Anales del Instituto de Actuarios Españoles, 4ª época, 28, 2022/101-126
ISSN: 0534-3232 / eISSN: 2531-2308/ D.L.: M-3160-1961 / DOI: 10.26360/2022_4

DO THE BUSINESS AND FINANCE SPANISH JOURNALS COVER HIGH-IMPACT TOPICS? A CO-KEYWORD ANALYSIS

¿LAS REVISTAS ESPAÑOLAS DE EMPRESA Y FINANZAS ESTUDIAN TEMAS DE MÁXIMA REPERCUSIÓN INTERNACIONAL? ANÁLISIS DE CO-PALABRAS CLAVES

Pilar Abad[a1], Jorge Cruz[b], M.-Dolores Robles[c]

[a]*Universidad Rey Juan Carlos de Madrid, España.*
https://orcid.org/0000-0002-0558-4055 | pilar.abad@urjc.es
[b]*Universidad Rey Juan Carlos de Madrid, España.*
https://orcid.org/0000-0002-4801-225X | j.cruzde@alumnos.urjc.es
[c]*Universidad Complutense de Madrid e Instituto Complutense de Análisis Económico, Madrid, España.*
https://orcid.org/0000-0001-6299-0300 | mdrobles@ucm.es

Fecha de recepción: 23/11/2022
Fecha de aceptación: 12/12/2022

Abstract

This paper investigates the extent to which the topics covered in Spanish journals from the field of Business and Finance published between 2010 and 2020 correspond to the topics with the highest international impact. To this end, the degree of similarity between the topics published in the journals with the highest international impact according to the JCR and SJR indexes and the topics published in a set of Spanish journals is analysed. The topics are identified using the keywords provided by the authors and the degree of similarity between the research topics is determined using a word co-occurrence analysis. The results show that although Spanish journals are broader in terms of topics, the topics covered in them are among the leading topics according to the JCR and SJR and that there is a direct relationship between the topics and their degree of influence or impact.

[1] Autora para correspondencia: pilar.abad@urjc.es

Keywords: Journal Citation Report; SCImago Journal & Country Rank; keywords; trending topic; co-occurrence.

Resumen

Este trabajo investiga en qué medida los temas tratados en las revistas españolas del ámbito de la Empresa y las Finanzas publicadas entre 2010 y 2020 se corresponden con los temas de mayor impacto internacional. Para ello, se analiza el grado de similitud entre los temas publicados en las revistas de mayor impacto internacional según los índices JCR y SJR y los temas publicados en un conjunto de revistas españolas. Los temas se identifican mediante las palabras clave proporcionadas por los autores y el grado de similitud entre los temas de investigación se determina mediante un análisis de co-ocurrencia de palabras. Los resultados muestran que, aunque las revistas españolas analizan un conjunto más amplio de temas, los temas tratados en ellas están entre los temas con más repercusión según el JCR y el SJR, y que existe una relación directa la repercusión de los temas y su grado de influencia o impacto.

Palabras clave: Journal Citation Report; SCImago Journal & Country Rank; palabras clave; *trending topic*; co-occurrencia.

1. Introduction

The main international multidisciplinary bibliographic databases are Web of Science (WoS), which contains information on more than 12,000 scientific journals and 120,000 conference proceedings, and Scopus, which collects bibliographic information on more than 16,000 journals. Both provide rankings for journals using impact indices based on the number of citations of papers (Journal Citation Report, JCR, and SCImago Journal & Country Rank, SJR). Both indices are commonly used by the international scientific community in evaluating the quality of scientific production. In the case of Spain, the National Agency for Quality Assessment and Accreditation (ANECA), the corresponding agencies of the Autonomous Communities and the National Commission for the Evaluation of Research Activity (CNEAI) also use them for their assessments.

However, despite the increasing inclusion of Spanish journals in these two databases encouraged by their use in accreditations, there are still many journals that are not included in WoS and/or Scopus, that even if they are

included in them, they are not included in the rankings or that are placed in the lower positions in the rankings. One explanation is that in many cases these journals publish research mainly focused on specific matters of the Spanish economy and markets. While representing a significant advance in knowledge, this research could not be of interest to Anglo-Saxon journals, heavily focused on the study of large economies such as the United States.[2] However, the subject coverage of both databases is not complete, especially in Social Sciences and the Humanities (Mongeon and Paul-Hus, 2016). Nor is it in terms of language. According to Vera-Baceta *et al.* (2019), more than 92% of the documents published in WoS and Scopus are written in English. Therefore, researchers working on specific topics and/or who prefer to use Spanish are forced to publish their research in national journals.[3]

In this paper, we investigate the extent to which research in Business and Finance carried out in Spain and published in Spanish journals is in line with leading international research, or to what extent it has a differentiated profile. The aim is to place the research published in these journals in the international context. To this end, we investigated the extent to which the most frequently researched topics in Spanish journals correspond to the trending international topics, i.e., the topics most frequently cited in top journals with a high impact factor.

To identify the topics studied, we analysed the keywords provided by the authors of the papers published between 2010 and 2020. Hu and Zhang (2015) show that author-provided keywords help identify the main topic because they summarise the underlying theme of a paper. Their use is valuable in finding the most relevant trends and topics across disciplines (see, e.g., Li, 2018 and Xu *et al*., 2018), compared to other alternatives, such as searching for key concepts in the abstract or main text of papers (Zhang *et al.*, 2016). We identify the most researched topics or trending topics as those with the highest frequency of occurrence. After identifying the most frequent topics in each journal category, we examined their diversity and conducted a word co-occurrence analysis between categories (Spanish journals versus JCR and SJR) to determine the degree of similarity between trending topics. We also examined their joint distribution to analyse the extent to which similarity is greater among the leading themes.

[2] For example, Argilés and Garcia-Blandon (2011) find that the affiliation of 82% of the authors writing for the highest impact journals according to the JCR in the area of accounting and finance is American. He *et al.* (2021) show that there is a bias in these publications towards Anglo-Saxon research, largely originating in the US, as well as the presence of US researchers on editorial boards.

[3] Arquero *et al.* (2017) find that research in Accounting in Spain has changed its subject matter due to the influence of teacher evaluation systems. There has been a reorientation towards Finance, with a greater presence in journals indexed in the JCR, due to the scarcity of Accounting journals occupying top positions in the ranking.

The results indicate that Spanish journals tend to cover similar topics to the journals with the highest impact according to international rankings (JCR and SJR). Moreover, the most frequently covered topics in both journal categories show a strong positive relationship according to the joint distribution of co-occurrence, i.e., the most frequent topics in one category tend to be the most frequent in the other. This shows that the research trending topics covered in the papers of the most internationally cited journals coincide with the topics that receive the most attention in this group of Spanish journals.

