Original Research Article

The impact of publishing houses and subjects on the citation quantity and frequency of books

Qianhe Gao

International Business School, Shandong Technology and Business University, Shijiazhuang City, Hebei Province, 050000, China

Abstract: This study probes the impact of publishers and subjects on book citation metrics. Despite prior research on citation influence, the effects of publishing houses and subjects on citation counts were underexposed. The research tasks involved analyzing data from CNKI's "Most Academic Influential Publishing House (2014 - 2023)" on 587 publishers, 17 disciplines, and numerous books. Use the R language for overall data analysis and apply the ANOVA model to test the correlation of these data, ultimately drawing conclusions through the tests. For cited book numbers, ANOVA shows both publishing house type (non-university presses higher) and subject (natural sciences and tech more) matter. For citation frequency per book, subject is key (philosophy and social sciences higher), not publishing house type. Discussion reveals these findings align with some literature, enriching academic evaluation theory. However, data reliance on CNKI limits generalization, and insufficient factor interaction analysis may miss nuances. Future research should expand factors and globally validate findings.

Keywords: Citation quantity; Citation frequ**ency;** Publishing houses; Subjects; Academic evaluation

1. Introduction

In the realm of academic research, citation rates have been a subject of extensive study. Previous investigations have delved into various factors influencing citation rates, such as the work of Onodera and Yoshikane (2014) on the complexity of academic impact measurement, Rostami et al. (2014) regarding the role of article titles, Liu et al. (2015) exploring the correlation between peer-reviewed scores and impact factors, and Hutchins et al. (2016) introducing the Relative Citation Ratio. Additionally, there is a substantial body of literature exploring the correlation between social media engagement and academic citations, as well as new indicators for academic evaluation and their relationship with traditional citation indicators.

However, despite the wealth of research in related areas, few studies have directly addressed the impact of different publishers and subjects on citation counts. In the current study, we aim to fill this gap by comprehensively investigating the effects of two crucial factors, namely publishing houses and subjects, on the number of cited books and the citation frequency per book.

2. Literature review

2.1. Factors affecting citation rates

Regarding the factors affecting citation rates, many predecessors have done a lot of studies. The research in this field explores the various elements contributing to the citation rates of scholarly articles. For instance, Onodera and Yoshikane (2014) conducted a comprehensive analysis of factors influencing citation rates, highlighting the complexity of academic impact measurement. Rostami et al. (2014) focused on the role of article titles, suggesting that specific characteristics of titles can significantly impact citation rates. Liu et al. (2015) ex-

plored the correlation between peer-reviewed scores and impact factors across different citation time windows for ophthalmic journals, providing insights into the temporal dynamics of citation impact. Hutchins et al. (2016) introduced the Relative Citation Ratio (RCR), a new metric that leverages citation rates to measure article-level influence, offering a fresh perspective on impact assessment.

2.2. Social media engagement and citation correlation

Though there are few studies directly addressing the impact of different publishers and subjects on citation counts, there is a substantial amount of literature exploring the correlation between social media engagement and academic citations. For example, Madhugiri and Venkatesan (2024) found that reader engagement with neurosurgery journal websites, as measured by views and downloads, correlates with the number of citations received by articles. Qian et al. (2017) performed a citation regression analysis on computer science publications, revealing the interplay between ranking categories, fields, and citation patterns. Fang (2018) utilized derivative analysis to study the yearly citation trends of "sleeping beauties" in science, shedding light on the long-term impact of publications.

2.3. Academic evaluation and impact metrics

Although there are few studied direct research on indicators that affect citation frequency, a large amount of literature has explored new indicators for academic evaluation and their relationship with traditional citation indicators. For example, Liu and Huang (2022) explored the relationship between altmetric counts and citations across different academic fields, using co-occurrence analysis to uncover the potential of altmetrics as academic evaluation indicators. Wu et al. (2024) investigated the citation advantage of positive words in scientific papers, discussing the predictability, temporal evolution, and universality of this phenomenon across journals of varying quality. McGillivray and Astell (2019) analyzed the relationship between usage and citations in an open access mega-journal, providing evidence on the impact of open access on citation rates. Yang et al. (2019) proposed a new approach to constructing journal co-citation matrices based on the number of co-cited articles, offering a novel method for journal comparison and evaluation.

3. The current study

In the present study, our primary objective was to comprehensively investigate the effects of two crucial factors, namely publishing houses and subjects, on the number of cited books and the citation frequency per book. For the number of cited books, we aimed to figure out the comparison between university presses, which are usually more academically and specially oriented, and non-university publishers in respect of the citation quantity of their books. Additionally, we aimed to analyze the differences between subject areas like philosophy and social sciences, which often involve theoretical and interpretative works, and natural sciences and technology, which are more centered around empirical research and technological advancements, in relation to the number of times their respective books are cited.

