclusion of journals with lower citation densities may therefore result in an inflation factor that can create some artefacts in longitudinal studies of the development in impact scores for policy units of analysis such as countries. Such an inflation factor may be hard to discern and it may interact with other time variant factors that influence citation scores such as developments in co-authoring or length of reference lists. However, disregarding such "database effects" can have important consequences for science policy as a country may seem to increase its impact, which could be interpreted as an effect of some policy initiatives, whereas in fact it is more likely an artefact due to a dynamic development in the database. Such policy interpretations have for instance been voiced in relation to the seemingly continuous growth of national Danish citation impact in the last 20-years.

The present study examines such potential "database effects" by comparing dynamic and fixed journals sets. Our point of departure is the apparent stability over time in ranking of countries based on relative citation impact previously found in the dynamic version of WoS (Aagaard & Schneider, 2015). We examine the development of impact for ten different countries by comparing their impact and presence in the dynamic version of WoS to a fixed version consisting of journals selected for inclusion up until 1981 based on their impact factor and continuously indexed until 2014. The fixed set constitutes a baseline where the discrepancy in journal coverage will be small in the beginning of the period examined and larger at the end. Our fixed set is not time invariant as we do not account for developments in volume, length of reference lists etc. However, we assume that these

[2] Notice, the mathematical property where the database average sums to one is only possible when fractional counting is applied (Waltman & van Eck, 2015).

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

developments are comparable to the dynamic set and that we therefore still are able to compare developments and examine potential differences between the two sets at the country level.

Data and Method

Our analyses are based on publications from the WoS citation database. More specifically, we use a version of the WoS database enhanced and preserved by CWTS at Leiden University, the Netherlands. This version contains publications from 1981 onwards. Two data sets are used: 1) the entire database comprising publications from Science Citation Index, Social Science Citation Index, as well as the Arts and Humanities Citation Index; and 2) a reduced version of the database based on a fixed set of journals from the three aforementioned citation indices.

In order to be included in the fixed set, we require journals to be indexed continuously in the WoS database from 1981 to 2014. As these journals were selected for inclusion in WoS up until 1981 their selection was most likely based on their Journal Impact Factor (JIF). Their continuous presence in WoS can only be subscribed to the JIF up until the time when journal inclusion in WoS became a commercial matter beginning in the mid-1990s. Notice, journals changing names or ISSN number, or journals split up, are all excluded for the analyses. The inclusion criteria entail that relatively recent and seemingly central journals such as the expanding Nature and PLOS families are not included in the fixed set. Consequently, the reduced database comprises a fixed set of 3232 continuously indexed journals. Based on the eligible fixed set journals and their publications, in this case research articles, reviews and letters as classified in WoS, a restricted citation databases is created which only contains citation traffic between these 3232 journals.

We run a number of different descriptive analyses with different parameter values. In the analyses, research articles and reviews are weighted as one publication unit and letters as 0.25[3]. We examine both full and fractional counts; as we focus our analyses at the country level, we apply fractional counting according to country affiliations in the publications. Citation analyses are performed with similar parameters and for all citation analyses we apply a four-year fixed citation window and we exclude self-citations. We normalize observed citation scores according to expected citation activities in the respective WoS journal subject categories[4] to which publications are linked through their journals, also taking into consideration comparable publication years (Waltman et al., 2012). We compare the developments in citation impact for ten different countries, a group of traditionally high performing countries with different sizes: USA, UK, Netherlands, Switzerland and Denmark. A group of medium performing European countries: Norway and Spain, and finally a group of relative large but lower performing BRICS countries: China, Brazil and South Africa. The choice of countries are convenient. We have data for all countries, but since the study should be manageable, we selected what we think are interesting countries from the three broad analytical groups described above.

Results

In this section, we present some of the results of our studies. First, we examine and compare the developments in the dynamic and fixed database when it comes to journals and publications. Next, we present the main results concerning the comparisons of impact development between the dynamic and fixed sets for ten individual countries.

---

[3] We follow the procedure used in the Leiden Ranking, where letters count as "one-fourth of an ordinary publication (i.e., an article or a review) of course involves some arbitrariness. We have chosen to use a weight of 0.25 for letters because in WoS a letter on average receives roughly one-fourth of the citations of an ordinary publication" (Waltman et al., 2012, p. 2423).

[4] Notice, these categories are adjusted to fit the restricted number of journals in the fixed set.

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

Developments in publications and journals

Figure 1 below illustrates the developments in publications and journals in the dynamic and fixed sets, as well as the overlap between them. We present results for both full and fractional counts. Panel (A) shows the development in publication counts from 1981 to 2014. Panel (B) shows the development in the number of journals in the dynamic set, obviously the fixed set is constant at 3232 journals as described above. Panel (C) shows the developments in the proportion of publication counts and journals covered by the fixed set compared to the dynamic set, or in other words the overlap between the two journal sets.