The main contributions of this work are twofold. First, it provides a bibliometric record of research in Business and Finance. This area covers a wide set of topics that include among many others Asset Management; Asset Pricing; Corporate Finance; Financial Markets; Investments; Performance Analysis; as well as Risk Management and Analysis; Political Risk; Systemic Risk; Insurance Economics; Distribution Forecasting; and Mathematical Methods for Insurance and Finances. Journal publications in this area are relevant to a wide range of economic actors, such as financial and insurance market practitioners, policymakers, and regulators, as well as corporate and household risk managers. However, as noted by Farinha *et al.* (2020), despite its importance, this field has received little attention in recent years and is traditionally underrepresented in international rankings (Mongeon and Paul-Hus, 2016, Larivière *et al.*, 2006). Second, the evidence found provides a better understanding of the differences and similarities between two research areas that are recognised to different degrees in research quality assessment systems.

The remainder of the paper is organised as follows: The following section describes the data collection process and outlines the research design. Section 3 presents the main results. Section 4 presents the most relevant conclusions. The paper concludes with an Appendix providing complementary information.

2. Methodology

2.1. Data

In this paper, we focus on topics published in the field of Business and Finance in the period 2010 - 2020. We selected papers published in academic journals as these are the predominant medium for original research in this area. The information for the study was obtained in three steps: 1) selection of journals, 2) collection of scientific papers published in them, and 3) extraction and processing of keywords provided by the authors.

A. Selection of journals

To identify internationally leading research topics, we selected the journals with the highest impact index according to the JCR as the source of keywords. The papers in these journals may well reflect the research priorities in the field (Xiong *et al.*, 2022). In addition, the JCR ranking of journals by their impact is the reference indicator for assessing research quality. We selected the journals in the Business and Finance category according to the 2018 Journal Impact Factor. To obtain the most relevant topics according to this criterion, we selected the journals that are in the first two quartiles of the ranking, i.e. the journals with an impact factor higher than the median value for each category. This procedure results in a set of 55 journals (see list in Table A.1 in the Appendix).

To overcome the problem of limited coverage of JCR in the field of our interest, as pointed out by Mongeon and Paul-Hus (2016) among others, we also selected high-impact journals according to the SJR index based on Scopus, which is much broader and gives us a more comprehensive overview.[4] To avoid overlap, we followed Ennas *et al.* (2015) and excluded journals that were already included in the JCR selection. Consequently, this selection eliminates the highest-impact journals from this database. Thus, the SJR category defined in this way allows the identification of topics with a lower impact than the JCR category.

We used the 2018 SCImago Journal Rank Indicator to rank the journals in the SJR category. To cover the field of Business and Finance, we selected the subcategories "Accounting", "Business and International Management", "Business, Management, and Accounting (miscellaneous)" within the class "Business, Management and Accounting" and the subcategories "Finance", "Economics, Econometrics, and Finance (miscellaneous)" within the class "Economics, Econometrics, and Finance". To make all considered subcategories comparable, we normalise the SJR index of journal *i* in subcategory *j* as:

$$SJR_{ij}^{N} = \frac{SJR_{ij} - \overline{SJR}_j}{S_{SJR_j}} \quad (1)$$

where $\overline{SJR}_j$ and S_{SJR_j} are the average and the standard deviation of the SJR indicator value of all journals in subcategory *j*, respectively. This normalisation allows for a uniform list of subcategories ordered by SJR^N to select the highest-impact journals. Given the wider reach of Scopus, we

[4] JCR and SJR present some differences in their calculation as well as in the journal base considered: WoS versus Scopus. JCR (SJR) considers citations from the last two (three) years and all citations have the same weight (they are weighted by the prestige of the journal they come from, i.e., the journal's ranking position).

decided to select the journals that are in the first decile of the generated ranking list, which in this case consists of 48 journals (see Table A.1 in the Appendix). In this way, we ensure that we select journals that allow us to identify the topics with the highest impact.

Finally, we selected a sample of Spanish journals (hereafter RENIR). The selection of Spanish journals in the field of Social Sciences is large and heterogeneous. Authors such as Rodríguez-Yunta and Giménez-Toledo (2013) have highlighted that Spanish Social Science journals are very diverse in terms of the parameters that determine the quality of a publication (periodicity, peer review, editorial board, etc.). Therefore, to select the most relevant journals with a homogeneous quality[5], we consulted the FECYT catalogue of Spanish scientific journals, from which we selected the subcategory of Economics Journals within the Humanities and Social Sciences category[6], and the IN-RECS, a citation-based ranking, from which we selected the category Social Sciences: Economics.[7] After merging the two lists of journals and eliminating overlaps, we obtain an initial selection of 55 journals. To achieve homogeneity, we selected among these journals a sample of them that are included in WoS and Scopus. In this way, the finally selected journals meet homogeneous quality standards and comparability of data, as both WoS and Scopus carry out a rigorous selection process for journals before including them in their databases. Finally, we select a group of 36 journals (see Table A.2 in the Appendix). Some of them are listed in the JCR and/or SJR and others are not listed in these rankings.

B. Compilation of the scientific papers

After identifying the journals in the three categories (JCR, SJR, and RENIR), we obtained all papers published in these journals between 2010 and 2020. For this purpose, we use the bibliographic manager EndNote.[8] This manager allows us to import bibliographic references, check duplicate entries, organize the search according to different criteria, and export searches and selected fields.

[5] The quality of a publication is determined by the fulfillment of guidelines related to its editorial quality, the scientific rigour of its contents, its visibility, and its dissemination, thereby guaranteeing the scientific rigour of the papers published in it (FEYCT, 2020).

[6] https://www.fecyt.es/es/noticia/fecyt-publica-un-ranking-de-visibilidad-e-impacto-de-revistas-cientificas-espanolas-de

[7] https://web.archive.org/web/20141013160418/http:/ec3.ugr.es/ec3/Economia.html

[8] EndNote is one of the most widely used bibliographic managers among researchers (Gallegos *et al.*, 2017). It allows references to be organised according to the database to which each paper belongs, as well as to group the bibliography according to their criteria (Cordón-García and Alonso-Arévalo, 2009).

The information was extracted from its primary source, i.e. the journal databases WoS and Scopus. Both allow the searching and filtering of papers according to various criteria such as geographical, linguistic, thematic, etc. (more details in Chaparro-Martínez and D'Armas-Regnault, 2017). In our case, we filtered by journal ISSN and year of publication and downloaded publications in .ciw (WoS, compatible with different applications) or .ris (research information systems file extension) format. Information is collected on 8 variables: Title of paper, author/s, keywords, abstract, journal title, year of publication, journal number, and journal volume. The final database, described in Table 1, consists of 60,775 papers: 33,608 from JCR, 20,305 from SJR, and 6862 from RENIR.

