



Tax Management Research, Bibliometric Analysis Using Keywords and Abstracts

Sonia Elizabeth Ramos-Medina^{1,*}

¹Universidad Autónoma de Sinaloa (México), Universidad de Salamanca, Spain
Email: sonia.ramos@uas.edu.mx

Abstract

The challenges derived from tax operations in different jurisdictions make the management of tax obligations increasingly complex. Taking this into account, a bibliometric analysis is proposed to shed light on the content of the publications from a word analysis to identify research topics in a set of publications and establish relationships between concepts. The findings show a dominance of two large clearly segmented clusters: on the one hand, multinational companies trying to better manage their profits; and on the other, the Tax Agency trying to avoid tax evasion at all costs.

Keywords: Tax management; Multinational Companies; Portfolio Choice; Tax Evasion

1. Introduction

In the examination of bibliographic material in each field, bibliometrics as an information science is fundamental. From a quantitative analysis [1], statistics are obtained that allow us to evaluate the most significant results of a data set [2,3]. Among them we can mention the impact and quality of scientific publications measured using bibliometric indicators [4].

Among the advantages of the method, we have its functionality in mapping the evolution of an area of knowledge, the ease in identifying new developments in research, the establishment of sources of information and, as we have mentioned, the speed in the evaluation of performance in terms of productivity, dissemination, and impact [5].

Our objective is to perform a bibliometric analysis that explores the content of publications on fiscal management. The main issue will be to show the development of this research area according to the information obtained from the Scopus database for the period 1965-2021. The textual statistical analysis considers titles, keywords, and abstracts.

From the set of publications, research concepts on which researchers have focused attention will be identified. This is important because, if different researchers use the same words to a relatively large degree, it could be inferred that they analyse the same problem sets, and thus the same research concepts [6]. Furthermore, if the content of individual publications converges with the results of the word analysis, it means that they also share a focus on the intellectual basis.

Our hypothesis holds that the words contained in both keywords and abstracts reflect the research topics, issues and concepts included in the publications.

The results of the bibliometric analysis show nine clusters or research areas into which the study on fiscal management can be broken down. Then, in a first processing of the data by examining the summaries, the predominance of two large clearly segmented clusters is identified (a binomial that relates to the State and the Firms) and two emerging clusters (related to investments and profits).

Applying Reinert's method confirms the predominance of the two segments plus an emerging one represented by portfolio diversification and insurance.

The sections are organized as follows: section 2 provides a brief review of the literature; section 3 describes the methodology used; section 4 shows the results: it develops graphical visualizations of the bibliographic data; section 5 discusses the findings; and finally, section 6 presents the conclusions.

2. Literature Review

Bibliometrics uses statistical methods and mathematics to analyse patterns in the literature [7], in turn, the review of scientific publications could be used as a channel of science communication [8,9,10]. This type of analysis has a long history, among its pioneers we can mention [11,12,13]. Although bibliometric methods offer a combination of several types of analysis, the most widely used are content and citation analysis.

The content analysis we propose here uses the occurrence of keywords to discern relationships among the publications analysed [14]. In the content analysis of documents, a word analysis allows the identification of research topics, coherences, and differences between them, as well as the establishment of consensus on concepts that relate to individual documents [16]. When certain concepts appear together in different documents, a relationship can be inferred.

Although word frequency analysis was initially used to examine titles of publications, its application was quickly extended to abstracts, citations, among others [16].

With advances in the availability of bibliographic data, computational linguistics and knowledge in databases, there was a shift from bibliometric analysis to data mining [17]. With analytical approaches one can also explore full text and abstracts, and from this, infer knowledge from the body of publications: mapping and profiling research domains [18, 19]; exploring the networks of a research community [20, 21].

In 2021, some of the publications that combine both methods include the following:

It uses the co-occurrence of keywords to analyse human mobility behaviour in the pandemic context [22]. Within her findings, she has identified four major research themes that allow her to propose an agenda for shaping transportation policies and practical responses to COVID-19.

Employing text mining on title content, abstracts, and keywords, [23] describe the evolution and state of the art related to Operations Research and Management Science. Their results yield 26 consolidated research fields.