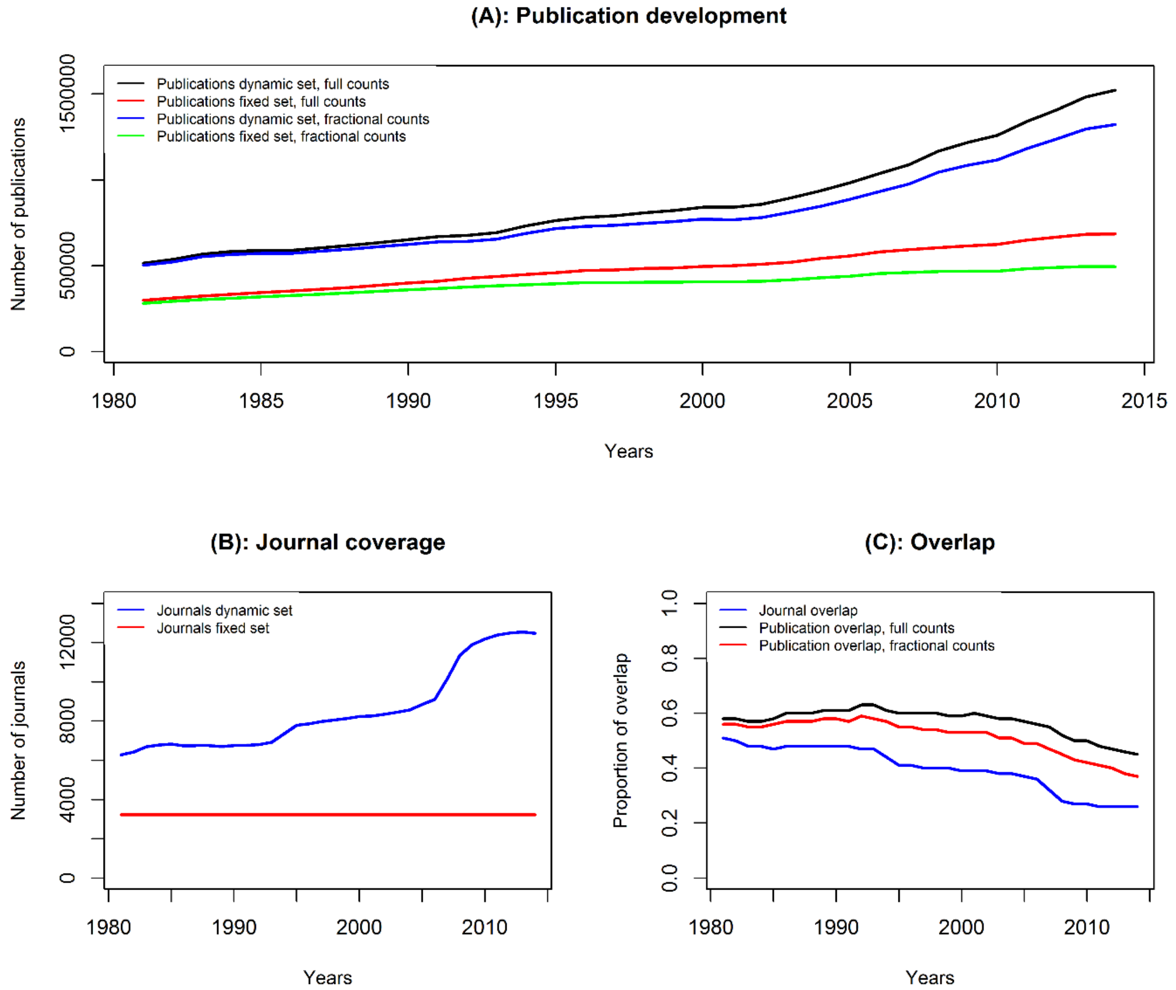


Figure 1. Development in journals and publications in the fixed and dynamic journal sets and between them.

Panel (B) shows the well-known growth in the WoS database with the sudden increases in journal coverage first around 1995 and more markedly around 2005/6. These changes in journal coverage is also reflected in panel (A). We see a small upward jump in publication counts for the dynamic database around 1995 and a more noticeable one around 2005/6 resulting in a subsequent steeper growth curve. When we compare the developments in publication counts for the fixed set with the dynamic set, we do see a continuous growth but clearly without sudden bursts. Consequently, accounting for the number of journals alone in a fixed set is in principle not enough in order to create an invariant data set. It is reasonable to assume, however, that the fixed set is influenced by the same mechanisms as the dynamic set when it comes to time variant influences on knowledge production, such as increasing number of authorships, international cooperation, average length of papers and

