"SEASONAL ADJUSTMENT" IN GOOGLE SCHOLAR AND MICROSOFT ACADEMIC

Andreea MIRICA
Assistant Lecturer, PhD,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: miricaandreea89@gmail.com

Octavian CEBAN
PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: octavianceban1995@gmail.com

Georgiana Andreea FERARIU
PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: ferariugeorgiana94@gmail.com

Viorel CÎRNU
PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: viorelcirnu8@gmail.com

Bogdan Ionut CHIPER
PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: bogdan.chiper@yahoo.com

Nicoleta Violeta VELISCA
PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: nicoleta.velisca@yahoo.com

Marian NECULA[1]
PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: neculamarian18@stud.ase.ro

Abstract

The aim of the present paper was to explore the visibility of the concept of "seasonal adjustment" with regard to research field, using the Publish or Perish 7 tool. The online scientific databases analyzed were Google Scholar and Microsoft Academic. Furthermore, the compared analysis provides consistent findings which will help the researchers to better explore this topic and some useful recommendations regarding the use of these databases for retrieving articles on seasonal adjustment will be pointed out.

Keywords: *seasonal adjustment; academic visibility; Google Scholar; Microsoft Academic*

Introduction

Academic visibility of a research field is mainly a measure encompassing all the activities and their results in promoting research and interest on the subject-matter, inside and outside the field. This measure is expressed by how much money that field attracts and the number of citations. In support of this argument, the US National Science Foundation, with a budget of 8.3 billion dollars for the fiscal year of 2020, allocates roughly 80 million dollars for social and economics sciences, compared to approximately 230 million dollars to mathematics or to the 960 million dollars to computer science and related fields (NSF, 2020). In Europe, the European Research Council, perhaps the largest governmental body for funding science with a budget of 13 billion Euros, has allocated for 2019 roughly 50 million euros for social sciences research projects (ERC, 2020). There are no evidence that this financing model is not adopted across the world at the state level. Although these figures are arguable, at least from different perspectives, they provide an objective measure where social sciences are positioned in terms of visibility, academic or non-academic, compared to other research fields. The second measure of academic visibility is provided by the number of citations a paper that a broad research field attracts. According to Patience et al. (2017), the number of average citations a paper from economics receives is about 80, but when a look at the mean number of agencies interested in financing economics we get 0.3., compared to papers in applied mathematics which get 70 citations on average, and with 1.6 agencies on average financing. According to the same source for political science or sociology, where for a mean number of citations around 35-40, the average number of agencies interested in financing these fields is approximately 0. Although these measures are not the only ones to measure academic visibility, as they are arguable of predicting the future importance of a research field, they provide some insights on the importance of academic visibility in providing a sustainable background.

Even though the bibliometric indicators for a broad field may provide useful insights with regard to its academic visibility, researchers may be interested in figures for more specific fields. That is because researchers in a field often compete for funds (Fang and Casadeval, 2015) and pursuing a topic that naturally attracts a high number of citations can be a competitive advantage. This paper provides an example of how the visibility of such a topic can be analyzed. As research is a way of bridging the gap between academia and experts (Williams, 2014), the topic chosen is more of a high interest for experts within the field (statisticians in this case) than for academia. That is because, as the paper will point out, major developments within this area occurred within organizations other than academia.

The aim of this paper is to explore the visibility of the concept of "seasonal adjustment" in two major online scientific databases: Google Scholar (GS), Microsoft Academic (MA). More specifically, various indicators regarding the number of papers, the number of citations as well as more advanced bibliometric indicators will be explored. Next some recommendations with regard to the use of these databases for retrieving articles on seasonal adjustment will be provided. Such an analysis is useful for researchers within the field of seasonal adjustments from two perspectives: firstly, it provides some insights with regard to the dynamics of the field and how this dynamics is reflected in various databases; secondly, the tool used in this example may have other applications as well, that will be stated in the last section of the paper.

Methodology

For the purpose of this research, we used the Publish or Perish 7 tool, provided by Harzing, A.W. (2007) Publish or Perish, available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020). This is one of the most widely used tool for retrieving data with regard to academic visibility of researchers, papers or even domains in various databases (Harzing, 2016a). Using this tool, we performed web crawling on Google Scholar and Microsoft Academics on 07/02/2020 (09:59) and 07/02/2020 (09:41) respectively. For Microsoft Academics we used the API for Project Academic Knowledge. Data from Google Scholar and Microsoft Academics are mined from the Google and Bing web services respectively (Microsoft, 2018; Google, 2019). One should note that the advantages and disadvantages of these two databases will be reflected in the results returned by Publish or Perish (Harzing, 2016b).