Perform a bibliometric analysis to evaluate the development of Topic modelling research [24]. Among their findings we can mention that topic modelling has been an important technique in natural and social sciences, and they also point out that the use of social networks and text analysis are areas of growing popularity.

In the study of the relationship between foreign direct investment and economic growth, [25] perform a bibliometric analysis to identify trends and explain economic reality phenomena. This analysis presents a detailed segmentation of the information by subject area, temporal and geographic distribution and content. Their findings confirm the significant relationship between these actors, corroborating some trends that relate it to ecology and environmental protection.

Analyse emerging trends in merchandising research [26]. In addition to identifying related topics, the authors suggest a trend of virtual and augmented reality applied to the e-merchandising context.

Apropos of the anniversary of The Journal of Intellectual Capital (JIC), [27] provide an overview of the journal's history and trend. They retrieve citation and textual data as a basis for using a combination

of bibliometric analysis, altmetric analysis and text mining. Their findings identify measurability domain, financial performance, and Intellectual Capital reporting.

All these studies show the potential of bibliometric analysis to clarify areas of research and identify emerging areas, which is why this exercise contributes to expanding knowledge on fiscal management, which is of great relevance due to its implications in the current economic, political, and social dynamics.

3. Research Methodology

For the construction of the database, we used Scopus records, one of the most important sources of citation and abstract data on linked academic literature in a variety of disciplines.

As a first step for the bibliometric analysis, a general systematic review was performed in the Scopus database. This search is intended to provide a longitudinal review of the state of the art in fiscal management research.

However, our intention is to delimit the search criteria. From the collection of published papers only articles were selected (conferences, book chapter, review, book, book, note, short survey, conference review, errata and retracted are not included). Research fields were limited to the following: Business, Management and Accounting (43.9%), Economics, Econometrics and Finance (43.6%) and Social Sciences (12.5%).

Once we narrowed down the criteria, we found 483 papers that matched our search. After processing and normalizing the data, we obtained the first evidence.

The first recorded article dates from 1965, published by The Modern Law Review under the title: "Finance Act, 1964 Finance (No. 2) Act, 1964 Income Tax Management Act, 1964".

Let us first examine the annual evolution of the number of publications. Figure 1 presents the results. Regarding the trend of publications over time, during the first years the volume has been low. However, the number of publications is accentuated from 2013, then begins a marked growth in the annual volume reaching its highest peak in 2020 with 67 articles.

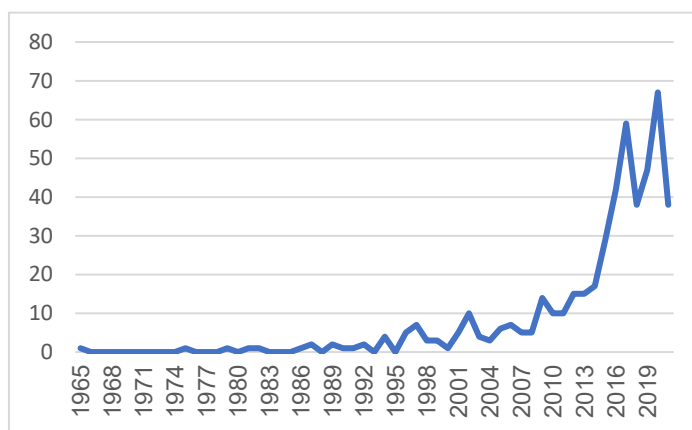


Figure 1: Annual publications on fiscal management

Table 1 summarizes the analysis design, procedures, and types of results. VOSviewer was developed by the Centre for Science and Technology Studies at Leiden University and applied to visualize co-occurrence networks of important terms extracted from the literature [28]. For the conceptual construction of scientific output, co-occurrence based on keywords, titles and abstracts is examined. This will require identifying the most relevant concepts through frequency, proximity, and distance [29].

From the data we have obtained, in this second phase we propose to create a data matrix, called a lexical table composed of the frequencies of the words of the textual content of the abstracts. This step will allow us to lemmatize, i.e., to eliminate words that have no meaning by themselves: articles, prepositions, and determiners. This table will be organized as a matrix (X) of data $I \times J$ where the rows I correspond to the 483 abstracts and the columns J to a selection of 924 active forms (words). With the active forms, a classification analysis was performed on text segments, using the Reinert method. The Iramuteq software was used for this purpose.

