

## **Analysis of Distribution Characteristics and Influencing Factors of Hot Papers: With Commentary on Time Windows and Inter-disciplinary Differences (Postprint)**

**Authors:** Song Chao, Chen Yue, Wang Ling, Zuo Jia, Liu Zeyuan

**Date:** 2023-07-26T00:00:00+00:00

### **Abstract**

**Purpose/Significance:** In recent years, hot papers have gradually attracted academic attention. While limited research has begun to explore the intrinsic characteristics of hot papers, investigations into influencing factors and underlying patterns remain insufficient. **Methods/Process:** This study employs the TF-IDF algorithm and negative binomial regression model to investigate the distribution characteristics, influencing factors, time window variations, and disciplinary differences of hot papers. **Results/Conclusions:** The findings indicate that hot papers are predominantly distributed in developed countries, renowned research institutions, interdisciplinary fields, and authoritative journals; and are influenced by factors such as concise titles, international collaboration, research-oriented outputs, open access, and high-impact-factor journals. Hot papers exhibit a time window effect, which even alters the specific pathways through which titles, abstracts, open access, and other factors influence them. Disciplinary differences significantly affect hot papers, with variations existing in titles, abstracts, research collaboration, document types, open access, and journal impact factors.

### **Full Text**

### **Preamble**

**Vol. 63 No. 16 August 2019**

**Analysis of Distribution Characteristics and Influencing Factors of Hot Papers—A Comment on Differences Under Time Window and Across Disciplines**

Song Chao, Chen Yue, Wang Ling, Zuo Jia, Liu Zeyuan

Institute of Science of Science and S&T Management & WISE Lab, Dalian University of Technology, Dalian 116024

## Abstract

**[Purpose/Significance]** In recent years, hot papers have gradually attracted academic attention. While existing research has begun to explore the characteristics of hot papers themselves, studies on influencing factors and underlying patterns remain insufficient. **[Method/Process]** This study employs the TF-IDF algorithm and negative binomial regression model to investigate the distribution characteristics, influencing factors, time window differences, and disciplinary variations of hot papers. **[Result/Conclusion]** The findings indicate that hot papers are predominantly distributed in developed countries, renowned research institutions, interdisciplinary fields, and authoritative journals. They are influenced by factors such as concise titles, international collaboration, research-type output, open access, and high-impact journals. Hot papers exhibit time window effects that even alter how factors like titles, abstracts, and open access influence citation trajectories. Disciplinary differences also impact hot papers, showing variations in titles, abstracts, research collaboration, literature types, open access, and journal impact factors.

**Classification Number:** G301

**Keywords:** hot papers, distribution characteristics, influencing factors, time window, disciplinary differences, negative binomial regression model

## 2. Literature Review and Commentary

### 2.1 Literature Review

**2.1.1 Hot Paper-Centered Research** Research on hot papers has examined not only their network impact in terms of citations but also differences in literature types, time windows, disciplines, journals, and countries. Hot papers are typically characterized by “rapid citation accumulation” [13]. Early studies in network analysis treated each paper as a node, where simultaneously created nodes had equal opportunities to be connected. Empirical research on scientific paper attractiveness demonstrated that hot papers receive more links or citations than contemporaneous papers [14-15]. Subsequent scholars used the “node popularity” metric to characterize node influence within a field, finding that regional collaboration primarily involves authors, countries, and institutions. Collaboration can enrich research perspectives [32], and empirical studies confirm that cooperation improves research performance and enhances the quality and impact of scientific output [33-34]. However, Professor Glänzel holds a different view, arguing that “a considerable portion of internationally co-authored papers have citation performance below the sample average” [35] and that “the myth that collaboration always guarantees success” [36] should be dispelled. Other scholars note uncertainty in the relationship between collaboration scale and citation frequency [37], or even no significant association

[38].