Perhaps the biggest advantage of Google Scholar is the perception of a user-friendly tool, which may be due to the popularity of the Google search engine. As one can observe, one can see articles from a certain area of interest and there is also a newsletter that announces about specific articles related to a topic of interest, which are other advantages of the engine. Similarly, it has the option to create alerts, which sends emails with news about a topic set when creating the alert. Moreover, using GS, one can see how many citations an article had and where and It also has a section where different articles can be saved for later reading. Moreover, according to McQuade Library (2019), GS has multiple partnerships with various providers such as PubMed and provides results regarding legal documents.

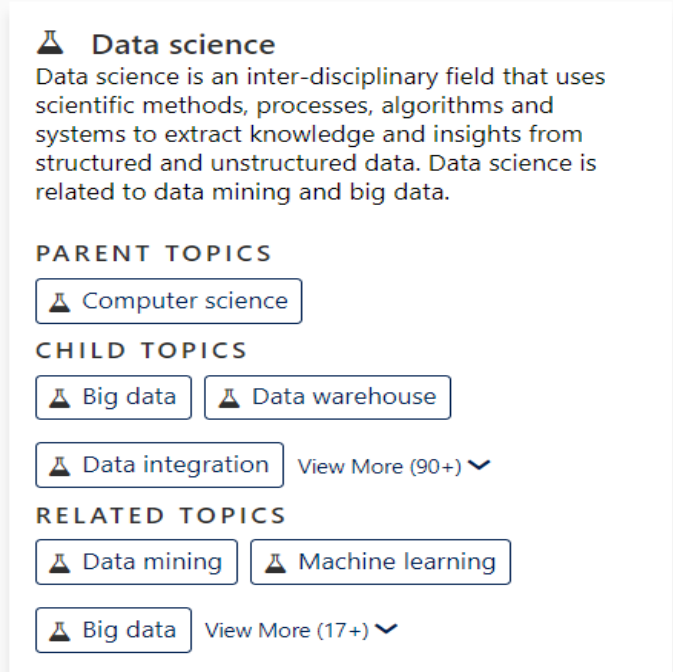

Figure 1. Additional results for "data science for football" search on Microsoft Academics
Source: Microsoft Academics: https://academic.microsoft.com/search?q=data%20science%20for%20football&f=&orderBy=4 accessed 15 February 2020

Among the disadvantages, as Université Bretagne Loire (2017) points out, perhaps the main one is that it is not comprehensive, as in certain searches the results incomplete due to the lack of articles in the database or the functioning of the search engine. Moreover,

according to the same source, sorting does not offer too many options and there are difficulties in searching for articles by the author.

Microsoft Academic Search (MAS) offers the possibility to query other documents besides articles. This information may help the user to better understand the subject and its connections, but also to filter the information. The information that also helps with filtering is Top Topics where topics are associated with what has been searched and can provide a broader perspective of what is being searched, Publication Type, Top Authors, Top Journals. In addition, in MAS there is a connection between topics, showing links between the concepts in the query. For example, when searching for "data science for football" there is an association between different terms, as shown in the figure below.

Unlike Web of Science and Scopus both GS and MAS databases are free, and this increases the audience and is welcome in developed countries (Ortega & Aguillo 2014). Moreover they are under continuous innovation (Waldrop, M. M. 2008).

The methodology of this research comprises of four steps: the analysis of descriptive indicators, graphical analysis on the evolution of the number of papers over time, graphical analysis of the number of papers by the number of authors and graphical analysis of the number of papers by the number of authors. Each of these steps is described in the following three paragraphs.

Firstly, some of the descriptive indicators provided by the program when running a query will be analyzed. More specifically we will compare Google Scholar and Microsoft Academic with regard to the following indicators for papers retrieved when entering the term “seasonal adjustment”: the year of the first article published, the year of the last article published, the number of papers, the number of citations, the number of citations/year, the number of citations/paper, the number of papers with Annual Citation Count higher than 1, 2, 5, 10 and 20 respectively and the number of authors/paper. The first three indicators will provide some insights on the coverage of these databases regarding the coverage of the topic in these databases. The next 8 selected indicators may be a measure of visibility and impact of research in this area, as they are based on the number of citations Ale Ebrahim (2014). Finally, the number of authors per paper may provide insights on the productivity as “there appears to be an inverse relationship between productivity and the number of authors per paper” (Sridhar, 2002, p.250).