In the last phase, the associated words are selected to represent graphically the areas that have originated the most research and in which direction. Through Correspondence Factor Analysis (CFA), the $I \times J$ data set is inspected. This technique allows the simultaneous representation of the classes with a minimum loss of information, starting from the bibliometric data matrix. This facilitates the identification of research topics.

Table 1: Research design for bibliometric analysis

Delimitation Criteria	Required		Results
	Information	Data analysis	
Database: Scopus	1) Abstract and keywords:	Excel: basic statistical data	Descriptive analysis
Search terms: "tax management"	Abstract, author keywords and index keywords.	VOSviewer:	A map based on bibliographic data: co-occurrence of keywords.
Document type: articles	2) Other information: Include references.	Type of analysis: co-occurrence.	A map based on text data: term occurrence map.
Period: 1965-2021		Iramuteq	Clusters and total link strength.
Subject area:		Type of analysis: lexical tables / Reinert method.	Create a map based on lexical tables.
· Business, Management and Accounting		AFC	
· Economics, Econometrics and Finance		Type of analysis: text segment representation	
· Social Sciences		Unit of analysis: all keywords, author keywords, index keywords; abstract.	

4. Interpretation of results

In a general overview, the articles related to fiscal management are summarized as follows: Grant Richardson of Macquarie Business School and Susan C. Borkowski of La Salle University are the

In order to interpret the figure, we must bear in mind that the size of the circles reveals the frequency of the concepts used, while the colors indicate a relationship between terms creating clusters indicating in turn the connections with other clusters.

Nine clusters have been identified that refer to the different sub-disciplines into which the study on "fiscal management" can be broken down. The red cluster refers to the impact of fluctuations in the capital structure, revenue, capital withdrawal and their effects on budgets. It also considers strategies to counteract the effects of these phenomena. Its main concepts: budget, capital structure, efficiency, financial difficulties, fiscal policy, revenues, leverage, management, management compensation, strategy, tax administration, tax management and taxes.

The green cluster is made up of concepts related to how firms manage their earnings through their accounting with knowledge of tax compliance and incentives: corporate income tax, intangible assets, accounting, income transfer, international taxation, auditing, tax enforcement, tax havens, tax incentives and transfer pricing.

The blue cluster relates the exercise of "tax management" to the object of corruption, avoidance, or evasion. It is composed of related terms such as: corporate tax avoidance, revenue management, effective tax rate, financial restrictions, foreign ownership, ownership structure, private companies, tax avoidance and evasion.

The yellow cluster covers insurance-related concepts: audit committee, CEO, innovation, life insurance, reinsurance, taxation. The purple cluster refers to the relationship between tax systems and multinational companies: corporate taxation, empirical analysis, company size, international transfer pricing, multinational companies, profitability, tax system.

The aqua cluster is made up of concepts associated with corporate governance: board of directors, family businesses, company value and tax aggressiveness. The orange cluster could be called agent-principal, with terms such as: agency theory, accounting tax differences, corporate social responses, earnings persistence, earnings quality, mergers, and acquisitions. The brown cluster is composed of tax planning terms: effective tax rates, governance, institutional ownership, tax compliance, tax planning. Finally, the pink cluster relates tax risk terms: firm performance, fiscal, tax risk.

B. Summaries: co-occurrence network and clustering

The review of the summaries shows two large, clearly segmented clusters and two emerging clusters. The results are shown in Figure 3.