Table 1
Description of papers, journals, and keywords by categories

	Journals	Papers	Papers by journal and year	Keywords	Keywords by journal and year	Filtered keywords
Research in international journals with the highest impact						
JCR	55	33,608	55.6	318,193	525.9	38,606
SJR	48	20,305	38.5	249,101	471.8	36,957
Research in Spanish journals						
RENIR	36	6,862	16.9	46,792	115.0	15,776
Total	140	60,775	42.2	614,086	42.64	91,339

Note: Authors' elaboration. Information from 2010 to 2020 (11 years). primary source of information: Web of Science and Scopus.

We note that JCR journals tend to publish more papers per year (the average is about 56), while RENIR journals publish the fewest (17). The number of keywords by journal and year is also much lower in RENIR than in the indexed journals, with the JCR journals containing more keywords.

C. Processing the keywords

After downloading the information, errors are corrected and the keywords are homogenised in the three databases. This thorough processing is crucial for further analyses and calculations. In addition to detecting and correcting errors (incomplete words and/or words without a clear meaning), homogenising plurals and singulars of the same word, homogenising differences in the translation of keywords between English and Spanish, etc., we take into account that different authors may use different words with the same meaning,

so we also merge these synonymous terms. This filtering process is done both mechanically and manually, first independently in each category and then by matching the information between the three databases. After cleaning, the final number of keywords is 91,339, of which 38,606 are from JCR journals, 36,957 from SJR, and 15,576 from RENIR (see Table 1).

As an illustrative example of the collection and depuration of keywords process applied, we analyze the frequency of keywords in one Spanish journal in the RENIR group with a "word cloud" (see Figure 1). We chose *Anales del Instituto de Actuarios Españoles,* and gather the most frequent keywords from the papers published in it from 2010 to 2020. The font size of each keyword in Figure 1 reflects their corresponding frequency of occurrence. We can observe that the topics most covered in this journal according to the most frequent keywords are *Solvency II, Risk, Social Security, Mortality, Chain-Ladder, Copula, Premium and Reserve risk.*

Figure 1. Keyword frequency in Anales del Instituto de Actuarios Españoles
Note: Word cloud analysis *from 2010 to 2020*.

2.2. Research design

In this paper, we propose a two-stage methodology based on bibliometric analysis. This type of analysis is often used to identify the distribution of research in a given field (see Donthu *et al.*, 2021, Farinha *et al*., 2020 and Cheng *et al*., 2020 for some recent examples). In our case, the aim is not so much to find out which topics are most frequently covered in the field of

Business and Finance, but rather to determine the degree of correspondence of leading topics between two categories of journals according to their degree of international influence.

In the context of this literature, we conducted the research in three phases: (1) identifying the most frequently occurring keywords representative of the most common research topics in each group of journals and examining their characteristics, (2) determining the degree of similarity between them, and (3) examining their joint distribution. To this end, we focused on the study of keywords and their co-occurrence among different groups of journals (co-keyword analysis). As pointed out by Li *et al.* (2020), the co-occurrence of keywords in different papers implies a thematic link between them and makes it possible to identify the trending topics of research in a field or discipline. We, therefore, consider that the analysis of the co-occurrence of keywords is the appropriate tool to look for links between the topics studied in the papers of Spanish journals and those published in the journals with the greatest influence at the international level.

The individual analysis for each journal category was carried out by examining the distribution of keywords observed in each journal category and their relative frequency (or "abundance"), and applying a set of diversity indicators commonly used to study biodiversity in biology studies (see Magurran, 2004). On the other hand, the analysis of the similarity between topics covered in two different journal categories is carried out using a set of indicators that describe the statistical congruence between topics published in the journals of both categories. These similarity indicators are commonly used in biology to assess species overlap (Tuomisto, 2017) and have also been widely used in scientometric analyses (Stirling, 2007). Table A.3 in the Appendix contains the mathematical expressions for both types of indicators (diversity in Panel A and similarity in Panel B).

3. Results

3.1. Selection of trending topics based on keywords

We have independently analyzed the JCR and SJR keyword databases to identify the top topics in each of them by analyzing their occurrence. The aim is to identify the most frequent keywords that signal the most researched topics in the period under consideration. For this purpose, a hierarchy is created according to the frequency of occurrence in each database. The most relevant

topics or trending topics are determined by those keywords whose occurrence is greater than 1‰ concerning the number of papers in which they could occur (total number of papers considered within a category).[9] Thus, in the JCR category, 999 different keywords are identified with an occurrence of 31 or more, while in the SJR category, a total of 957 keywords are identified with an occurrence of 27 or more. In both cases, this selection means that the keyword with the lowest occurrence in each base has a relative frequency of 0.02% of the total number of keywords. To improve comparability, we applied the same criteria to the RENIR category and retained keywords with a relative frequency of 0.02% or more of the total number of final keywords in the database. Thus, the number of unique keywords in this category is 1190 with an occurrence of 6 or more. A descriptive analysis of the final keyword databases can be found in Table 2.

Table 2
Description of the keyword sample by category

	JCR		**SJR**		**RENIR**	
	Keywords	*Relative frequency*	*Keywords*	*Relative frequency*	*Keywords*	*Relative frequency*
Total	179,781		135,074		24,621	
Different	999		957		1,190	
Mean	180.0	0.10%	141.1	0.10%	20.7	0.08%
Median	76	0.04%	65	0.05%	11	0.04%
Maximum	4,107	2.28%	3,742	2.77%	373	1.51%
Minimum	31	0.02%	27	0.02%	6	0.02%
SD	362.2	0.20%	247.9	0.18%	30.2	0.12%
CV	2.01		1.76		1.46	

Note: Authors' elaboration. Information from 2010 to 2020. Primary sources of information: Web of Science and Scopus. Total indicates the sum of the absolute frequency of keywords in a category, while Different indicates the number of distinct keywords in a category. Relative frequency is the number of times a keyword is repeated out of the total number of keywords in the category. SD denotes standard deviation and CV coefficient of variation.

The results in Table 2 provide interesting information that allows us to make a first comparison between the three categories. As we mentioned earlier, JCR is the category with the highest number of keywords, followed by SJR and

[9] One per thousand of the 33,608 papers collected in the JCR category are 34 papers. In the case of the 20,305 SJR papers, there are 20 papers. Consequently, any keyword appearing in at least 34 (20) is considered a JCR category leading topic (SJR).

RENIR. The average occurrence is also higher in this database, 180, compared to 141.1 in SJR or 20.7 in the case of RENIR. This result seems to indicate that papers in journals indexed in the first places by impact in JCR are more likely to cover the same topic than papers in the other categories, i.e. they show a higher degree of specialization. The differences in the analysis of mean and median frequencies are not so great, with similar numbers in all three categories. The highest frequency is found in the SJR category, where a keyword occurs at 2.77%, compared to 2.28% in JCR or 1.77% in RENIR. In terms of variability, it can be seen that the JCR category has the highest dispersion, both in terms of standard deviation and coefficient of variation, while RENIR has the lowest variability. This suggests that the attention given to leading issues in the JCR category is more variable than in the other categories.