Secondly, the paper explores the evolution of the number of articles throughout the entire time span each database covers. Moreover, the results, especially very high or very low values will be correlated with important events in the field such as the release of a new software tool.

Thirdly, the evolution of the number of papers by the number of authors per paper will be analysed. More specifically, major trends in the field with regard to productivity will be pointed out.

Finally, an analysis on the number of papers by the number of citations will be performed. This analysis will reveal major trends with regard to the visibility of research in this area.

Results

Table 1 presents several descriptive indicators for articles on seasonal adjustment present on Google Scholar and Microsoft Academic respectively. As one can observe,

Google Scholar covers 123 years of publications, 1.4 wider compared to Microsoft Academic of 86 years. Moreover the number of papers as well as the number of citations is considerable higher for Google Scholar. Consequently, the number of citations per year as well as per paper is higher. The results are consistent with the number of papers by Annual Citation Count. With regard to the number of authors, one can observe that the values for the two databases are approximately the same.

Table 1. Descriptive indicators for articles on seasonal adjustment retrieved from the two databases; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

	Google Scholar	Microsoft Academic
Year of the first article published	1896	1933
Year of the last article published	2019	2019
Number of papers	990	614
Number of citations	53447	10535
Cites/year	431.02	121.09
Cites/paper	53.98687	17.15798
Authors/paper	1.96	2
Papers with ACC higher than 1	386	89
Papers with ACC higher than 2	255	53
Papers with ACC higher than 5	127	24
Papers with ACC higher than 10	56	9
Papers with ACC higher than 20	23	2

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

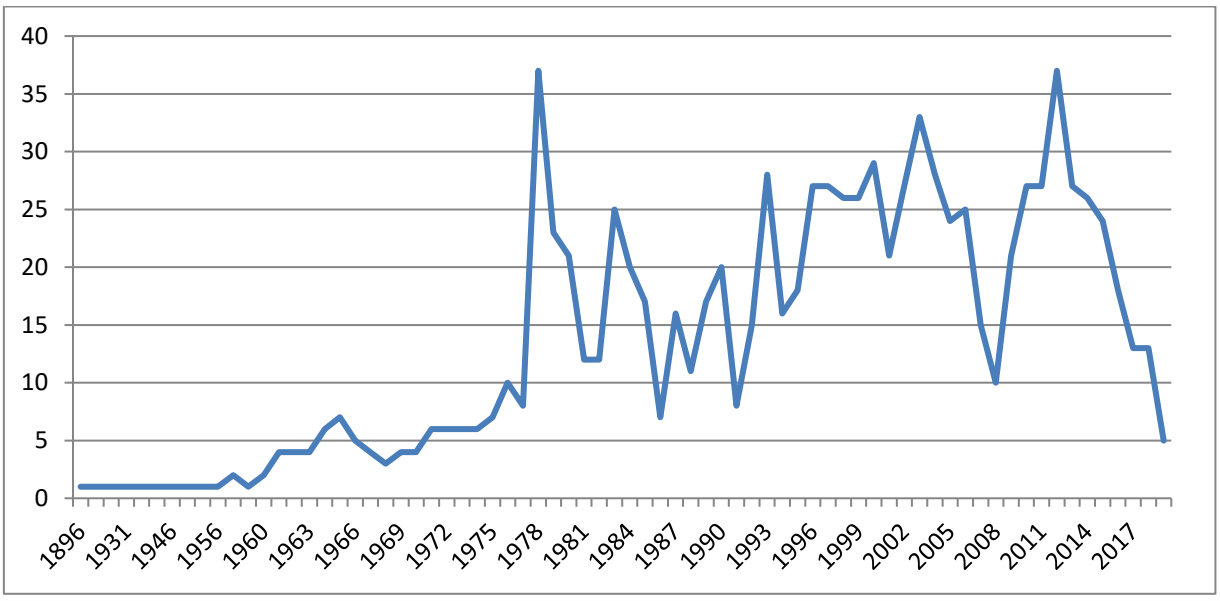

Figure 2. The evolution of the number of articles on seasonal adjustment in Google Scholar, by year of publication; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