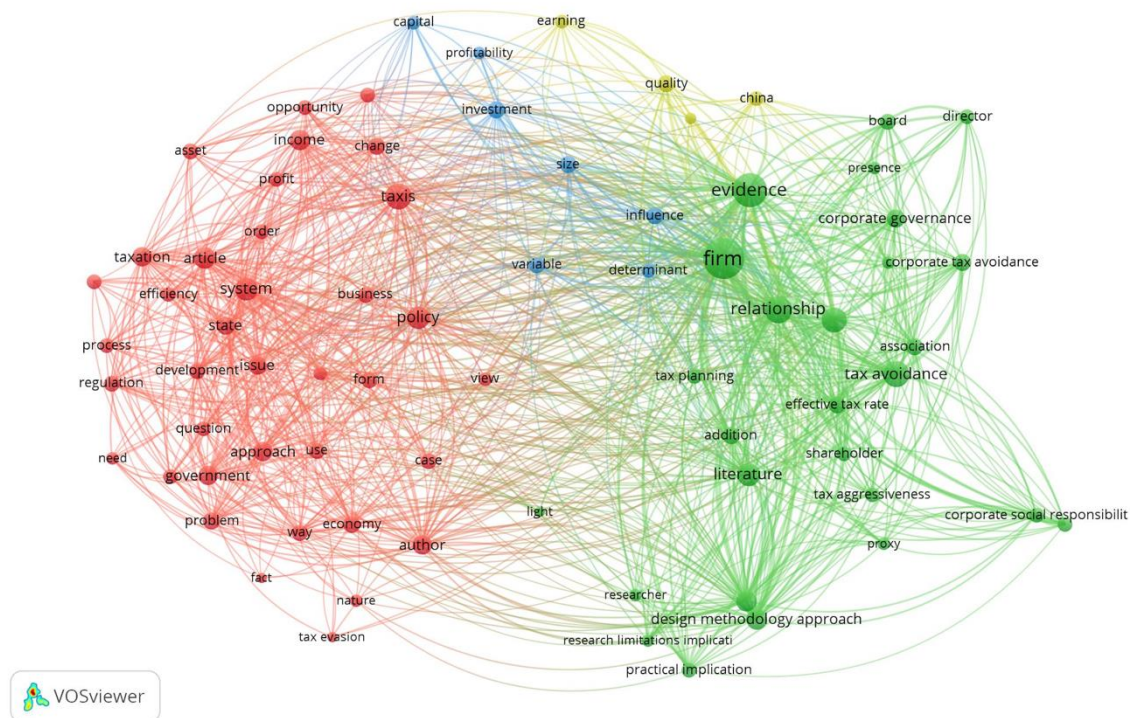


Figure 3: Co-occurrence of terms based on titles and abstract fields.

Note: Counting method: binary count. Minimum number of occurrences of a term: 15. Of the 8827 terms, 125 meet the threshold. For each of the 125 terms, a relevance score was calculated. Based on this score, the most relevant terms were selected. The default option is to select 60% of the most relevant terms (75). Method: strength of association. Weights: occurrences.

The red cluster shows the relationship between accounting efficiency and opportunities to reduce the impact of fluctuations in income or assets, under the framework of a clearly defined system, regulation, tax authority, transfer pricing. On the one hand, companies trying to better manage their profits; and on the other hand, the Tax Agency trying to avoid tax evasion at all costs. Some concepts of the cluster are accounting, business, development, economy, efficiency, government, opportunity, policy, problem, process, profit, regulation, state, system, tax authority, tax evasion, taxes, transfer pricing.

The green cluster has reproduced the concerns related to corporate social responsibility, the implications derived from its study such as the existence of limited evidence, or the participation of directors or CEOs. Some of the related concepts are corporate governance, corporate social responsibility, corporate tax avoidance, design methodology, director, effective tax rate, evidence, originality value, practical implication, relationship, research limitation implication, researcher, shareholder, tax aggressiveness, tax avoidance and planning.

In relation to emerging clusters, blue relates return on capital, investments to the following concepts: Capital, determinant, influence, investment, profitability, size, variable. Finally, yellow relates to revenue, revenue management and quality. Studies in these areas attempt to find empirical evidence of the effect of revenue growth on the possible transfer of wealth from companies to other stakeholders. In other words, from a fiscal management point of view, revenue enhancement is used to transfer the potential welfare of companies to stakeholders. Revenue enhancement techniques are often associated with opportunistic behavior by managers.

C. Textual statistical analysis

Continuing with the systematic review of literature associated with fiscal management, we have elaborated a classification in clusters by analyzing the content of the articles. Once we have processed and normalized the data, we propose to classify them on text segments. The objective is to obtain a data matrix with only active forms. The Reinert method is used for this purpose. The results are shown in Figure 4.

