A bibliometric study of global research activity in relation to the use of partial least squares for policy evaluation

Rosanna Cataldo, Laura Antonucci, Corrado Crocetta,
Maria Gabriella Grassia, Marina Marino

1. Partial Least Squares for policy evaluation

Structural Equation Modeling (SEM), especially Partial Least Squares - Path Modeling (PLS-PM) has become a mainstream method in many fields of research. Indeed, PLS-PM has been used in the social and behavioral sciences, rooted in psychometrics and in the literature on causal modeling. In the last few years it has been increasingly disseminated in a variety of disciplines, and, in particular, has been extensively used in the business and management sciences. Within these research projects, PLS-PM has been applied successfully in studies concerning the measurement of intangibles like customer and employee perceptions (e.g. satisfaction, motivation and loyalty). These kinds of model are becoming crucial to managers in order to improve their decision making processes and increase their organization's profitability. In every time and place the decision making process has always been complex. Generally, it applies evaluation principles and methods to examine the content, implementation or impact of a policy or a decision. In the last few years, researchers have been promoting statistical methods such as SEM and PLS-PM for the evaluation of policies, especially in the context of decision making. In the literature, empirical approaches which have applied PLS-PM to decision making have been identified through a systematic literature search. To better understand and characterize this trend, a bibliometric study of international papers on this subject has been developed in order to describe the use of SEM and PLS-PM approaches in policy evaluation during the last twenty years.

2. Study Methodology

A bibliometric analysis has been used to analyse the trends in the field of SEM in the context of decision making. Bibliometric analysis is a quantitative approach for the analysis of academic literature using bibliographies to provide the description, evaluation and monitoring of the published research (Garfield et al., 1964); (White and McCain, 1989). The methodological aim is to analyse publications, citations and sources of information (Rodriguez-Soler et al., 2020). Bibliographic data are processed through a workflow: study design, data collection, data analysis, data visualization and interpretation. The analysis has been performed using the Bibliometrix R-Tool (Aria and Cuccurullo, 2017), a recent R-package which facilitates a more complete bibliometric analysis employing specific tools for both bibliometric and scientometric quantitative research. The Bibliometrix R-package (http://www.bibliometrix.org) provides a set of tools for quantitative research in bibliometrics and scientometrics, supporting scholars in three key phases of analysis: 1) data importing and conversion to the R format; 2) bibliometric analysis of a publication dataset; and 3) building matrices for co-citation, coupling, collaboration and co-word analysis. The R program and the bibliometrix codes have been used to produce a descriptive bibliometric analysis and to construct the matrices. In addition, "biblioshiny" (Aria and Cuccurullo, 2017), a Bibliometrix web-interface, has been used to build a

Rosanna Cataldo, University of Naples Federico II, Italy, rosanna.cataldo2@unina.it, 0000-0002-6324-8252
Laura Antonucci, University of Foggia, Italy, laura.antonucci@unifg.it, 0000-0003-0211-2578
Corrado Crocetta, University of Foggia, Italy, corrado.crocetta@unifg.it, 0000-0001-9059-5092
Maria Gabriella Grassia, University of Naples Federico II, Italy, mariagabriella.grassia@unina.it, 0000-0002-7128-7323
Marina Marino, University of Naples Federico II, Italy, marina.marino@unina.it, 0000-0002-0742-5912

conceptual map and network for co-citation. Matrices are the input data for the performance of network analysis, multiple correspondence analysis and certain data reduction techniques (Aria and Cuccurullo, 2017).

3. Data Collection

With the aim of understanding how the research on SEM and PLS-PM issues has evolved, the data were retrieved from two main databases commonly used by researchers: Scopus and Web of Science (WoS). Scopus and WoS are the world's most trusted independent global citation database. They are recognised as covering a broad range of relevant journals and peer-reviewed articles of high quality (Skute et al., 2019). These databases have already been used in bibliometric analysis in different disciplines, sometimes individually, in the case of WoS (Diem and Wolter, 2013); (Falagas et al. , 2006) and Scopus (Maharana, 2013); (Morandi et al., 2015), and sometimes in combination (Rodriguez-Soler et al., 2020).