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

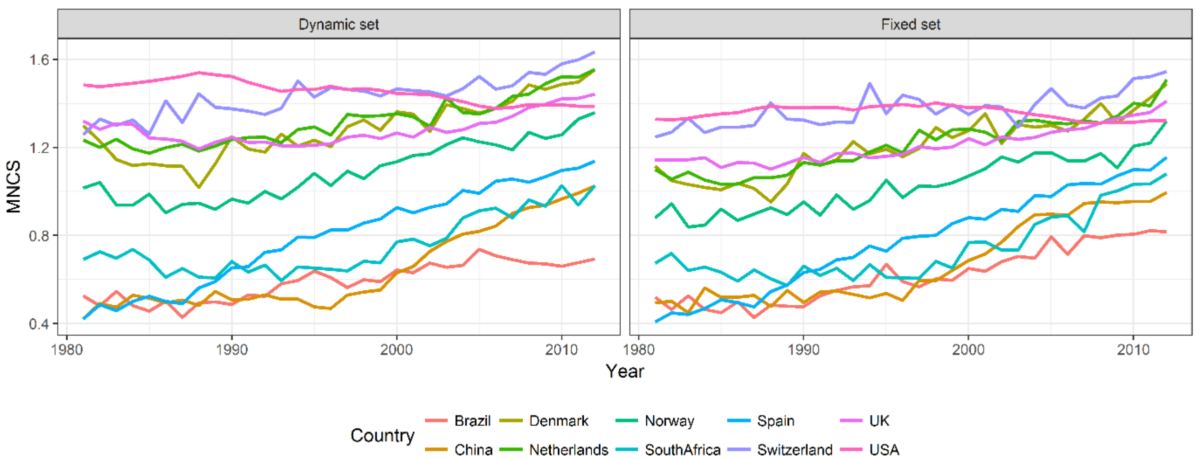
references lists, information and communication technology etc. The growth in volume is thus "variance unaccounted for" and, if necessary, will have to be dealt with either statistically and/or pragmatically. We can infer that the developments in the two sets are indeed influenced by similar mechanisms by comparing the widening gap between full and fractional counts. Fractional counting is done at the country level, hence the widening gap is a result of more internationalization which seems to affect both sets in a similar manner.

Panel C illustrates the developments in the overlaps between the two sets. For journals we see that in 1981 the overlap proportion was 0.51 (i.e. half of the journals indexed up to 1981 are covered in the database for the whole period examined) and in 2014 this has dropped to 0.26 due to the marked inclusion of new journals in the dynamic set. The overlap proportion for full counted publications was 0.58 in 1981, increasing somewhat until the early 1990s where after we see a steady decline to 0.45 in 2014. The trend for fractional counts is similar albeit the proportions are somewhat smaller, 0.56 in 1981 to 0.37 in 2014.

Consequently, in the early period examined, the fixed set included around half the journals included in the dynamic set, but more than half of publications. At the end of the period, the fixed set only contains a quarter of the journals indexed by dynamic set, but still covers 45% or 37% of the publication counts respectively. The fixed set can be seen as a core set of continuously indexed primarily Anglo-American journals that have grown steadily in volume and therefore contains a considerable amount of the total citation traffic. New journals have been added to the dynamic database. While a few of them such as the Nature and PLOS families accrue considerable citation traffic, the majority are smaller specialty journals with lesser citation traffic. We should therefore expect a substantial degree of stability in impact trends between the two journal sets for the ten countries examined, but the question is to what extent the distribution of missing citation activity outside the fixed set influences the impact scores for individual countries.

Developments in relative citation impact

Figure 2 shows the developments in relative citation impact in the dynamic and fixed journal sets for all 10 countries using *full counting*. The patterns in the dynamic set are well-known, most countries have continuously growing impact scores culminating in the most recent years and the exceptions being the US and Brazil. The fixed set produces a similar overall representation albeit with a noticeable difference. The impact scores are generally much closer to each other in the last decade examined and the scores for so-called high-performing countries are generally somewhat lower compared to the dynamic set.



This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

Figure 2. Mean normalized citation scores (MNCS) based on full counts for all 10 countries based on the dynamic and fixed versions of WoS.

Figure 3 shows the developments in relative citation impact in the dynamic and fixed journal sets for all 10 countries using *fractional counting*.

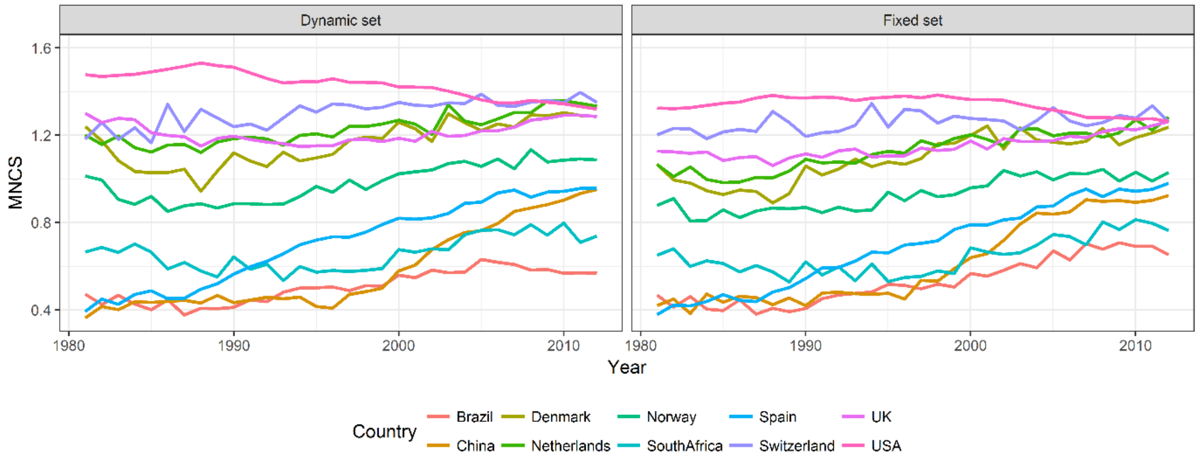


Figure 3. Mean normalized citation scores (MNCS) based on fractional counts for all 10 countries based on the dynamic and fixed versions of WoS