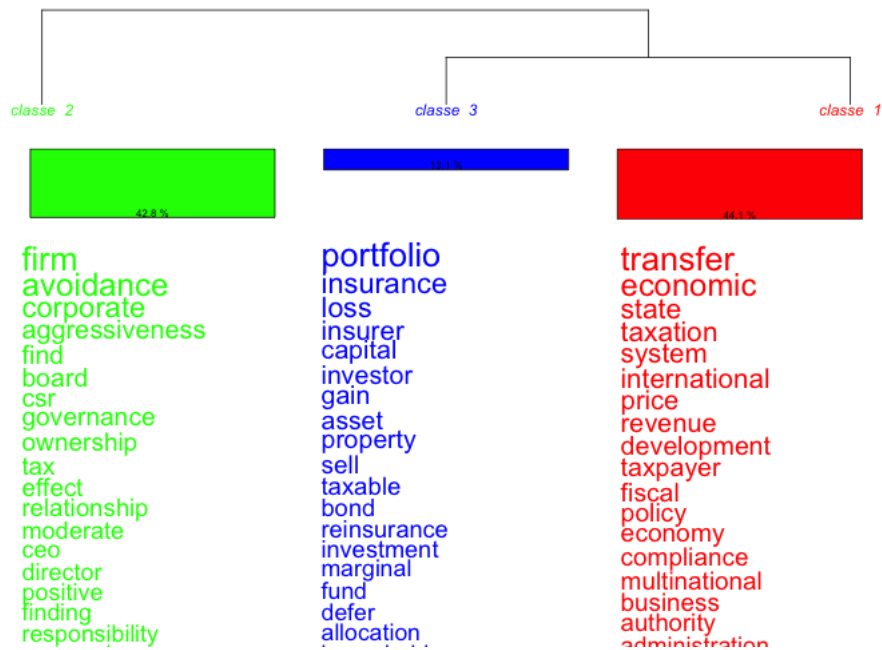


Figure 4: Dendrogram of active shapes

The results are consistent with the clustering in the previous section. The information was classified into three clusters. The green cluster is made up of concepts related to corporate social responsibility. The largest cluster, the red cluster, includes terms related to economics, the state and taxation. It shows a close relationship between this cluster and the blue cluster, two distinct clusters that derive from the same branch. This cluster, being the smallest, could be taken as evidence of emerging research, among which the terms "portfolio" and "insurance" stand out.

D. Correspondence Factor Analysis (CFA)

In order to obtain a simultaneous representation of the classes from the $I \times J$ data set, a correspondence factor analysis (CFA) is performed. Figure 5 shows the factorial representation of the plane 1 (56.11%) - 2 (43.89%), the cumulative inertia amounts to 100% of the variability of the data.

The words indicate their position in the plane, while the size corresponds to their frequencies. The distances between words are inversely related to their similarity; similar being the closest ones. Therefore, the formation of class segments around specific themes indicates the degree of relationship between words and their dimension. The similarity of classes represented with the closest ones could also be identified. The green and red clusters are positively related, but in the opposite direction. The blue cluster, which derives from the red cluster, indicates a different trend: emerging themes.