Figures 2 and 3 present the yearly number of papers on seasonal adjustment within the two databases. The results are consistent with one another. Both graphs display peaks in 1978, 1996 and 2012. There the explosion of the number of papers in 1978 might be explained by programmes initiated by statistical offices such as the Census Bureau Programs

to Measure Consumer Purchase Expectations (see McNeil, 1974). Later, in 1980 the Statistics Canada developed the X11 method and in 1990 it was further refined by the Census Bureau (Australian Bureau of Statistics, 2017). Next, in 1996, the TRAMO – SEATS method was developed (International Monetary Fund, 2017). Also, one should note that several developments occurred in the field starting 2008 such as the introduction of guidelines and procedures in several public organisations (Australian Bureau of Statistics, Statistics Canada, Bank of Spain) as well the release of Demetra+ in 2012 (Eurostat, 2020).

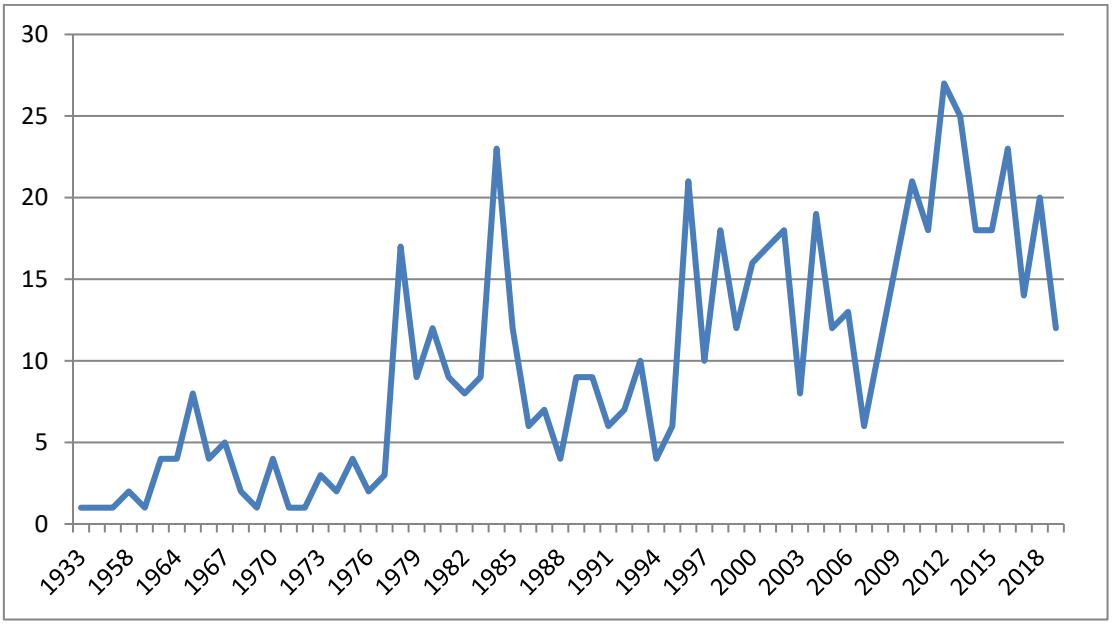

Figure 3. The evolution of the number of articles on seasonal adjustment in Microsoft Academic, by year of publication; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