Again, the patterns in the dynamic set are well known. A continuous rise in international co-authoring, especially among the medium and high-performing countries brings them closer together especially during the last decade resulting in miniscule differences in impact scores. Contrary to the developments for the full counting approach in Figure 2, fractional counting results in much lower growth rates. . For several high-performing countries, developments over time seem much more flat. Noticeable exceptions are Spain and China. Also, US performance does not decline in the same manner as with full counts testifying to the country's central collaborative role. Similar to Figure 2, the fixed set produces a comparable overall representation of impact development albeit again with the noticeable difference that the impact scores are generally much closer to each other in the last decade examined and the scores for so-called high-performing countries are generally somewhat lower compared to the dynamic set.

Overall, when we compare the developments in the rankings between the countries in the dynamic and fixed sets in Figures 2 and 3, we can infer that they are very stable and very similar. What differs is the density between impact scores, which are closer in the fixed set albeit still with a recognizable gap between two performance groups, most pronounced in Figure 3 where the concentration effects of international co-authoring are most noticeable.

Figures S4 and S5 in the supplementary material contrasts the developments in impact scores for the individual countries. The patterns are similar between full and fractional counting for the individual countries. So-called high-performing countries, e.g. Denmark, Netherlands, Switzerland, UK and the US, generally have slightly lower impact scores in the fixed set. There are individual differences, e.g. the gap for the UK and the US is largest in the early period examined which is most likely a result of excluding many Anglo-American journals that were indexed in the database early on but not continuously up until 2014. On the other hand, Brazil and China seem to benefit from the fixed set at certain times. Brazil have higher impact scores in the fixed set in the last period, whereas China also has larger impact scores in the fixed set in a decade long period from around 1995 onwards.

The implications of these findings are discussed in the next final section.

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

Discussion

The use of fixed journal sets in publication analyses at the country level have been fiercely debated in the scientometric community (e.g., Martin et al., 1987; Leydesdorff, 1988, 1989, 1991). Indeed, the choice of journal sets was only one among several important parameter choices in such studies and the debate clearly demonstrated that the outcomes of publication analyses at the country level are highly dependent on these choices, (Braun, Glänzel, & Schubert, 1989). Obviously, it is not appropriate that outcomes are so volatile and dependent on parameter choices, especially in policy matters.

The important question is to what extent such volatile outcomes also apply to longitudinal analyses of relative citation scores at the country level. We therefore set out to examine this in the present study. Our aim was to examine to what extent recent longitudinal findings in citation impact development for countries are influenced by the fact that they are derived from dynamic journal sets as well as to examine the influence of counting methods. Would we see similar volatile outcomes as in the previous cases with publication analyses?

Hence, we constructed a fixed journal set comprising of 3232 journals continuously indexed in the WoS from 1981 to 2014. From this restricted set, a citation database depicting the citing relations between the journal publications was created. Relative citation scores, their developments and the rankings between the countries were then compared to the traditional, dynamic version of the WoS.

Main findings

Overall, we show that the fixed journal sets largely mirrors the dynamic sets in terms of the long-term developments in relative citation scores at the country level. Consequently, the ranking between countries remains stable over time, albeit the gap in impact scores between higher and lower performing countries diminishes somewhat. This is true for both full and fractional counting. Consequently, we demonstrate that one of the main issues previously debated in relation to publication analyses, choice of journal set, does not seem to influence the general outcomes in relation to relative citation scores at the country level. Stable results are important especially for science policy. However, the results also show that the high-performing countries generally have somewhat lower citation scores in the fixed set compared to their scores in the dynamic set, hence while seemingly not influencing the ranking between the countries, the density between citation scores and countries become smaller in the fixed set, especially in the most recent period. It is also evident that the trajectories of the citation profiles and the density between them are somewhat different when we compare full and fractional counting, albeit not between the dynamic and fixed sets. This is not surprising in as much as fractional counting adjusts for international collaboration in our study. Clearly, full and fractional counting measure different constructs and relative citation indicators are also to some extent influenced by their parent counting method.

We can thus conclude that publication analyses are much more vulnerable to choices of journals sets as publication behaviour obviously change over time. Most interestingly, the same does not seem to be the case with citation behaviour which seem to be much stable. In this study, we are able to reproduce the main citation profiles for countries in a fixed journal set despite considerable changes in publication behaviour both in the fixed and dynamics sets. We therefore conclude that the recent longitudinal findings in citation impact development for countries in a dynamic database are stable.

We will now elaborate on these findings.

Some explanations

The apparent isomorphism in country profiles between the dynamic and fixed sets is interesting as the journal overlap between the two sets decreases substantially over time from 0.51 in 1981 to 0.26

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

in 2014. In other words, in spite of the overall growth in the dynamic set, the fixed set with its 3232 journals still contains the vast majority of the most visible and important journals in the WoS. Many of them are old Anglo-American journals which continue to generate considerable citation activity.