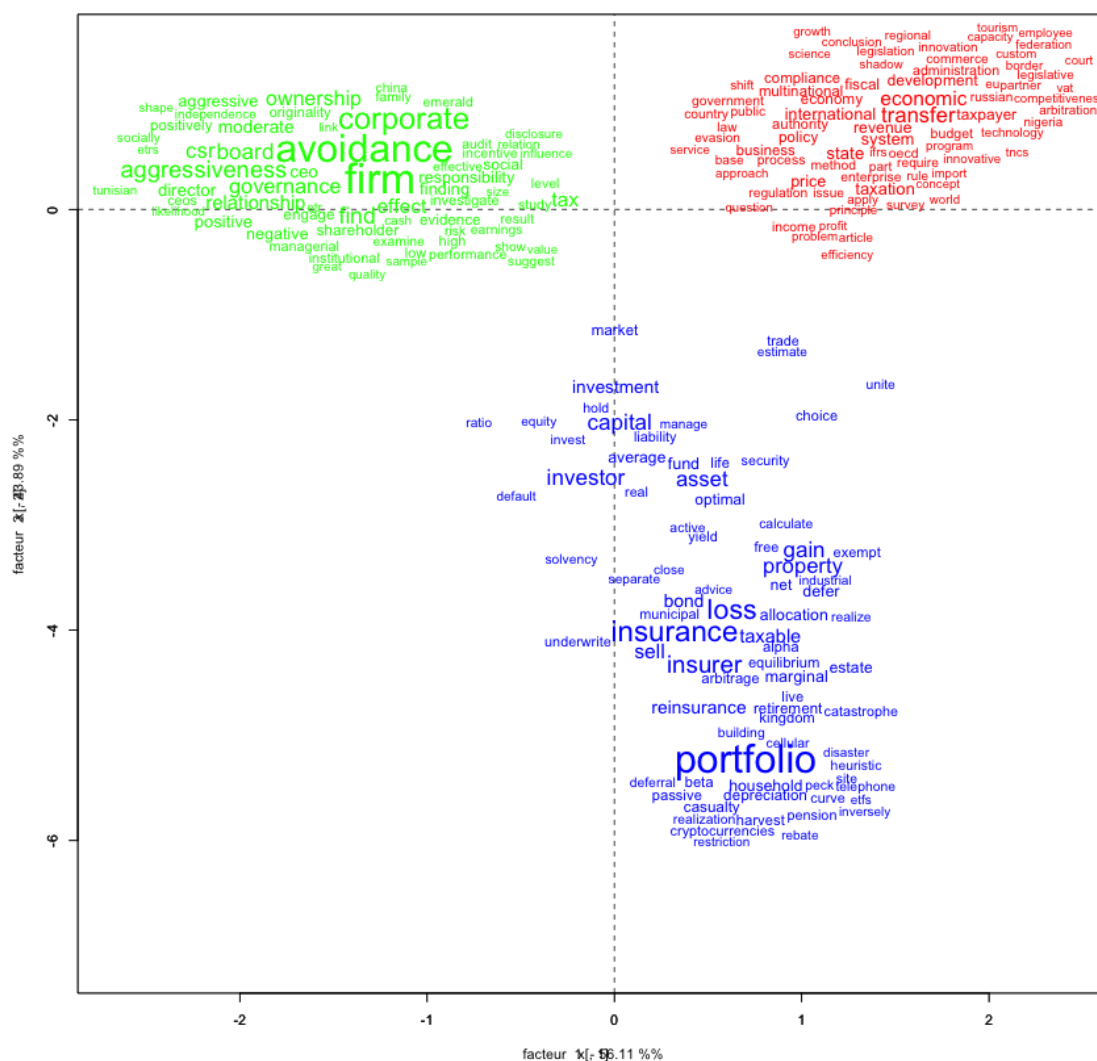


Figure 5: Resulting factorial representation of the flat CFA 1-2

5. Discussion

Our results have highlighted some of the sub-disciplines from which fiscal management could be studied. First, the problem related to capital fluctuations leads to the establishment of strategies to minimize these effects. When referring to a national budget, the withdrawal of capital could be identified as a threat to the balance of payments and its capacity to finance the budget deficit. Fiscal policy will then act as a macroeconomic stabilizer: on the one hand, by establishing a tax policy framework consistent with the national plan; on the other hand, it will consider within its budget fiscal transfers to help vulnerable populations.

From the point of view of private companies, in the absence of capital market pressures, they find incentives to manage profits. In a context where tax regulation strongly influences financial accounting, companies see it as an advantage to report their losses in response to tax incentives that reduce profits to companies with upward management.

Several studies relate tax avoidance, tax management and corporate social responsibility. They analyse the effects of a company's social actions on tax avoidance. The evidence indicates the likelihood that more socially responsible companies show less tax avoidance.

The arguments point to the conception of tax avoidance as a way to prevent the transfer of wealth from companies to governments. Transparency of information could mitigate corporate tax avoidance. Based on these manifestations, tax authorities have focused their efforts on identifying inappropriate transfer pricing used by transnational corporations. Several countries have created partnerships to combat income shifting, improve cross-border information flows and develop conciliatory relationships. Due to the exchange of information, transfer pricing audits have increased.

Although the arm's length principle has been adopted by most economies around the world, establishing guidelines and doctrine on transfer pricing, there is a tendency not to follow free market rules to artificially shift profits or losses from one company to another.

The purpose of transfer pricing is to determine the profits subject to taxation, thus avoiding tax avoidance. The issue is to prevent companies with international operations that generally have deductible expenses in one jurisdiction from transferring them to other jurisdictions with a lower tax burden. Tax Agencies pay special attention to related party transactions when any of them is domiciled in a tax haven.