We extracted articles published between 2000 and 2020 (incl.) which contained the topic "decision making" with the following keywords in the title or abstract: "PLS-PM"; "PLS Path modeling"; "PLS-Path modeling"; "SEM-PLS" ("decision making" AND "PLS-PM" OR "PLS Path modeling" OR "PLS-Path modeling" OR "SEM-PLS"). The data were downloaded on December 5, 2020. Only articles, reviews, proceedings papers and book chapters were included, with document types such as editorials, notes and corrections were excluded from the study. By merging the Scopus and WoS databases, 93 duplicate documents were removed. This process resulted in a final sample of 451 articles, which constitute the core material of this study, relating to 1,308 authors and 323 sources.

4. Analysis and Discussion

In the analysis of the data, a descriptive analysis was initially performed. Next, bibliometric techniques were developed using conceptual, intellectual or social networks.

Figure 1 shows the growth of publications from 2000 to 2020.
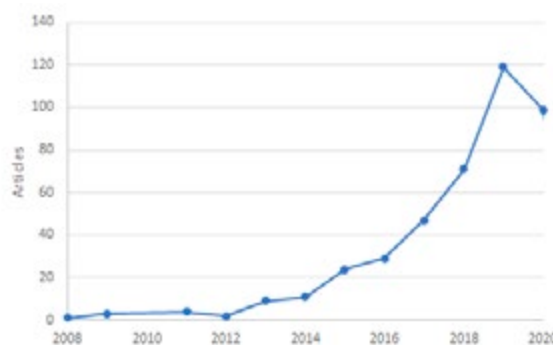
Figure 1: Growth trajectory of the literature relating to the use of PLS-PM in decision making, 2008 - 2020

As can we see, the first studies dealing with issues related to decision making were carried out in 2008. In the first years of the analysis (2008 - 2012) the number of publications is very low, emphasizing the fact that the topic was probably not very well developed and addressed by researchers. In 2019 we see a peak in the number of publications relating to the PLS-PM

approach as a statistical methods in the context of decision making.

Concerning the sources, the distribution of the articles does not present any significant concentration. The journals which included the most frequently quoted articles, containing "decision making" and "PLS-PM" as keywords, are presented in Figure 2, with the largest number of articles, namely 17, published in the *Sustainability* journal, followed by the *International journal of environmental research and public health*, a journal which deals with issues related to environmental health sciences and public health, measurement and monitoring models.

Figure 2: The most relevant sources

As regards provenance, the research activity of countries in terms of their publication output on this theme was examined. Figure 3 shows the top 20 most productive countries in terms of publication output and scientific collaborations during the period 2008-2020. In particular, the left-hand side of the Figure 3 shows the number of articles produced by the authors of different countries and the rate of cooperation of each country's authors with those of other countries, while the right-hand side of the Figure 3 shows the cooperations and networking among researchers working on and studying this subject in different countries.

Figure 3: Country production and country collaboration networks

The authors who have distinguished themselves in terms of the number of publications related to this topic come mainly from China, Malaysia, the USA and Italy. The authors from China and Malaysia have produced the same numbers of articles (35) (not shown here), but the rate of Chinese authorship with other countries is about 46% while the rate of MCP (Multiple Country Pubblications) of Malaysia is 17%. This demonstates that Chinese authors collaborate

extensively with authors from other countries. Italian authors ranked third with 30 papers, and the authorship rate for contributing articles to other authors from other countries is 27.6%. As we can see in the ranking, Italy is the European country that has most significantly increased its publication output in relation to policy evaluation in recent years, indicating that Italian researchers have been promoting statistical methods such as SEM and PLS-PM for the evaluation of policies.