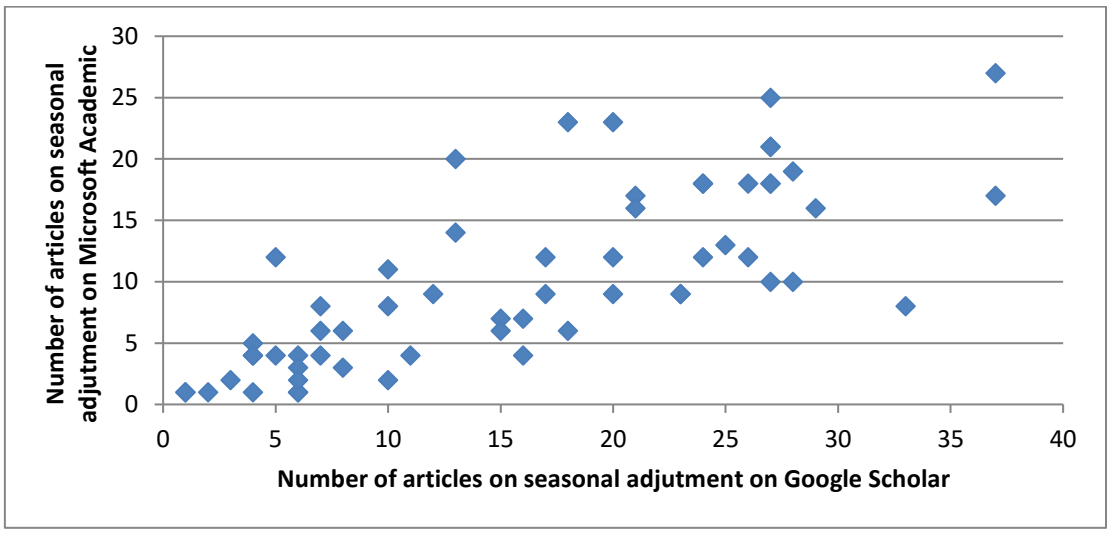

Figure 4. Number of articles on seasonal adjustment on Google Scholar and Microsoft Academic; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

The relationship between the number of articles in Google Scholar and the number of articles in Google Academics is displayed in figure 4. Only those years that are common among the two databases were kept. As one can observe, the relationship is a linear one. Moreover, the Pearson correlation coefficient is 0.76, suggesting a strong positive correlation.

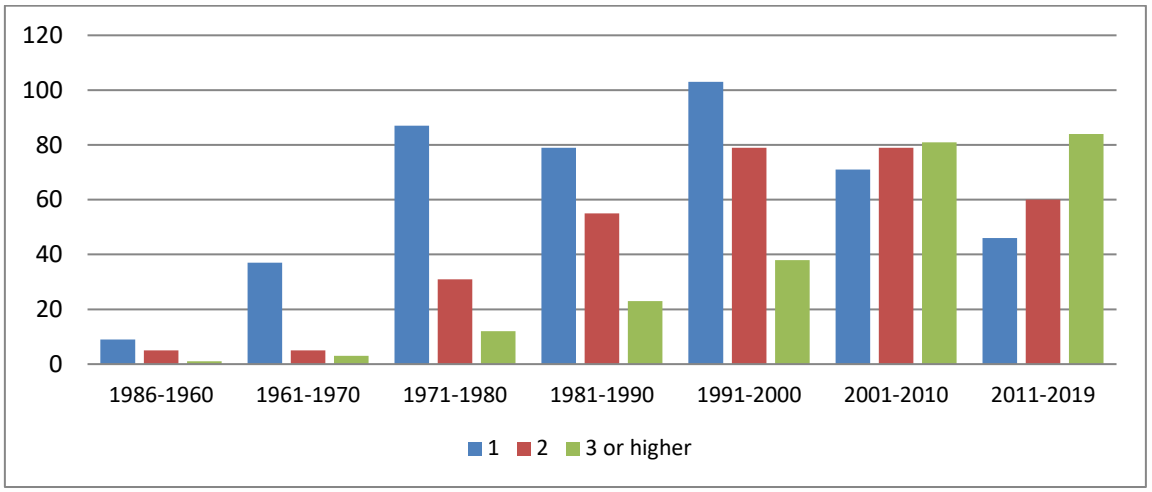

Figure 5. Number of papers on seasonal adjustment in Google Scholar by the number of authors and period published; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

Figures 5 and 6 present the number of papers by the number of authors in different periods of time in Google Scholar and Microsoft Academics. As one can observe in both databases there is a clear trend of increasing the number of papers with 3 or more authors. These findings are consistent with the one in the scientific literature (Enago Academy, 2019).