Figure 2 shows the cumulative frequency of keywords in each category. We note that the most frequently occurring words in the JCR category have a higher frequency than those in SJR and RENIR. In RENIR, the frequency of the most frequent keywords is lower, indicating a greater variety of topics or less specialization than in the JCR and SJR journals.

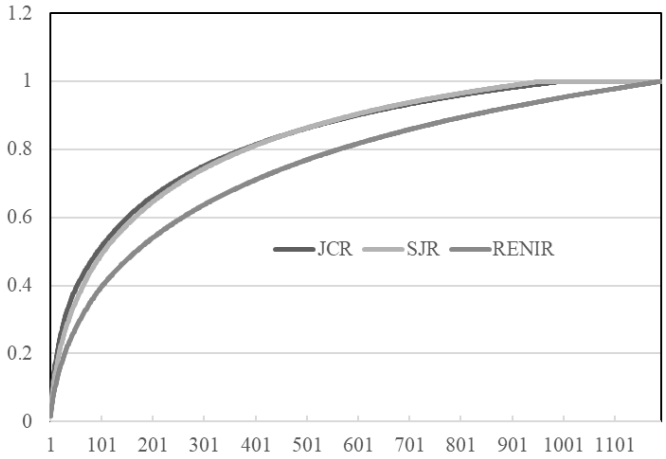

Figure 2. Cumulative frequency of Keywords

We deepen the analysis of three categories (JCR, SJR, and RENIR) in terms of thematic diversity by analyzing the diversity indicators listed in Panel A of Table A.3 in the Appendix. These take into account both the diversity of topics (their number) and their relative frequency (their abundance). In particular, we use the Shannon entropy index and the transformed version of it (which

allows us to determine effective diversity), the Berger-Parker index (which takes into account the higher frequency in each database), and the Gini-Simpson index (see Tuomisto, 2017 for a detailed description). The results are presented in Table 3.

Table 3
Thematic diversity indicators

	JCR	SJR	RENIR
Shannon's Entropy	6,110	6,179	6,550
Gini-Simpson	0.995	0.996	0.997
Shannon transformed	450.4	482.7	701.0
Berger-Parker	2.28%	2.77%	1.77%

Note: Authors' elaboration. Information from 2010 to 2020. Primary sources of information: Web of Science and Scopus.

Based on the Shannon entropy index values (values above 4 are considered to indicate high diversity), we found a high diversity of topics in all categories. The values are difficult to compare between categories because this index is influenced by the sample size. To compare categories, we use the transformed Shannon index, which gives us the "effective number" of topics, defined as the number of equally frequent topics needed to obtain the average observed frequency (see Hill, 1973, Tuomisto, 2017). According to this indicator, diversity is higher in the case of RENIR (701 topics compared to 482.7 for SJR and 450.4). On the other hand, RENIR is the category with the lowest value for the Berguer-Parker indicator (1.77% compared to 2.28% and 2.77% for JCR and SJR respectively). The values of the Gini-Simpson index are similar in all three categories, although slightly higher in the case of RENIR. In summary, these indicators of thematic diversity show that although there are no major differences in the diversity of topics covered in the three journal categories, RENIR has the greatest diversity and JCR the greatest specialization.

3.2. Thematic Similarity Analysis

We analyse here the co-occurrence of the selected keywords between the categories JCR and SJR with the keywords of the category RENIR. Matching keywords indicate a thematic connection between the two journal categories and allow us to determine the degree of similarity between the research topics they address. In other words, it is possible to determine the degree to which

the research topics published in Spanish journals correspond to the research in Business and Finance with the highest impact at the international level.

To assess the degree of similarity between the JCR (or SJR) category and RENIR, we used the indicators described in Panel B of Table A.3 in the Appendix, based on both the occurrence (degree of similarity, Jaccard, Sorensen-Dice, and Russel-Rao) and the frequency (cosine distance, Horn, and Morisita-Horn). We also used these indicators to determine the similarity between JCR and SJR. The results are presented in Table 4.

We find that 488 of the keywords in the JCR category and 495 in SJR are also included in RENIR. The Similarity-1 indicator suggests that 48.8% of the JCR keywords are also found in RENIR, i.e., 48.8% of the leading topics covered in the highest impact journals according to JCR are also covered in Spanish journals. 41% of the topics in RENIR are also explored in JCR (see similarity-2). These percentages of similarity are 51.7% and 41.6% in the case of SJR, i.e., the similarity of the topics explored in RENIR with those of SJR is slightly higher than with those of JCR. Note that the similarity between JCR and SJR is much lower, with 394 common words and similarity degrees 1 and 2 of 41.2% and 39.4%, respectively. Recall that the SJR category excludes many journals that are already included in the JCR selection to avoid overlap. Figure 3 illustrates the degree of similarity between the three categories using the co-occurrence of keywords in a Venn diagram. We can see that 319 topics are shared by the three categories.

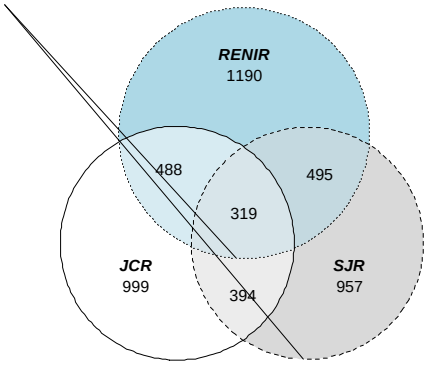

Figure 3. Degree of similarity between the three categories.

The Jaccard, Sorensen-Dice, and Russel-Rao values are in the same direction, i.e., they indicate a greater similarity of topics addressed by RENIR with SJR,

followed by RENIR with JCR, with the JCR and SJR categories showing the least similarity.

Indicators based on keyword abundance point in the same direction. These indicators show a medium or low level of similarity and confirm that the similarity in the topics covered is higher between RENIR and SJR. The value of the indicators is 0.789 (cosine distance), 0.292 (Horn), and 0.753 (Morisita-Horn), while the values in the comparison between RENIR and JCR are 0.609, 0.266, and 0.560, respectively. In JCR and SJR, these indicators have higher and lower values (0.66, 0.258, and 0.656, respectively), which does not suggest that the similarity between JCR and SJR is lower than that between RENIR and JCR when abundance is considered.