Therefore, in the study of international tax avoidance, transfer pricing is a mechanism for scrutinizing foreign sales, corporate governance, and ownership by foreign investors. Although these factors are not the only ones to consider, it is clear that transfer pricing is induced by the practices of global corporations, and that strengthening corporate governance will reduce the practice of tax avoidance. As we have shown empirically, there is an increase in research examining the role of corporate governance with Corporate Social Responsibility (CSR) and the possible relationship of the size of these international operations with earnings management.

On the other hand, the concepts of tax avoidance and tax evasion seem somewhat blurred, sometimes used interchangeably. Measuring the level of tax avoidance and evasion, as well as their costs, presents several difficulties. Although it has not been possible to establish precise figures, some estimates have been made indicating that the annual loss of revenue is as high as €1 trillion [32]. Research examining motivations, opportunity and institutional standardization mechanisms agrees that one of the main motivations for breaking the law depends on the possibility of being caught. Since only a small portion of those of tax crimes come to the attention of the authorities [33].

Finally, the above issues are related to the Agency Theory, the principal-agent problem considers those situations that arise when an economic actor depends in a certain sense on the action or morals of another about which it has no particular information.

6. Conclusion

In this exercise, a systematic literature review was proposed, applying a bibliometric analysis with a text mining approach to identify research concepts related to fiscal management.

After applying the delimitation criteria, 483 articles were standardized. The first article for which the Scopus database has a record date back to 1965, however, during the last 15 years the annual volume of publications has increased, reaching its highest peak in 2020 with 67 articles.

Briefly, the most productive authors, most cited articles, institutions, countries, and journals with the highest number of publications were identified.

We have been able to confirm a rapid emergence of different sub-disciplines into which we can break down the study of fiscal management. By examining the network of co-occurrence of keywords and clustering, nine clusters or research areas have been identified. In turn, the study of the summaries and the co-occurrence and clustering networks shows a dominance of two major clusters clearly segmented (binomial relating State - Firms) and two emerging ones related to revenue enhancement and the possible transfer of wealth from firms to other stakeholders.

This trend is confirmed by Reinert's method. The information was classified into three clusters that include concepts related to a) corporate social responsibility, b) state and taxation, and c) a portfolio of assets. So, it could be taken as evidence of emerging research in the area.

It is important to remember that the results are dynamic and could be modified with the emergence of new related topics or with variables increasing their position on a daily basis.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

- [1] Broadus, R. N. Toward a definition of bibliometrics. *Scientometrics*, 12(5-6), 373-379, 1987.
- [2] Goode, C. J., McCarty, L. B., Fink, R. M., Oman, K. S., Makic, M. F., Krugman, M. E., & Traditi, L. Mapping the organization: A bibliometric analysis of nurses' contributions to the literature. *Journal of Nursing Administration*, 43(9), 481-487, 2013.
- [3] Alfonzo, P. M., Sakraida, T. J., & Hastings-Tolsma, M. (2014). Bibliometrics: Visualizing the impact of nursing research. *Online Journal of Nursing Informatics*, 18(1), 1-17, 2014.
- [4] Moed, H. F. New developments in the use of citation analysis in research evaluation. *Archivum immunologiae et therapiae experimentalis*, 57, 13-18, 2009.
- [5] Wallin, J. Bibliometric methods: pitfalls and possibilities. *Basic and Clinical Pharmacology and Toxicology*, 97(5), 261-275, 2005.
- [6] Braam, R. R., Moed, H. F., & Van Raan, A. F. Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for information science*, 42(4), 233-251, 1991.
- [7] McBurney, M. K., & Novak, P. L. What is bibliometrics and why should you care?. In *Proceedings. IEEE international professional communication conference*, 108-114, 2002.
- [8] Pritchard, A. Statistical bibliography or bibliometrics. *Journal of documentation*, 25(4), 348-349, 1969.
- [9] Mingers, J., & Leydesdorff, L. A review of theory and practice in scientometrics. *European journal of operational research*, 246(1), 1-19, 2015.
- [10] Van Raan, A. F. J. (2004). *Measuring Science*. In: Moed, H., Glänzel, W., & Schmoch, U., *Handbook of Quantitative Science and Technology Research*. New York: Kluwer Academic Publishers, 19-50.
- [11] De Solla Price, D. J. (1963). *Little science, big science*. New York: Columbia University Press.
- [12] Garfield, E., Sher, I. H., & Torpie, R. J. (1964). *The use of citation data in writing the history of science*. Philadelphia PA: Institute for Scientific Information Inc.
- [13] Small, H. Cocitation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265-269, 1973.
- [14] Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. From translations to problematic networks: An introduction to co-word analysis. *Social science information*, 22(2), 191-235, 1983.
- [15] Small, H., & Griffith, B. C. The structure of scientific literatures, I: Identifying and graphing specialties. *Science Studies*, 4(1), 17-40, 1974.
- [16] Mullins, N., Snizek, W., & Oehler, K. (1988). The structural analysis of a scientific paper. In Van Raan A. F. J., ed. *Handbook of Quantitative Studies of Science & Technology*. Amsterdam: Elsevier Science Publishers, 85-101.
- [17] Kongthon, A. (2004). *A text mining framework for discovering technological intelligence to support science and technology management*. Georgia Institute of Technology.
- [18] Börner, K., Chen, C., & Boyack, K. W. Visualizing knowledge domains. *Annual review of information science and technology*, 37(1), 179-255, 2003.
- [19] Porter, A. L., Kongthon, A., & Lu, J. C. J. Research profiling: Improving the literature review. *Scientometrics*, 53(3), 351-370, 2002.
- [20] Gerdri, N., Kongthon, A., & Vatananan, R. S. Mapping the knowledge evolution and professional network in the field of technology roadmapping: a bibliometric analysis. *Technology Analysis & Strategic Management*, 25(4), 403-422, 2013.