There are however, some minor, but systematic, discrepancies when it comes to actual citation scores and their developments. In particular, for the high-performing countries it is observed that the citation scores in general are marginally lower in the fixed set. There may be several intertwined expiations for this. For one thing, citation traffic is not symmetric between units. High-performers tend to receive citations from the entire performance strata, whereas their own citing activity typically is more concentrated mainly in the higher performance strata. It is therefore likely that high-performing countries will lose relatively more citations going from a dynamic to a fixed set of well-established journals, compared to lower performing countries. In addition to losing citations, cumulative effects beneficial to high-performing countries are most likely also affected by this move (M Bonitz, 1997; M Bonitz, Bruckner, & Scharnhorst, 1997; Manfred Bonitz, 2005).

Like Althouse and colleagues (2009) and Stahlschmidt and Hinze (2016) we also find it plausible that some of the discrepancies can be ascribed to small "inflation effects". Althouse and colleagues (2009, p. 27) estimate the average increase in a weighted impact factor for all fields in a dynamic version of WoS to be on average 2.6% per year over the period 1994–2005. They claim that the predominant factor responsible for this inflation effect is the fact that the average number of references in papers has increased gradually in the period under examination. Further, the authors also claim that the growth rate in the dynamic database has had very little influence on impact factor inflation. More recently, Stahlschmidt and Hinze (2016) have also examined the seemingly unbroken impact growth in recent decades among most countries. Knowing that not every country can improve its relative performance to all other countries in the database, Stahlschmidt and Hinze (2016) also turn to the dynamic nature of the database as an explanatory factor. In particular, they focus upon the exponential growth of Chinese publications and the influence this may have on the supposed "inflation factor". To examine this, the authors construct a "counterfactual" database by removing all Chinese publications from WoS and compare differences in mean normalized citation scores (MNCS) based on fractional counts between the two versions. Like the present findings, the overall difference in MNCS decreases over time, but the effect is weak and there are country variations. According to the authors the explanation of the (limited) inflation effect is, that Chinese publications cite developed countries' publications above average and that the high performing countries therefore benefit the most from the growth of Chinese publications in the WoS.

As stated above, and similar to Stahlschmidt and Hinze (2016), we also ascribe some of the effects found in our study to an increase in obtained citations for the high-performing countries in the dynamic set compared to the fixed set. But like other studies, we also find this effect to be relatively insignificant. Potential database effects at the country level thus seem to be of very limited importance compared to the stability of citation networks.

Hence, the most interesting finding of the study is not the small inflation effects, but rather the fact that with the fixed set, we are able to reproduce the relative developments in impact scores and the rankings between the countries found in the constantly growing dynamic set. A relatively small "core set" of only 3232 journals thus still dominate the global scholarly communication patterns and its citation activity. An interesting question, which we will dive deeper into in further studies, is then whether an even smaller number of "core journals" also can replicate the same developments? From early on the SCI was constructed around a core set of journals that accounted for the vast majority of the perceived citation traffic within science. Garfield and others have repeatedly claimed that "[a] small number of journals accounts for the bulk of significant scientific results" (1996, p. 13), and that

This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3

this core set has been remarkable stable over the years (see also M. Bonitz, 2002; Ioannidis, 2006). Obviously, factors such as increases in volume, co-authorships and length of reference lists may have increased citation traffic considerably over the years. Yet, what seems evident is that citation distributions are not only skewed and highly concentrated, but also extremely stable. This fact perhaps more than any other explains why we can replicate the developments in WoS with a smaller journal set.

Acknowledgements

The research was funded by the Research Council of Norway, grant number 256223 (the R-QUEST centre).

References

Aksnes, D. W., Schneider, J. W., & Gunnarsson, M. (2012). Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalised counting methods. *Journal of Informetrics, 6*(1), 36-43. doi:http://dx.doi.org/10.1016/j.joi.2011.08.002

Althouse, B. M., West, J. D., Bergstrom, C. T., & Bergstrom, T. (2009). Differences in impact factor across fields and over time. *Journal of the American Society for Information Science and Technology, 60*(1), 27-34. doi:10.1002/asi.20936

Bonitz, M. (1997). The scientific talents of nations. *Libri, 47*(4), 206-213. doi:10.1515/libr.1997.47.4.206

Bonitz, M. (2002). Ranking of nations and heightened competition in Matthew core journals: Two faces of the Matthew effect for countries. *Library Trends, 50*(3), 440-460.

Bonitz, M. (2005). Ten years Matthew effect for countries. *Scientometrics, 64*(3), 375-379. doi:10.1007/s11192-005-0256-5

Bonitz, M., Bruckner, E., & Scharnhorst, A. (1997). Characteristics and impact of the matthew effect for countries. *Scientometrics, 40*(3), 407-422. doi:10.1007/BF02459289

Braun, T., Glänzel, W., & Schubert, A. (1989). Assessing assessments of British science. Some facts and figures to accept or decline. *Scientometrics, 15*(3), 165-170. doi:10.1007/bf02017195

Braun, T., Glänzel, W., & Schubert, A. (1991). The bibliometric assessment of UK scientific performance—Some comments on Martin's “reply”. *Scientometrics, 20*(2), 359-362. doi:10.1007/bf02017525

Cimini, G., Gabrielli, A., & Sylos Labini, F. (2014). The Scientific Competitiveness of Nations. *PLoS ONE, 9*(12), e113470. doi:10.1371/journal.pone.0113470

Garfield, E. (1955). Citation indexes to science: a new dimension in documentation through association of ideas. *Science, 122*(3159), 108-111.