The networking analysis emphasizes the strong collaboration from the Chinese researchers with those from countries such as the USA and Australia (the strength of the collaboration is indicated by the thickness of the links), while European researchers prefer to collaborate with each other. In particular, there is a strong collaboration between France, Spain and Italy. The size of the name of the country is related to the number of works published on the analyzed topic, while the different colors of the countries and of the links represent the clusters that have been formed, as determined by the program algorithm.

Figure 4 highlights some of the most frequently used topics in studies associated with "decision making" and "PLS-PM" during this period. As can be observed from Figure 4, topics relating to evaluation start to appear in 2018 ("life", "prevention", "trasportation"). The frequency increases with the passing of the years. The words most commonly used by researchers who have applied PLS-PM in their studies during the last two years have been "students", "education", "university", "perceptions", "job", "learning", "growth" and "country", topics associated with policy evaluation and decision making issues.

Figure 4: Trend of the topics over the time period

The final figure, Figure 5, shows the keywords considered as themes, classified by different levels of density and centrality in the network of scientific keywords. In the strategic diagram presented in Figure 5, the vertical axis measures the density, namely the strength of the internal links within a cluster represented by a theme, and the horizontal axis the centrality, namely the strength of the internal links within a cluster represented by a theme, and the horizontal axis the centrality, namely the strength of the links between the theme and other themes in the

map (Pourkhani et al., 2019). A thematic map is a very intuitive plot, enabling an analysis of themes according to the quadrant in which they are placed, namely: (1) the upper-right quadrant: motor-themes; (2) the lower-right quadrant: basic themes (3) the lower-left quadrant: emerging or disappearing themes; (4) the upper-left quadrant: very specialized/niche themes (Cataldo et al., 2019).
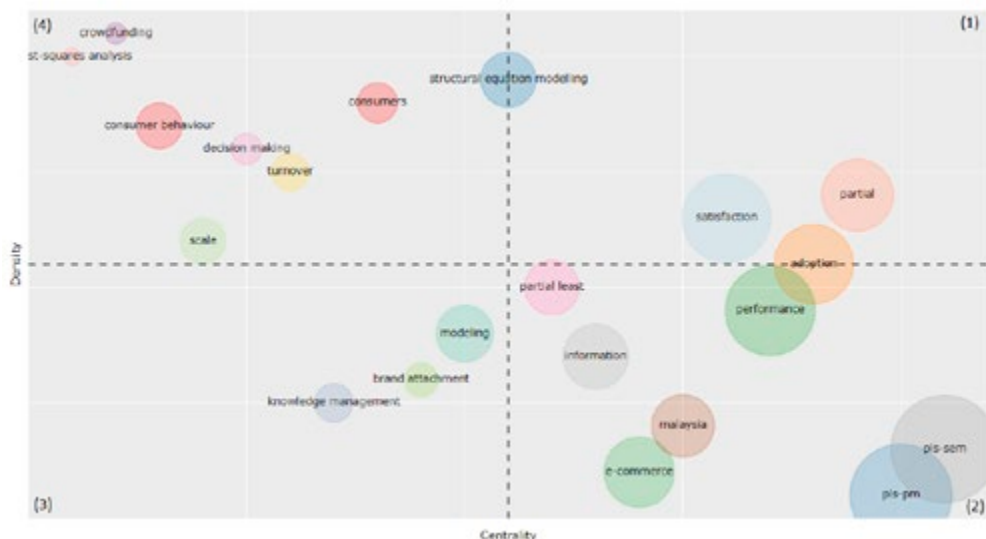

Figure 5: Thematic map

Author's keywords linked to "satisfaction" appear as a motor theme, emphasizing how in the last few years researchers have focused their attention on this theme; words like "trust", "service quality", "customer loyalty", "relationship" and "perceived value" appear in this cluster. Themes with a higher centrality include "pls-sem", "pls-pm", topics that appear ubiquitously in different scientific works and can be considered a common synthesis of the content expressed in the literature. The topic "Malaysia" appears also in this quadrant, while "China", "engagement", "finantial performance" and "risk perception" are other author keywords presented in this cluster, highlighting the predominance of Malaysian and Chinese pubblications in relation to the evaluation theme. Keywords such as "consumer behaviour", "decision-making", "consumer", "crowfunding", despite having a low centrality, have a higher frequency, showing that these themes are considered very specialized topics in these scientific works.