Table 4
Indicators of topic similarity between categories

	JCR vs RENIR	**SJR vs RENIR**	**JCR vs SJR**
Panel A. Binary indicators			
Co-occurrences	488	495	394
Similarity-1	0.488	0.517	0.412
Similarity-2	0.410	0.416	0.394
Jaccard	0.287	0.300	0.252
Sorensen-Dice	0.446	0.461	0.403
Russel-Rao	0.223	0.231	0.201
Panel B. Frequency-based indicators			
Cosine	0.609	0.789	0.660
Horn	0.266	0.292	0.258
Morisita-Horn	0.560	0.753	0.656

Note: Authors' elaboration. Information from 2010 to 2020. Primary sources of information: Web of Science and Scopus.

An analysis of the joint distribution of keywords among categories follows. To do this, we grouped the keywords into quartiles according to their abundance within each category. Thus, in group C1, we count the first quartile, i.e., 25% of the keywords with the highest occurrence within the corresponding category and, consequently, the most frequent keywords that determine the trending topics. C2 denotes the second quartile (the next 25% of keywords ranked by occurrence), C3 the third quartile, and C4 the fourth quartile. We then created a 4x4-dimensional matrix showing the joint distribution of occurrence, i.e., the joint occurrence in each quartile that

crosses, and we tested independence in the distribution of keywords between categories using Pearson's chi-square test:

$$Q_R^B = \sum_{j=1}^{4}\sum_{i=1}^{4}\frac{(O_{i,j}-E_{i,j})}{E_{i,j}} \sim^{si\ H_0} \chi^2_{(4-1)(4-1)} \quad (6)$$

where Oi,j is the frequency of co-occurrence of keywords in quartile Ci of the RENIR category and quartile Cj of base B (i =1,...,4; j =1,...,4; B=JCR, SJR), Eij is the frequency of co-occurrence expected under the null hypothesis of independence. If the most frequently published research topics in RENIR match the international trending topics, we would expect the highest frequency of co-occurrence to occur at the intersection of the C1 quartiles, which refutes the null hypothesis of independence.

Table 5 shows the results of the comparison between JCR and RENIR. We find that about 23% of the coincidences between the keywords with the highest frequency in JCR and RENIR, i.e., fall in the C1 quartile of both. This number increases to 50% when we consider co-occurrence in C1+C2 quartiles ($\sum_{j=1}^{2}\sum_{i=1}^{2} O_{i,j} / \sum_{j=1}^{4}\sum_{i=1}^{4} O_{i,j}$). These results show that the most frequently covered topics in Spanish journals include, to a large extent, the topics included in the journals with the highest impact according to the JCR. Furthermore, if we look at the topics with the highest occurrence in each category, i.e. those in the C1 quartile, we find that about 56% are ranked C1 in RENIR and JCR (55.7% and 56% respectively), indicating that the most researched topics in both categories have a large overlap. The results indicate a strong positive correlation between the most discussed topics in the two categories, which is confirmed by the result of the independence test. The chi-square test rejects the hypothesis of independence. Consequently, we found a significant direct relationship between the distributions of the research topics according to their importance in both categories.

Table 5
Joint distribution of co-occurrences: JCR compared to RENIR

JCR \ RENIR	C1	C2	C3	C4	*Total*
C1	112	43	30	15	*200*
C2	51	38	32	21	*142*
C3	18	24	22	19	*83*
C4	20	21	12	10	*63*
Total	*201*	*126*	*96*	*65*	***488***

Chi-squared	39.94***
(p-value)	(0.000)

Note: Each element (Oi,j) shows the co-occurrences of the keywords in quartile i of the JCR category and quartile j of the RENIR category. The last row (column) shows the total number of co-occurrences of the JCR category (RENIR) by quartile. Chi-square is Pearson's test of independence for the null hypothesis of independence between categories. *** indicates significance at 1%.

The matrix of co-occurrence of SJR and RENIR (Table 6) also shows that a high percentage of the trending topics (keywords in C1) according to SJR (about 61%) are also in the C1 quartile of the RENIR category (57.8%) and a high percentage of co-occurrence (24.4%) are in the first quartile of both categories (55.5% when we consider C1+C2). The result of the chi-square test for independence rejects the null hypothesis in this case as well. This indicates that there is a positive and significant correlation between the impact of trending topics in Spanish journals and the journals with the highest impact according to the SJR ranking.

Table 6
Joint distribution of co-occurrences: SJR compared to RENIR

RENIR / SJR	C1	C2	C3	C4	*Total*
C1	121	49	20	10	*200*
C2	61	44	27	18	*150*
C3	17	25	30	15	*87*
C4	11	17	15	15	*58*
Total	*211*	*135*	*92*	*58*	***495***
Chi-squared	76.84***				
(p-value)	(0.000)				

Note: Each element (Oi,j) shows the co-occurrences of the keywords in quartile i of the JCR category and quartile j of the RENIR category. The last row (column) shows the total number of co-occurrences of the JCR category (RENIR) by quartile. Chi-square is Pearson's test of independence for the null hypothesis of

independence between categories. *** indicates significance at 1%.

In summary, the results of the thematic similarity analysis show that although there is an alignment between the topics covered in Spanish journals and the topics with the highest international impact, it is far from complete. RENIR journals show a greater presence of trending topics identified in the SJR category than in the JCR. Given the process used to create the categories, these topics are of lesser importance than those in the JRC. However, a large number of research topics are common to all three categories. Similarly, the topics with the greatest impact in the JCR and SJR categories are often also those with the greatest impact in RENIR.

4. Conclusions

Assessing the quality of scientific output is a major challenge that has been much discussed recently. The assessment processes are based on indicators of the influence or impact of publications and their presence in international rankings, mainly JCR and SJR. Publications that are not included in these rankings or are placed in the lower positions in the rankings are valued less, regardless of the type of research they do. This paper contributes to this debate by providing evidence on the degree of coincidence between the most analyzed research topics in a set of Spanish journals (RENIR) and the leading topics in journals with the highest international impact in the field of Business and Finance according to JCR or SJR.

We examined the coincidence of keywords in papers published between 2010 and 2020 in 36 Spanish journals from the field, compared to papers published in 103 journals that hold the top positions in the JCR and SJR rankings. The results show that RENIR has a higher degree of subject diversity than JCR and SJR. The evidence on the similarity of the topics covered in the RENIR journals with the other categories indicates an intermediate degree of similarity, which is slightly lower in JCR than in SJR, i.e., it increases with decreasing impact.

On the other hand, the most frequently discussed topics show a strong positive relationship according to the joint distribution of common occurrences between the categories. A high proportion of the most frequent topics in one category are also the most frequent in the other category. This shows that the research trends or trending topics covered in the papers of the most

internationally cited journals largely coincide with the topics that Spanish journals pay the most attention to.