Garfield, E. (1996). The significant scientific literature appears in a small core of journals. . *The Scientist, 10*(17), 13-15.

Garfield, E., & Sher, I. H. (1963). *Genetics Citation Index*. http://www.garfield.library.upenn.edu/essays/v7p515y1984.pdf, Philadelphia, PA.

Gauffriau, M., Larsen, P., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics, 73*(2), 175-214. doi:10.1007/s11192-007-1800-2

Gauffriau, M., Larsen, P., Maye, I., Roulin-Perriard, A., & von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics, 77*(1), 147-176. doi:10.1007/s11192-007-1934-2

Gauffriau, M., & Larsen, P. O. (2005). Counting methods are decisive for rankings based on publication and citation studies. *Scientometrics, 64*(1), 85-93. doi:10.1007/s11192-005-0239-6

This is the accepted version of the publication. The final published version is available here: https://doi.org/10.1007/s11192-019-03082-3

Glanzel, W., Danell, R., & Persson, O. (2003). The decline of Swedish neuroscience: Decomposing a bibliometric national science indicator. *Scientometrics, 57*(2), 197-213. doi:10.1023/a:1024185601555
Hagen, N. T. (2009). Credit for Coauthors. *Science, 323*(5914), 583-583.
Ioannidis, J. P. A. (2006). Concentration of the Most-Cited Papers in the Scientific Literature: Analysis of Journal Ecosystems. *PLoS ONE, 1*(1), e5. doi:10.1371/journal.pone.0000005
Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18. doi:10.1016/s0048-7333(96)00917-1
Kiesslich, T., Weineck, S. B., & Koelblinger, D. (2016). Reasons for Journal Impact Factor Changes: Influence of Changing Source Items. *PLoS ONE, 11*(4), e0154199. doi:10.1371/journal.pone.0154199
Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics, 84*(3), 575-603. doi:10.1007/s11192-010-0202-z
Leydesdorff, L. (1988). Problems with the 'measurement' of national scientific performance. *Science and Public Policy, 15*(3), 149-152. doi:10.1093/spp/15.3.149
Leydesdorff, L. (1989). The science citation index and the measurement of national performance in terms of numbers of scientific publications. *Scientometrics, 17*(1), 111-120. doi:10.1007/bf02017727
Leydesdorff, L. (1991). On the "scientometric decline" of British science. One additional graph in reply to Ben Martin. *Scientometrics, 20*(2), 363-367. doi:10.1007/bf02017526
Leydesdorff, L. (2002). Dynamic and evolutionary updates of classificatory schemes in scientific journal structures. *Journal of the American Society for Information Science and Technology, 53*(12), 987-994. doi:10.1002/asi.10144
Leydesdorff, L. (2017). The positive side of discursive disagreements in the social sciences. *Journal of Informetrics, 11*(4), 1043. doi:https://doi.org/10.1016/j.joi.2017.09.006
Leydesdorff, L., & Wagner, C. (2009). Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics, 78*(1), 23-36. doi:10.1007/s11192-008-1830-4
Martin, B. R. (1991). The bibliometric assessment of UK scientific performance a reply to Braun, Glänzel and Schubert. *Scientometrics, 20*(2), 333-357. doi:10.1007/bf02017524
Martin, B. R. (1994). British science in the 1980s — Has the relative decline continued? *Scientometrics, 29*(1), 27-56. doi:10.1007/bf02018382
Martin, B. R. (2017). When social scientists disagree: Comments on the Butler-van den Besselaar debate. *Journal of Informetrics, 11*(3), 937-940. doi:https://doi.org/10.1016/j.joi.2017.05.021
Martin, B. R., Irvine, J., Narin, F., & Sterritt, C. (1987). The continuing decline of British science. *Nature, 330*(6144), 123-126.
Michels, C., & Schmoch, U. (2012). The growth of science and database coverage. *Scientometrics, 93*(3), 831-846. doi:10.1007/s11192-012-0732-7
Mirowski, P. (2010). Bibliometrics and the Modern Commercial Regime. *Archives Europeennes De Sociologie, 51*(2), 243-270. doi:10.1017/s0003975610000123
Mirowski, P. (2011). *Science-Mart. Privatizing American Science*. Cambridge, MA & London, UK: Harvard University Press.
Moed, H. F. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics, 74*(1), 153-161. doi:10.1007/s11192-008-0108-1
Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Washington, DC: Computer Horizons Inc.
Narin, F., Stevens, K., Anderson , J., Collins, P., Irvine , J., Isard, P., & Martin, B. R. (1988). On-line approaches to measuring national scientific output: a cautionary tale. *Science and Public Policy, 15*(3), 153-161.

This is the accepted version of the publication. The final published version is available here: https://doi.org/10.1007/s11192-019-03082-3

Neff, B. D., & Olden, J. D. (2010). Not so fast: Inflation in impact factors contributes to apparent improvements in journal quality. *BioScience, 60*(6), 455-459.