Similarly, regarding the citation frequency per book, we intended to explore how the characteristics of different publishing houses and subject areas influence the frequency with which individual books are cited. University presses may have certain editorial and academic standards that could potentially affect the citation frequency of their publications differently from non-university publishers. Likewise, the nature of research and knowledge dissemination in philosophy and social sciences might lead to a different citation frequency pattern compared to that in natural sciences and technology.

The research questions that will be addressed in this study are as follows:

Research Question 1: How do different publishing houses and subjects impact the number of cited books?

Research Question 2: How do different publishing houses and subjects influence the citation frequency per book?

4. Research methods

4.1. Materials

The data was collected from the selection of the "Most Academic Influential Publishing House (2014-2023)" released by the China National Knowledge Infrastructure (CNKI) Evaluation Center. The CNKI Evaluation Center collected and analyzed data from various sources. The bibliographic data of books was sourced from the National Library of China. This comprehensive collection provided detailed information about the books published in China, serving as a crucial foundation for the research. The citation statistical sources were journal papers, conference papers, and doctoral and master's degree papers included in the CNKI Knowledge Resource Database. And experts conducted two rounds of reviews on the lists of publishers. These academic documents were used to determine the citation frequencies of the books, thus enabling an assessment of their academic influence. The "Citation Analysis Database of Chinese Books" was developed by the CNKI Evaluation Center. This database was based on the Citation Data of the Chinese Academic Evaluation Support Platform and was used to regularly count and publish the publication situation and citation frequency of books in mainland China from 1949 to the present. During the period of 2014 - 2023, it covered data on 587 book publishing units, 17 disciplinary fields, approximately 1, 570, 000 published books, a total citation frequency of approximately 3, 850, 000 times, and about 340, 000 books that were cited at least once.

4.2. Procedure

Obtain relevant data on the «Most Influential Academic Publishers (2014-2023)» in 17 disciplinary fields published by CNKI Evaluation Center. Firstly, classify and code the disciplines, including economics, political science, sociology, management, literature, history, law, and education, into the category of philosophy and social sciences, and uniformly code them as 1 (n=118). Classify the nine disciplinary fields of industrial technology, medicine, health, mathematical and chemical sciences, agricultural sciences, transportation, biological sciences, astronomy, earth sciences, environmental sciences, and aerospace as natural science and technology, and code them as 2 (n=75). Next, classify and code the publishing house into two categories based on their nature: university presses and non-university presses, and code them as 1 (n=50) and 2 (n=143) respectively. Input the classified and encoded data into a spreadsheet for subsequent data analysis and processing.

4.3. Data analysis

After processing the data, treat publishing house and subject were considered as independent variables, while the number of cited books and average citation frequency per book were considered as dependent variables. Use the R language for overall data analysis and apply the ANOVA model to test the correlation of these data, ultimately drawing conclusions through the tests.

5. Statistical results and analysis

5.1. Experiment 1: Investigating the influence of publishing house and subject on the number of cited books

This study first explores the impact of subjects and publishing houses on the number of cited books separately, then examines their joint influence (without interaction), and finally studies the interaction effect, calculating the AICc-values of each model. The results showed that the model of subjects and publishing houses without

interaction effects had the lowest AICc-value, and the best balance between goodness of fit and complexity was achieved, which could more effectively explain quantitative changes. Therefore, this study chose this model to explore its impact relationship in depth.

Table 1. Effects of publishing houses and subjects on number of cited books.

Factor	Df	Sum of square	Mean Square	F value	Pr(>F)
publishing	1	4905751	4905751	4.760	0.0303*
subject	1	3334724	3334724	3.236	0.0736.

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 17

To explore the effects of publishing houses (university presses and non-university presses) and subjects (philosophy and social sciences, and natural science and technology) on number of cited books, and ANOVA was performed, and the results are listed in **Table 1**. **Table 1** indicate that publishing house (F=4.760, p<0.05) , and subject (F=3.236, p<0.1) had a significant effect on number of cited books.

Table 2. The mean number of books cited by each type of publishing house.

Publishing	Mean	Std. Deviation	N	*P*-value
Group 1	577.1800	765.0838	50	0.0313*
Group 2	941.0769	1095.6179	143	

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

To further investigate whether the observed differences between the two groups of publishing houses were significant, the Tukey HSD (Honestly Significant Difference) method was used as part of post hoc multiple comparison tests, which was analyzed through PostHocTest. In **Table 2**, the results signify that, at a 95% confidence level, there exists a statistically significant difference in the pairwise comparisons of publishing house types. Given that the p - value of 0.0313 is less than the conventionally adopted significance level of 0.05, the null hypothesis can be rejected. It can thus be inferred that there are probable genuine differences between the specific pairs of publishing house types as presented. The mean number of cited books in group 2 (non-university presses, *mean*= 941.0769, *sd*=1095.6179) is significantly higher than that in group 1 (university presses, *mean*= 577.1800, *sd*=65.0838), indicating that non-university presses might have a greater influence in terms of the quantity of cited books.