5. Conclusion

The decision making process has always been complex. In the last few years, researchers have been promoting the use of statistical methods such as SEM and PLS-PM for the evaluation of policies, especially in the context of decision making. To better understand and characterize the trend of the scientific pubblication relating to this theme, a bibliometric study of international papers on this subject has been developed highlighting the use of SEM and PLS-PM approaches in policy evaluation during the last twenty years. The data were retrieved from two main databases commonly used by researchers, Scopus and Web of Science, and the analysis of 451 articles was performed using bibliometrix R-Tool (Aria and Cuccurullo, 2017). The results suggest that the interest in research on this topic has increased in recent years, particularly between 2015 and 2019, indicating that this issue has become a significant topic of attention

among researchers in this period. Globally, China is ranked first in terms of production, while in Europe Italian researchers are the most prominent in the promotion of statistical methods such as SEM and PLS-PM for policy evaluation, also collaborating with scholars in Spain and France. The words most frequently used in the last two years by researchers who deal with PLS-PM in their studies have been "students", "education", "university", "perceptions", "job", "learning", "growth" and "country", topics associated with policy evaluation and decision making issues. This study has analysed scientific pubblications on databases being constantly updated. Therefore, a bibliometric analysis regarding an emergent theme may, in a few years, be subject to substantial variations. Furthemore, the present study has analysed a particular theme using two different databases. Despite them being two of the most influential databases, the global perspective could be improved with the inclusion of other databases. However, the results obtained from this analysis may assist researchers in investigation this theme and in focusing on developing the PLS-PM approach for policy evaluation and decision making in many fields of research.

References

Aria, M., Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, Elsevier, **11**(4), pp. 959- 975.

Cataldo, R., Grassia, M.G., Lauro, C.N., Marino, M., Voytsekhovska, V. (2019). A bibliometric study of the global research activity in sustainability and its dimensions. In *International Conference on Data Science and Social Research*, pp. 91-102. Springer, Cham.

Diem, A., Wolter, S.C. (2013). The use of bibliometrics to measure research performance in education sciences. *Research in Higher Education*, **54**(1), pp. 86-114.

Falagas, M. E., Karavasiou, A.I., Bliziotis, I.A. (2006). A bibliometric analysis of global trends of research productivity in tropical medicine. *Acta Tropica*, **99** (2-3), pp. 155-159.

Garfield, E., Sher, I.H., Torpie, R.J. (1964). The use of citation data in writing the history of science. Institute for Scientific Information Inc Philadelphia PA

Maharana, R.K. (2013). Bibliometric analysis of Orissa University of agricultural technology's research output as indexed in Scopus in 2008-2012. *Chinese Librarianship: An International Electronic Journal*, **36**.

Morandi, G., Guido, D., Tagliabue, A. (2015). A bibliometric study of scientific literature on the dietary therapies for epilepsy in Scopus. *Nutritional Neuroscience*, **18**(5), pp. 201-209.

Pourkhani, A., Abdipour, K., Baher, B., Moslehpour, M. (2019). The impact of social media in business growth and performance: a scientometrics analysis. *International Journal of Data and Network Science*, **3**(3), pp. 223-244.

Rodriguez-Soler, R., Uribe-Toril, J., Valenciano, J.D.P. (2020). Worldwide trends in the scientific production on rural depopulation, a bibliometric analysis using bibliometrix R-tool. *Land Use Policy*, **97**, pp. 104787.

Skute, I., Zalewska-Kurek, K., Hatak, I., de Weerd-Nederhof, P. (2019). Mapping the field: a bibliometric analysis of the literature on university-industry collaborations. *The Journal of Technology Transfer*, **44**(3), pp. 916-947.

White, H.D., McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology*, **24**, pp. 119-186.