Schneider, J. W. (2010). *Bibliometric Research Performance Indicators for the Nordic Countries*. Retrieved from https://www.nordforsk.org/en/publications/publications_container/bibliometric-research-performance-indicators-for-the-nordic-countries:

Schneider, J. W., & Aagaard, K. (2015). *Developments in Danish research performance*. Retrieved from http://ufm.dk/publikationer/2015/filer/dfir_scientometric_analysis_final.pdf:

Stahlschmidt, S., & Hinze, S. (2016). *How does the scientific progress in developing countries affect bibliometric impact measures of developed countries? A counterfactual case study on China*. Paper presented at the 21st International Conference on Science and Technology Indicators (STI), Universitat Politècnica de València, Spain.

Studer, K. E., & Chubin, D. E. (1980). *The Cancer Mission: Social Contexts of Biomedical Research*. London, UK: Sage.

Testa, J. (2011). *The globalization of Web of Science: 2005-2010*. Retrieved from

Tijssen, R. J. W., & Winnink, J. (2016). Twenty-first century macro-trends in the institutional fabric of science: bibliometric monitoring and analysis. *Scientometrics, 109*(3), 2181-2194. doi:10.1007/s11192-016-2041-z

van Leeuwen, T., & Tijssen, R. (2000). Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows. *Research Evaluation, 9*(3), 183-187. doi:10.3152/147154400781777241

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., . . . Wouters, P. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419-2432. doi:10.1002/asi.22708

Waltman, L., & van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics, 9*(4), 872-894. doi:http://dx.doi.org/10.1016/j.joi.2015.08.001

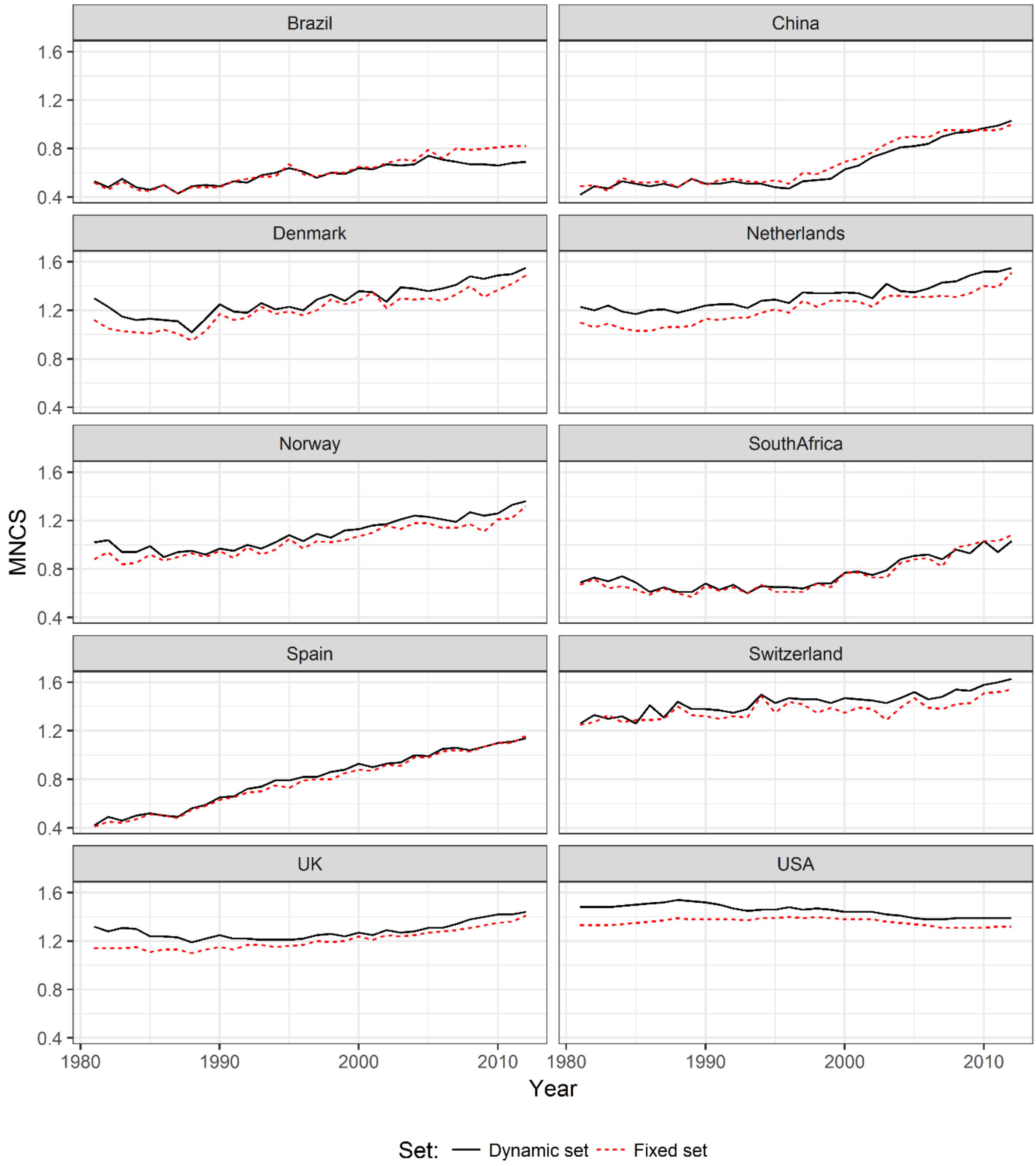
Wilson, A. E. (2007). Journal Impact Factors Are Inflated. *BioScience, 57*(7), 550-551. doi:10.1641/B570702