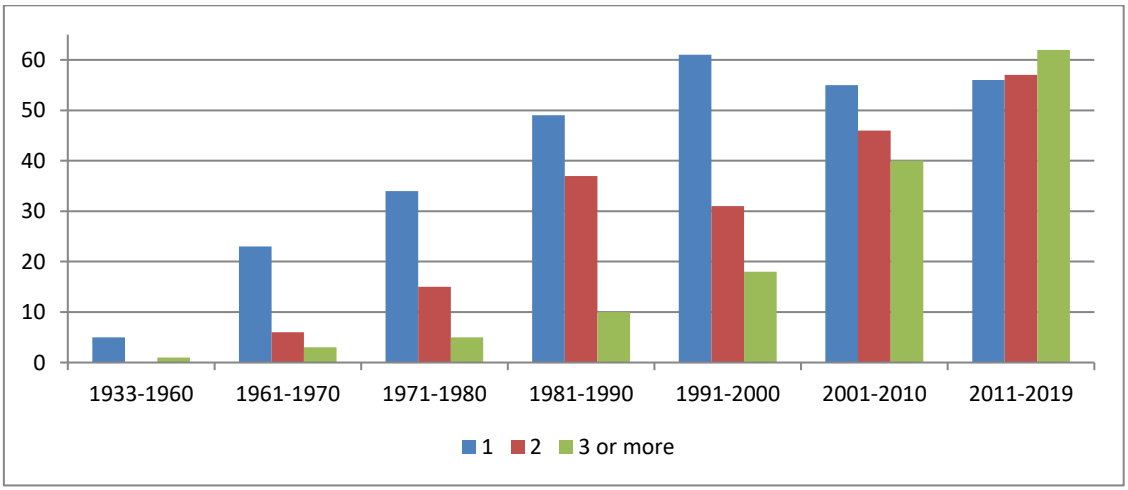

Figure 6. Number of papers on seasonal adjustment in Microsoft Academic by the number of authors and period published; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

Table 2 presents the number of articles by number of citations per year in Google Scholar and Microsoft Academic. As one can observe, there are high discrepancies between the two databases. Firstly, most of the articles in Microsoft Academic have no citations. Secondly, the number of articles with more than 1 citation per year is 4 times higher in Google Scholar than in Microsoft Academic.

Table 2. Number of articles by number of citations per year in Google Scholar and Microsoft Academic; source: designed by the authors based on data provided by Harzing, A.W. (2007) Publish or Perish

Number of citations per year	Number of articles in Google Scholar	Number of articles in Microsoft Academic
0	32	273
higher than 0 less than 1	572	252
1 and above	386	89

Note: Available from https://harzing.com/resources/publish-or-perish (accessed 7 February 2020)

Conclusions, recommendations and future research directions

Based on the importance of academic visibility, the aim of this paper was to explore the visibility of the concept of "seasonal adjustment" in two major online scientific databases (Google Scholar and Microsoft Academic), through a number of indicators such as: number of papers, numbers of citations and other advanced bibliometric indicators. The research provides not only an insight into the dynamics of the field as reflected in various databases but also a tool that may be useful for other applications.

The tool used for this research was Publish and Perish and the web crawling on the two databases was performed on 07.02.2020. The results show that the concept of seasonal adjustment is more visible on Google Scholar compared to Microsoft Academic, as the number of papers and the number of citations is considerable higher for Google Scholar, though the number of authors is practically the same. Regarding the number of papers on seasonal adjustment, the results for the two databases display peaks in 1978, 1996 and 2012, which are explained primarily by the programmes initiated by some statistical offices, and later by new methodologies (eg. TRAMO – SEATS developed by IMF in 2017) and software solutions (Demetra+, 2012). The relationship between the number of articles in Google Scholar and in Microsoft Academics is a linear one, the Pearson correlation coefficient of 0.76 suggesting a strong positive correlation.

The results obtained from the comparison of the two databases indicate that Google Scholar is a better tool for finding articles in seasonal adjustment. The descriptive analysis reveals that the number of articles available on Google Scholar is higher, but also that it covers a longer period. This is useful for a more accurate documentation when preparing the literature review.

However, one should note that this research did not perform any qualitative analysis upon the articles that were queried in the two databases. Such an analysis would reveal to what extent the articles retried are accurate towards the topic.

Although, the Science Citation Index (SCI) was developed in the '60, followed in the '70 by the Journal Citation Reports, the true revolution started along with the World Wide Web (Garfield, 2006). The development of more structured online research databases such as Web of Knowledge (2002) and Scopus (2004) encouraged and sustained the use of web scrapping instruments as well as productivity tools (Li et. al 2017). Expanding the meth-

odology and applying the used instruments for Web of Science and Scopus databases could add valuable data and provide a clearer image of this topic. A different perspective of seasonal adjustment can be explored by using instruments like Plum Analytics or Altmetrics for future studies.