The evidence presented in this paper is of interest to academics in this research area (Business and Finance) and academia in general, as it highlights the value of research published in national journals in recent years. These journals also deal with topics of international importance, although they show greater diversity in terms of including topics of more local interest. This result also helps insurance and financial market practitioners and policy makers in choosing where to look for recent research relevant to their interests.

Finally, it is worth pointing out some limitations and future extensions that result from the work carried out. One of the possible limitations is the specification of the research topics based on the keywords provided by the authors. A natural extension would be to start the analysis by identifying themes based on a textual analysis of the content of the papers. It would also be interesting to explore the temporal dimension, which would examine the evolution of the degree of similarity and dependence between themes over time, as well as the geographical dimension, which would provide measurements by region or research center, and finally the individual dimension, which would provide results by journal.

5. Acknowledgments

This research has been funded by the Spanish Ministry of Science and Innovation / Spanish State Research Agency (PID2020-113367RB-I00).

6. References

Argilés, J. M.; Garcia-Blandon, J. (2011). Accounting research: A critical view of the present situation and prospects. *Revista de Contabilidad*, *14*(2), 9-34. https://doi.org/10.1016/S1138-4891(11)70026-7

Arquero, J. L.; Jiménez-Cardoso, S. M.; Laffarga-Briones, J. (2017). Research Patterns of successful accounting academics in Spain. *Spanish Journal of Finance and Accounting*, *46*(3), 327-368. https://doi.org/10.1080/02102412.2017.1287462

Chaparro-Martínez, E.; Álvarez-Muñoz, P.; D'Armas-Regnault, M. (2017). Gestión de la información: Uso de las bases de datos Scopus y Web of Science con fines académicos. *Universidad Ciencia y Tecnología*, *20*(81).

http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1316-48212016000400003

Cheng, Q.; Wang, J.; Lu, W.; Huang, Y.; Bu, Y. (2020). Keyword-citation-keyword network: a new perspective of discipline knowledge structure analysis. *Scientometrics*, *124*(3), 1923-1943. http://doi.org/10.1007/s11192-020-03576-5

Cordón-García, J. A.; Martín-Rodero, H.; Alonso-Arévalo, J. (2009). Gestores de referencias de última generación: análisis comparativo de RefWorks, EndNote Web y Zotero. *El profesional de la información*, *18*(4), 445-454. https://doi.org/10.3145/epi.2009.jul.14

Donthu, N.; Reinartz, W.; Kumar, S.; Pattnaik, D. (2021). A retrospective review of the first 35 years of the International Journal of Research in Marketing. *International Journal of Research in Marketing*, *38*(1), 232-266. http://doi.org/10.1016/j.ijresmar.2020.10.006

Ennas, G.; Biggio, B.; Di Guardo, M. C. (2015). Data-driven journal meta-ranking in business and management. *Scientometrics*, *105*(3), 1911-1929. http://doi.org/10.1007/s11192-015-1751-y

Farinha, L.; Sebastião, J. R.; Sampaio, C.; Régio, M. (2020). Flipping the Best Journals in Accounting, Business, and Finance: What Fruits Do We Collect? In *Handbook of Research on Accounting and Financial Studies*, 1-21. IGI Global. https://doi.org/10.4018/978-1-7998-2136-6.ch001

Gallegos, M. C.; Peralta, C. A.; Guerrero, W. M. (2017). Utility of Bibliographic Managers in Organization of Information for Research Purposes. *Formación universitaria*, *10*(5), 77-87. https://scielo.conicyt.cl/pdf/formuniv/v10n5/art09.pdf

He, F.; Fu, T.; Lucey, B. (2021). Geographic diversity in academic finance editorial boards-A discussion. *Finance Research Letters*, 43, 102006.

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54, 427–432.

Hu, J.; Zhang, Y. (2015). Research patterns and trends of Recommendation System in China using co-word analysis. *Information processing & management*, *51*(4), 329-339. https://doi.org/10.1016/j.ipm.2015.02.002

Larivière, V.; Archambault, É.; Gingras, Y.; Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*(8), 997-1004. https://doi.org/10.1002/asi.20349

Li, M. (2018). Classifying and ranking topic terms based on a novel approach: Role differentiation of author keywords. *Scientometrics*, *116*(1), 77–100. http://10.1007/s11192-018-2741-7

Li, Q.; Zhang, H.; Hong, X. (2020). Knowledge structure of technology licensing based on co-keywords network: A review and future directions.

International Review of Economics & Finance, 66, 154-165. http://10.1016/j.iref.2019.11.007

Magurran, A.E. (2004). *Measuring biological diversity*. Blackwell, Malden http://www.bio-nica.info/Biblioteca/Magurran2004MeasuringBiological.pdf

Mongeon, P.; Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics, 106*(1), 213-228. https://doi.org/10.1007/s11192-015-1765-5

Rodríguez-Yunta, L.; Giménez Toledo, E. (2013). Fusión, coedición o reestructuración de revistas científicas en humanidades y ciencias sociales. *El profesional de la información, 22*(1), 36-45. https://doi.org/10.3145/epi.2013.ene.05

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface, 4*(15), 707-719. https://doi.org/10.1098/rsif.2007.0213

Tuomisto, H. (2017). Defining, Measuring, and Partitioning Species Diversity, Reference Module in Life Sciences. Elsevier.

Vera-Baceta, M. A.; Thelwall, M.; Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121, 1803–1813. https://doi.org/10.1007/s11192-019-03264-z

Xiong, T.; Zhou, L.; Zhao, Y.; Zhang, X. (2022). Mining semantic information of co-word network to improve link prediction performance. *Scientometrics*, 1-24. https://doi.org/10.1007/s11192-021-04247-9

Xu, J.; Bu, Y.; Ding, Y.; Yang, S.; Zhang, H.; Yu, C.; Sun, L. (2018). Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics*, *117*(2), 973–995. http://doi.org/10.1007/s11192-018-2897-1

Zhang, J.; Yu, Q.; Zheng, F.; Long, C.; Lu, Z.; Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, *67*(4), 967–972. https://doi.org/10.1002/asi.23437