Aagaard, K., & Schneider, J. W. (2015). Research funding and national academic performance: Examination of a Danish success story. *Science and Public Policy*. doi:10.1093/scipol/scv058

This is the accepted version of the publication. The final published version is available here: https://doi.org/10.1007/s11192-019-03082-3

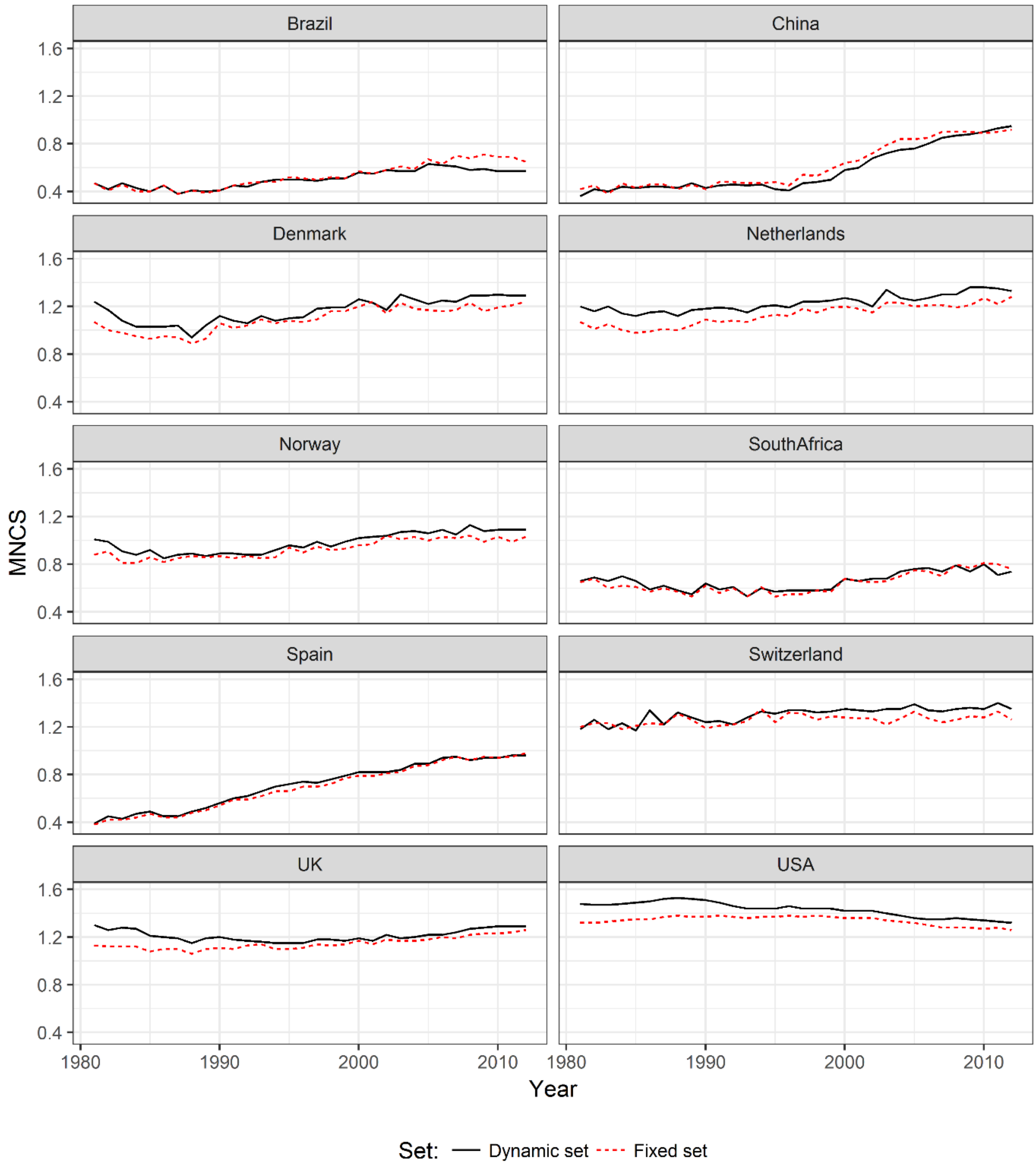
Supplementary material

Figure S1. Comparison of mean normalized citation scores (MNCS) based on full counts for 10 individual countries between the fixed and the dynamic versions of WoS.



This is the accepted version of the publication. The final published version is available here: https://doi.org/10.1007/s11192-019-03082-3

Figure S2. Comparison of mean normalized citation scores (MNCS) based on fractional counts for 10 individual countries between the fixed and the dynamic versions of WoS.



This is the accepted version of the publication. The final published version is available here:
https://doi.org/10.1007/s11192-019-03082-3