Table 3. The mean number of books cited by each type of subject.

Subject	Mean	Std. Deviation	N	*P*-value
Group 1	737.661	853.3027	118	0.0649 .
Group 2	1018.520	1247.8848	75	

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

To further investigate whether the observed differences between the two groups of subjects were significant, the Tukey HSD (Honestly Significant Difference) method was used as part of post hoc multiple comparison tests, which was analyzed through PostHocTest. In **Table 3**, the results signify that, at a 90% confidence level, there exists a statistically significant difference in the pairwise comparisons of subject types. Given that the p - value of 0.0649 is less than the conventionally adopted significance level of 0.1, the null hypothesis can be rejected. It can thus be inferred that there are probable genuine differences between the specific pairs of subject types as presented. The mean quantity of subject in group 2 (natural science and technology, *mean* = 1018.520, *sd*=1247.8848) is greater than that in group 1 (philosophy and social sciences, *mean* = 737.661, *sd*=853.3027). This disparity might suggest that group 2 has a more substantial number of cited books relative to group 1.

5.2. Experiment 2: Investigating the influence of publishing house and subject on citation frequency per book

This study first explores the impact of subjects and publishing houses on citation frequency per books separately, then examines their joint influence (without interaction), and finally studies the interaction effect, calculating the AICc-values of each model. The results showed that the model of subjects and publishing houses without interaction effects had the lowest AICc value, and the best balance between goodness of fit and complexity was achieved, which could more effectively explain quant imitative changes. Therefore, this study chose this model to explore its impact relationship in depth.

Table 4. Effects of publishing houses and subjects on citation frequency per book.

Factor	Df	Sum of square	Mean Square	F value	Pr(>F)
publishing	1	2192	2192	0.375	0.5408
subject	1	25604	25604	4.384	0.0376*

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

To explore the effects of publishing houses (university presses and non-university presses) and subjects (philosophy and social sciences, and natural science and technology) on citation frequency per book, and ANOVA was performed, and the results are listed in **Table 4**. **Table 4** indicate that subject (F=4.384, p<0.05) had a significant effect on number of cited books. However, the factor of publishing house, with an F value of 0.375 and a p - value greater than 0.1, did not exhibit a significant effect on the number of cited books.

Table 5. The mean number of citation frequency per book by each subject.

Subject	Mean	Std. Deviation	N	*P*-value
Group 1	33.54263	97.418726	118	0.0394*
Group 2	10.16680	5.216121	75	

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

To further investigate whether the observed differences between the two groups of subject were significant, the Tukey HSD (Honestly Significant Difference) method was used as part of post hoc multiple comparison tests, which was analyzed through PostHocTest. In **Table 3**, the results signify that, at a 95% confidence level, there exists a statistically significant difference in the pairwise comparisons of subject types. Given that the p - value of 0.0394 is less than the conventionally adopted significance level of 0.05, the null hypothesis can be rejected. It can thus be inferred that there are probable genuine differences between the specific pairs of subject types as presented. The mean quantity of subject in group 1 (philosophy and social sciences, *mean* = 33.54263, *sd*=97.418726) is higher than that in group 2 (natural science and technology, *mean* = 10.16680, *sd*=5.216121). This difference could imply that group 1 has a larger citation frequency per book compared to group 2

6. Discussion

For the first experiment, the study reveals that both the type of publishing house and the subject area significantly influence the number of cited books. Notably, books published by non-university presses have a significantly higher number of citations compared to those published by university presses. Additionally, books in the natural sciences and technology fields receive more citations than those in philosophy and social sciences. These findings align with the existing literature on the impact of publishers and subject areas on citation rates. For instance, the study by Onodera and Yoshikane (2014) highlighted the complexity of academic impact measurement, and our research further clarifies the specific roles of publishers and subject areas in this context. The results of our study offer a fresh perspective on how publishers and subject areas can affect the citation count

of academic books. It suggests that the academic orientation of publishers and the nature of research within subject areas are key factors influencing citation numbers. While our study is based on data from China, the influence of publisher type and subject area on citation counts may have universal implications, as these are characteristics shared across the global academic community.