References

1. Australian Bureau of Statistics. **Time Series Analysis: Seasonal Adjustment Methods**, 2017, https://www.abs.gov.au/websitedbs/d3310114.nsf/4a256353001af3ed4b2562bb00121564/c890aa8e65957397ca256ce10018c9d8!OpenDocument Accessed February 2020.
2. Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi, F., Gholizadeh, H., & Motahar, S. M. **Visibility and citation impact**. International Education Studies, Vol. 7, No. 4, 2014, pp. 120-125.
3. Enago Academy **Academic Authorship: How Many Authors Are Too Many?**, 2019, https://www.enago.com/academy/academic-authorship-how-many-authors-are-too-many/ Accessed February 2020.
4. European Research Council. **ERC funded projects**. 2020. Available at: https://erc.europa.eu/projects-figures/erc-funded-projects Accessed February 2020.
5. Eurostat **Seasonal Adjustment - Introduction and General Description (Theme).** 2020. https://ec.europa.eu/eurostat/cros/content/seasonal-adjustment-introduction-and-general-description-theme_en Accessed February 2020.
6. Fang, F. C. & Casadevall, A. **Competitive science: is competition ruining science?** 2015
7. Garfield, E. **The History and Meaning of the Journal Impact Factor**. JAMA, Vol. 295, No. 1, 2006, pp. 90-93.
8. Google Scholar **Stand on the shoulders of giants.** 2020. https://scholar.google.com/intl/en/scholar/about.html Accessed February 2020.
9. Harzing, A.W. **Publish or Perish**, 2007. available from https://harzing.com/resources/publish-or-perish Accessed February 2020.
10. Harzing, A.W. **Publish or Perish in the news.** 2016a. https://harzing.com/resources/publish-or-perish/in-the-news Accessed February 2020.
11. Harzing, A.W. **Google Scholar.** 2016b. https://harzing.com/resources/publish-or-perish/tutorial/google-scholar Accessed February 2020.
12. International Monetary Fund **Seasonal Adjustment**, 2017. https://www.imf.org/external/pubs/ft/qna/pdf/2017/chapter7.pdf Accessed February 2020.
13. Li, K., Rollins, J., & Yan, E. **Web of Science use in published research and review papers 1997–2017: a selective, dynamic, cross-domain, content-based analysis**. *Scientometrics*, Vol. *115, No.* 1, 2018, pp. 1-20.
14. McNeil, J. **Federal programs to measure consumer purchase expectations**, **1946–1973: A post-mortem**. Journal of Consumer Research, Vol. 1, No. 3, 1974, pp. 1-10.

15. McQuade Library, 2019. https://libguides.merrimack.edu/GoogleScholar/Strengths Weaknesses Accessed February 2020.
16. Microsoft **Academic Knowledge API,** 2018. https://docs.microsoft.com/en-us/azure/cognitive-services/academic-knowledge/home Accessed February 2020.
17. National Science Foundation **Fiscal Year 2020 Budget Request to Congress**. 2020. Available at: https://www.nsf.gov/about/budget/fy2020/tables.jsp#cise Accessed February 2020.
18. Ortega, Jose & Aguillo, Isidro. **Microsoft Academic Search and Google Scholar Citations: Comparative Analysis of Author Profiles**. Journal of the American Society for Information Science and Technology. Vol. 65, 2014, 10.1002/asi.23036.
19. Patience, G. S., Patience, C. A., Blais, B., & Bertrand, F. **Citation analysis of scientific categories.** Heliyon, Vol. 3, No. 5, 2017, pp. e00300. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446096/ Accessed February 2020.
20. Sridhar, M. S. **Library use and user research: with twenty case studies**. Concept publishing company. 2002
21. Université Bretagne Loire. 2017. https://guides-formadoct.u-bretagneloire.fr/c.php?g=654626&p=4598757 Accessed February 2020.
22. Williams S. **The gap between academics and practitioners is a reflection of the underlying tensions of academic belonging**. 2014. https://blogs.lse.ac.uk/impactofsocialsciences/2014/05/23/the-paradoxes-of-academic-and-practitioner-relationships/.
23. Waldrop, M. M. **Science 2.0**. Scientific American, Vol. 298, 2008, pp. 68 – 73.

[1]To whom correspondence should be addressed.