- [21] Porter, A. L., Garner, J., & Crowl, T. The RCN (Research Coordination Network) experiment: Can we build new research networks?. *BioScience*, 62, 282-288, 2012.
- [22] Benita, F. Human mobility behavior in COVID-19: A systematic literature review and bibliometric analysis. *Sustainable Cities and Society*, 70, 102916, 2021.
- [23] Romero-Silva, R., & De Leeuw, S. Learning from the past to shape the future: a comprehensive text mining analysis of OR/MS reviews. *Omega*, 102388, 2021.
- [24] Li, X., & Lei, L. A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science*, 47(2), 161-175, 2021.
- [25] Cicea, C., & Marinescu, C. Bibliometric analysis of foreign direct investment and economic growth relationship. A research agenda. *Journal of Business Economics and Management*, 22(2), 445-466, 2021.
- [26] Muñoz-Leiva, F., López, M. E. R., Liebana-Cabanillas, F., & Moro, S. (2021). Past, present, and future research on self-service merchandising: a co-word and text mining approach. *European Journal of Marketing*.
- [27] Givi, M. E., Saberi, M. K., Talafidaryani, M., Abdolhamid, M., Nikandish, R., & Fattahi, A. Assessment of the history and trends of “The Journal of Intellectual Capital”: a bibliometrics, altmetrics and text mining analysis. *Journal of Intellectual Capital*, 2021.
- [28] Van Eck, N. J., & Waltman, L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2), 1053–1070, 2017.
- [29] Zupic, I., & Čater, T. Bibliometric methods in management and organization. *Organizational Research Methods*, 18 (3), 429–472, 2015.
- [30] Barbieri, K., Keshk, O. M. & Pollins, B. M. Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science*, 26(5), 471-491, 2009.
- [31] Armstrong, C. S., Blouin, J. L., Jagolinzer, A. D., & Larcker, D. F. Corporate governance, incentives, and tax avoidance. *Journal of Accounting and Economics*, 60(1), 1-17, 2015.
- [32] Taxation and Customs Union - European Commission. A Huge Problem - Taxation and Customs Union - European Commission. [online] Disponible en: < <https://bit.ly/3picQcu> > [Accessed 20 January 2021].
- [33] Ramos-Medina, S. E. (2018). Defraudación fiscal ¿Comportamiento racionalmente oportunista?. In: Ramos-Medina, S. E., & Martín, E., (eds.) *Pasado y presente de la política fiscal en Iberoamérica*, Buenos Aires: L&C pp. 111-120.