For the second experiment, the results indicate that subject areas significantly affect the citation frequency per book, while the type of publishing house does not have a significant impact. Books in the philosophy and social sciences field have a higher citation frequency per book than those in the natural sciences and technology field. This finding is consistent with the existing literature on the impact of subject areas on citation rates. For example, the study by Wu et al. (2024) investigated the citation advantage of positive words in scientific papers, and our research further illuminates the role of subject areas in this regard. Our study provides new insights into how subject areas can influence the citation frequency of academic books. It suggests that the nature of research and knowledge dissemination within subject areas are critical factors affecting citation frequency. The influence of subject areas on citation frequency may be universally applicable, as these are characteristics inherent to academic disciplines worldwide.

7. Conclusion

This study centered on the impact of publishers and subject areas on the number of cited books and the citation frequency per book, filling a research gap in the academic field. Regarding the number of cited books, both publisher type and subject area have significant effects. For citation frequency per book, subject area is a significant determinant, with philosophy and social sciences showing higher frequencies than natural sciences and technology, while publisher type has no significant impact. These findings corroborate and expand on prior research, shedding new light on the role of these factors in academic citation.

Theoretically, it deepens the understanding of academic influence components, emphasizing the importance of publisher orientation and subject nature in citation metrics, thus enriching the academic evaluation framework. Practically, it guides publishers' strategies and scholars' choices, optimizing academic resource allocation and dissemination.

This study has two main limitations. Firstly, it depends on CNKI data, confining the sample to Chinese academic books. This may bias results towards the Chinese context, so using international data for validation is needed to enhance generalization. Secondly, the data analysis didn't deeply examine interactions among some factors. This might overlook how complex variable relations subtly affect outcomes.

Future research could investigate other factors that might affect citation frequency. These could include the quality of book content, the reputation of authors, and the timing of publication. Additionally, further exploration could be done on additional factors that influence citation counts, and studies could also be conducted on a global scale to test the universality of these findings.

About the author

Gao Qianhe (2004, 12-), female, is from Shijiazhuang City, Hebei Province, China. Her main research areas are in the fields of financial market risk management and business administration.

References

[1] Fang, H. (2018). Analysing the variation tendencies of the numbers of yearly citations for sleeping beauties in science by using derivative analysis. Scientometrics, 115(1), 1051–1070. https://doi.org/10.1007/s11192-018-2687-9

[2] Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative Citation Ratio (RCR):

A New Metric That Uses Citation Rates to Measure Influence at the Article Level. PLoS Biology, 14(9), e1002541. https://doi.org/10.1371/journal.pbio.1002541

[3] Liu, C., & Huang, M.-H. (2022). Exploring the relationships between altmetric counts and citations of papers in different academic fields based on co-occurrence analysis. Scientometrics, 127, 4939–4958. https://doi.org/10.1007/s11192-022-04456-w

[4] Liu, X.-L., Gai, S.-S., Zhang, S.-L., & Wang, P. (2015). An Analysis of Peer-Reviewed Scores and Impact Factors with Different Citation Time Windows: A Case Study of 28 Ophthalmologic Journals. PLoS ONE, 10(8), e0135583. https://doi.org/10.1371/journal.pone.0135583

[5] Madhugiri, V. S., & Venkatesan, S. (2024). Does reader engagement with neurosurgery journal websites correlate with the number of citations received by articles? Neurology India, 72(2), 352-357. https://doi.org/10.4103/ni.ni_38_22

[6] McGillivray, B., & Astell, M. (2019). The relationship between usage and citations in an open access mega-journal. Scientometrics, 121(2), 817–838. https://doi.org/10.1007/s11192-019-03228-3

[7] Onodera, N., & Yoshikane, F. (2014). Factors affecting citation rates of research articles. Journal of the Association for Information Science and Technology, 66(4), 739–764. https://doi.org/10.1002/asi.23209

[8] Qian, Y., Rong, W., Jiang, N., Tang, J., & Xiong, Z. (2017). Citation regression analysis of computer science publications in different ranking categories and subfields. Scientometrics, 110(2), 1351–1374. https://doi.org/10.1007/s11192-016-2235-4

[9] Rostami, F., Mohammadpoorasl, A., & Hajizadeh, M. (2014). The effect of characteristics of title on citation rates of articles. Scientometrics, 98(2), 2007–2010. https://doi.org/10.1007/s11192-013-1118-1

[10] Wu, D., Wu, H., & Li, J. (2024). Citation advantage of positive words: predictability, temporal evolution, and universality in varied quality journals. Scientometrics, 129, 4275–4293. https://doi.org/10.1007/s11192-024-05074-4

[11] Yang, L., Han, L., & Liu, N. (2019). A new approach to journal co-citation matrix construction based on the number of co-cited articles in journals. Scientometrics, 120(2), 507–517. https://doi.org/10.1007/s11192-019-03141-9