Appendix A

Table A.1. Selected JCR and SJR journals

		JCR	SJR
1	ABACUS-A JOURNAL OF ACCOUNTING FINANCE AND BUSINESS STUDIES	x	
2	ACADEMY OF MANAGEMENT ANNALS		x
3	ACADEMY OF MANAGEMENT JOURNAL		x
4	ACADEMY OF MANAGEMENT PERSPECTIVES		x
5	ACADEMY OF MANAGEMENT REVIEW		x
6	ACCOUNTING AND BUSINESS RESEARCH	x	
7	ACCOUNTING AND FINANCE	x	
8	ACCOUNTING AUDITING & ACCOUNTABILITY JOURNAL	x	
9	ACCOUNTING FORUM	x	
10	ACCOUNTING ORGANIZATIONS AND SOCIETY	x	
11	ACCOUNTING REVIEW	x	
12	AMERICAN ECONOMIC JOURNAL: APPLIED ECONOMICS		x
13	AMERICAN ECONOMIC JOURNAL: ECONOMIC POLICY		x
14	AMERICAN ECONOMIC JOURNAL: MACROECONOMICS		x
15	AMERICAN ECONOMIC JOURNAL: MICROECONOMICS		x
16	ANNUAL REVIEW OF FINANCIAL ECONOMICS	x	
17	AUDITING-A JOURNAL OF PRACTICE & THEORY	x	
18	AUSTRALIAN ACCOUNTING REVIEW	x	
19	BRITISH ACCOUNTING REVIEW	x	
20	BRITISH JOURNAL OF MANAGEMENT		x
21	BROOKINGS PAPERS ON ECONOMIC ACTIVITY		x
22	BUSINESS AND SOCIETY		x
23	CONTEMPORARY ACCOUNTING RESEARCH	x	
24	CORPORATE GOVERNANCE-AN INTERNATIONAL REVIEW	x	
25	CRITICAL PERSPECTIVES ON ACCOUNTING	x	
26	EMERGING MARKETS REVIEW	x	
27	ENTREPRENEURSHIP: THEORY AND PRACTICE		x
28	EUROPEAN ACCOUNTING REVIEW	x	
29	EXPERIMENTAL ECONOMICS		x
30	FINANCE AND STOCHASTICS	x	

31	FINANCE RESEARCH LETTERS	x	
32	FINANCIAL ANALYSTS JOURNAL	x	
33	FINANCIAL MANAGEMENT	x	
34	FORBES	x	
35	IMF ECONOMIC REVIEW	x	
36	INTERNATIONAL JOURNAL OF ACCOUNTING INFORMATION SYSTEMS	x	
37	INTERNATIONAL JOURNAL OF PRODUCTION ECONOMICS		x
38	INTERNATIONAL PUBLIC MANAGEMENT JOURNAL		x
39	INTERNATIONAL REVIEW OF FINANCIAL ANALYSIS	x	
40	JOURNAL OF ACCOUNTING & ECONOMICS	x	
41	JOURNAL OF ACCOUNTING AND PUBLIC POLICY	x	
42	JOURNAL OF ACCOUNTING RESEARCH	x	
43	JOURNAL OF BANKING & FINANCE	x	
44	JOURNAL OF BUSINESS ETHICS		x
45	JOURNAL OF BUSINESS FINANCE & ACCOUNTING	x	
46	JOURNAL OF BUSINESS LOGISTICS		x
47	JOURNAL OF BUSINESS VENTURING		x
48	JOURNAL OF COMMON MARKET STUDIES		x
49	JOURNAL OF CONFLICT RESOLUTION		x
50	JOURNAL OF CONSUMER RESEARCH		x
51	JOURNAL OF CORPORATE FINANCE	x	
52	JOURNAL OF FINANCE	x	
53	JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	x	
54	JOURNAL OF FINANCIAL ECONOMETRICS	x	
55	JOURNAL OF FINANCIAL ECONOMICS	x	
56	JOURNAL OF FINANCIAL INTERMEDIATION	x	
57	JOURNAL OF FINANCIAL SERVICES RESEARCH	x	
58	JOURNAL OF FINANCIAL STABILITY	x	
59	JOURNAL OF FUTURES MARKETS	x	
60	JOURNAL OF HUMAN CAPITAL		x
61	JOURNAL OF INTERACTIVE MARKETING		x
62	JOURNAL OF INTERNATIONAL BUSINESS STUDIES		x
63	JOURNAL OF INTERNATIONAL ECONOMICS		x
64	JOURNAL OF INTERNATIONAL FINANCIAL MANAGEMENT & ACCOUNTING	x	

65	JOURNAL OF INTERNATIONAL FINANCIAL MARKETS INSTITUTIONS & MONEY	x	
66	JOURNAL OF INTERNATIONAL MARKETING		x
67	JOURNAL OF INTERNATIONAL MONEY AND FINANCE	x	
68	JOURNAL OF MANAGEMENT		x
69	JOURNAL OF MANAGEMENT STUDIES		x
70	JOURNAL OF MARKETING		x
71	JOURNAL OF MARKETING RESEARCH		x
72	JOURNAL OF MONETARY ECONOMICS	x	
73	JOURNAL OF MONEY CREDIT AND BANKING	x	
74	JOURNAL OF POLICY ANALYSIS AND MANAGEMENT		x
75	JOURNAL OF PUBLIC ECONOMICS		x
76	JOURNAL OF RISK AND INSURANCE	x	
77	JOURNAL OF RISK AND UNCERTAINTY	x	
78	JOURNAL OF SMALL BUSINESS MANAGEMENT		x
79	JOURNAL OF SUPPLY CHAIN MANAGEMENT		x
80	JOURNAL OF THE ACADEMY OF MARKETING SCIENCE		x
81	JOURNAL OF THE EUROPEAN ECONOMIC ASSOCIATION		x
82	JOURNAL OF WORLD BUSINESS		x
83	LEADERSHIP QUARTERLY		x
84	MANAGEMENT ACCOUNTING RESEARCH	x	
85	MARKETING SCIENCE		x
86	MATHEMATICAL FINANCE	x	
87	MATHEMATICS AND FINANCIAL ECONOMICS	x	
88	QUANTITATIVE MARKETING AND ECONOMICS		x
89	REAL ESTATE ECONOMICS	x	
90	RESEARCH IN INTERNATIONAL BUSINESS AND FINANCE	x	
91	REVIEW OF ACCOUNTING STUDIES	x	
92	REVIEW OF FINANCE	x	
93	REVIEW OF FINANCIAL STUDIES	x	
94	SMALL BUSINESS ECONOMICS		x
95	SOCIO-ECONOMIC REVIEW		x
96	STRATEGIC ENTREPRENEURSHIP JOURNAL		x
97	STRATEGIC MANAGEMENT JOURNAL		x
98	STRATEGIC ORGANIZATION		x
99	STRUCTURAL EQUATION MODELING		x

100	SUPPLY CHAIN MANAGEMENT		x
101	SUSTAINABILITY ACCOUNTING MANAGEMENT AND POLICY JOURNAL	x	
102	THEORETICAL ECONOMICS		x
103	WORLD BANK ECONOMIC REVIEW	x	

Table A.2. Selected RENIR journals

		JCR	SJR
1	ANALES DEL INSTITUTO DE ACTUARIOS ESPAÑOLES		
2	CIRIEC - ESPAÑA. REVISTA DE ECONOMÍA PÚBLICA, SOCIAL Y COOPERATIVA		x
3	CUADERNOS DE ECONOMÍA: SPANISH JOURNAL OF ECONOMICS AND FINANCE		x
4	CUADERNOS DE GESTIÓN		x
5	DE COMPUTIS: REVISTA ESPAÑOLA DE HISTORIA DE LA CONTABILIDAD		
6	DIRECCIÓN Y ORGANIZACIÓN: REVISTA DE DIRECCIÓN, ORGANIZACIÓN Y ADMINISTRACIÓN DE EMPRESAS (REVISTA CEPADE)		x
7	ECONOMÍA AGRARIA Y RECURSOS NATURALES		x
8	EKONOMIAZ: REVISTA VASCA DE ECONOMÍA		x
9	ESIC MARKET		
10	ESTUDIOS DE ECONOMÍA APLICADA		x
11	ESTUDIOS ECONÓMICOS REGIONALES Y SECTORIALES / REGIONAL AND SECTORAL ECONOMIC STUDIES		x
12	EUROPEAN JOURNAL OF MANAGEMENT AND BUSINESS ECONOMICS		x
13	EUROPEAN RESEARCH ON MANAGEMENT AND BUSINESS ECONOMICS	x	x
14	FUZZY ECONOMIC REVIEW		x
15	HACIENDA PÚBLICA ESPAÑOLA	x	x
16	HISTORIA INDUSTRIAL. ECONOMÍA Y EMPRESA	x	
17	INTANGIBLE CAPITAL		x
18	INVESTIGACIONES DE HISTORIA ECONÓMICA: REVISTA DE LA ASOCIACIÓN ESPAÑOLA DE HISTORIA ECONÓMICA		x
19	INVESTIGACIONES EUROPEAS DE DIRECCIÓN Y ECONOMÍA DE LA EMPRESA		
20	INVESTIGACIONES REGIONALES		x
21	RECT@: REVISTA ELECTRÓNICA DE COMUNICACIONES Y TRABAJOS DE ASEPUMA		x
22	REVESCO: REVISTA DE ESTUDIOS COOPERATIVOS		x
23	REVISTA DE CONTABILIDAD	x	x

24	REVISTA DE ECONOMÍA APLICADA	x	x
25	REVISTA DE ECONOMÍA FINANCIERA (THE SPANISH REVIEW OF FINANCIAL ECONOMICS)		x
26	REVISTA DE ECONOMÍA MUNDIAL	x	x
27	REVISTA DE HISTORIA ECONÓMICA - JOURNAL OF IBERIAN AND LATIN AMERICAN ECONOMIC HISTORY	x	x
28	REVISTA DE HISTORIA INDUSTRIAL	x	x
29	REVISTA DE MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA		x
30	REVISTA EMPRESA Y HUMANISMO		
31	REVISTA EUROPEA DE DIRECCIÓN Y ECONOMÍA DE LA EMPRESA		
32	REVISTA GALEGA DE ECONOMÍA: PUBLICACIÓN INTERDISCIPLINAR DA FACULTADE DE CIENCIAS ECONÓMICAS E EMPRESARIAIS		x
33	SERIES : JOURNAL OF THE SPANISH ECONOMIC ASSOCIATION	x	
34	SPANISH JOURNAL OF FINANCE AND ACCOUNTING / REVISTA ESPAÑOLA DE FINANCIACIÓN Y CONTABILIDAD	x	x
35	THE INTERNATIONAL JOURNAL OF DIGITAL ACCOUNTING RESEARCH		
36	UNIVERSIA BUSINESS REVIEW	x	x

Note: x denotes whether the Spanish journal is included in the JCR or SJR rankings.

Table A.3. Diversity and similarity indicators used

	Formulation	**Interpretation**
Panel A. Diversity indicators		
Shannon's Entropy	$ES_s = -\sum_{i=1}^{M_s} p_i Ln(p_i)$	Uncertainty in the identity of the subject of a randomly chosen element of the category.
Gini-Simpson	$GS_s = 1 - \sum_{i=1}^{M_s} p_i^2$	Probability that two randomly chosen elements of the data set (with replacement) represent different subject matter.
Shannon transformed	$HT_s = e^{H_s}$	Effective diversity in Hill's sense: number of equally frequent topics needed to obtain the average frequency observed in the category.
Berger-Parker	$BP_s = Max\left(\{p_i\}_{i=1}^{M_s}\right)$	Relative frequency of the most prevalent topic

Panel B. Similarity indicators

B.1. Occurrence-based indicators

Similarity-1	$GS_{sr} = \frac{Co_{sr}}{\min(M_s, M_r)} x100$	Percentage of co-occurrences to the number of possible co-occurrences between two categories
Similarity-2	$GS_{sr} = \frac{Co_{sr}}{\max(M_s, M_r)} x100$	Percentage of co-occurrences to the number of co-occurrences in the broadest category
Jaccard	$J_{sr} = \frac{Co_{sr}}{M_s + M_r - Co_{sr}}$	Percentage of co-occurrences to the number of different words between the two categories
Sorensen-Dice	$SD_{sr} = \frac{2Co_{sr}}{M_s + M_r}$	Percentage of co-occurrences to the average number of different words in the two categories
Russel-Rao	$RR_{sr} = \frac{Co_{sr}}{M_s + M_r}$	Percentage of co-occurrences to the sum of the number of different words in the two categories

B.2. Abundance-based indicators

Cosine	$C_{sr} = \frac{\sum_{i=1}^{Co_{sr}} p_{i,s}\ p_{i,r}}{\left(\sum_{i=1}^{Co_{sr}} p_{i,s}^2 \sum_{i=1}^{Co_{sr}} p_{i,r}^2\right)^{1/2}}$	Measurement of similarity in angle of category vectors
Horn	$H_{sr} = \frac{1}{Ln2}\sum_{i=1}^{Co_{sr}} p_{i,s} Ln\left(1 + \frac{p_{i,r}}{p_{i,s}}\right) + p_{i,r} Ln\left(1 + \frac{p_{i,s}}{p_{i,r}}\right)$	Extension of Shannon's entropy to measure similarity
Morisita-Horn	$MH_{sr} = \frac{2\sum_{i=1}^{SCo_{sr}} p_{i,s}\ p_{i,r}}{\sum_{i=1}^{Co_{sr}} p_{i,s}^2 + \sum_{i=1}^{Co_{sr}} p_{i,r}^2}$	Probability that items extracted from each category belong to different topics, relative to the extraction of each category separately.

Note: $o_{i,s}$ is the number of times word *i* (its occurrence) is repeated in category *s*; $p_{i,s}$ is the relative frequency (abundance) of word *i* in category *s*; M_s is the number of distinct words in category *s*; N_s is the total number of words in category *s* ($N_s = \sum_{i=1}^{M_s} o_{i,s}$). Co_{sr} is the number of co-occurring distinct words in the categories *s* and *r*.