Literature type also differs among hot papers. If review articles are more likely to become hot papers than research articles, it may indicate a scarcity of high-quality original research [17]. The effect of time windows cannot be ignored and varies significantly by discipline. A publication's true impact can only be determined after a considerable period—for instance, biomedicine and multidisciplinary sciences typically show a 3-year citation window, while humanities and mathematics require 7 years [18]. Some studies even find that due to the highly specialized experimental methods and complex theories in physical chemistry, the true impact of publications in this field is difficult to measure in the short term. Rapid recognition in this field likely benefits from a large number of active researchers [19]. In contrast, certain disciplines (e.g., biochemistry, molecular biology, immunology, cell biology), journals (e.g., PRL, PNAS), and countries (e.g., USA) may produce more “hot papers” [20].

**2.1.2 Research on Hot Paper Influencing Factors** Research specifically on hot paper influencing factors remains scarce, but literature on citation impact factors provides valuable references. Most studies examine citation frequency influences from both paper-internal and external information perspectives—here termed internal and external factors. Internal factors refer to information expressed by the paper itself, such as titles, abstracts, collaboration (author, country, institution), literature type, funding, references, article length, and publication year. External factors include journal impact factor, open access availability, article-level usage metrics [21], and number of disciplines. While research on these influencing factors is common, conclusions vary significantly.

Regarding internal factors, title impact on citations shows three main conclusions (excluding editorial length restrictions) [22]: positive correlation [23-24] (longer titles receive more citations), no relationship [25], and negative correlation [26-28] (shorter titles receive more citations). Research on abstract impact is less common. Some scholars use the Flesch Reading Ease index [29-30] to analyze abstract readability, finding that papers whose abstracts contain frequently used words receive higher evaluations—shorter abstracts using more common vocabulary may be easier to read and thus gain more citations [31]. Research collaboration, encompassing authors, countries, and institutions, can enrich research perspectives [32], and empirical studies confirm that collaboration improves research performance and output quality [33-34]. However, Professor Glänzel argues that “a considerable portion of internationally co-authored papers have citation performance below the sample average” [35] and that “the myth that collaboration always guarantees success” [36] should be dispelled. Other scholars note uncertainty in the relationship between collaboration scale and citation frequency [37], or even no significant association [38]. References constitute a paper's knowledge base—more references often lead to more citations [39]. Article length (page count) [43] is considered an important factor affecting citations, though some argue there is no relationship [44]. The exist-

tence of citation half-life (approximately 5.6 years) [45] confirms that different citation time windows [46] affect citation frequency.

External factors show varying impacts. While high-impact journals are generally believed to increase citation likelihood [47], some find that journal impact factor and its annual changes often have no direct effect on citations [48], with others arguing that using impact factor to judge articles is “putting the cart before the horse” [49]. Open access publishing [50] and disciplinary differences [51] also affect citations. Recent research has incorporated “article-level usage” metrics from indexing databases, suggesting usage counts could serve as an early indicator of academic quality [52], with high early usage predicting later high citations [53-54]. However, studies on the relationship between usage and citations yield mixed results: positive correlation [55-56], weak correlation [27,57], and even negative correlation [58]. While older highly cited papers have higher usage, readers prefer more recent literature [59]. Disciplinary differences in readership and literature types [55] also influence citations.

## 2.2 Commentary

The review reveals that few studies have quantitatively analyzed hot paper distribution characteristics and influencing factors. Moreover, most research on citation impact factors uses fixed time windows or single disciplines, making cross-disciplinary and time-series comparisons difficult due to publication year and disciplinary differences, yielding inconsistent conclusions. This study posits that time window and disciplinary differences may moderate how other factors influence citations. Therefore, this research focuses on hot paper distribution characteristics and influencing factors, with comparative analysis across time windows and disciplines.

## 3. Research Questions and Design

### 3.1 Research Questions

Hot papers rapidly emerge from vast scientific literature, likely following certain scientometric patterns. This study investigates: (1) distribution characteristics by country, institution, discipline, and journal; (2) internal and external influencing factors; and (3) differences across time windows and discipline types (see research logic framework in Figure 1 [Figure 1: see original paper]).

### 3.2 Research Design

**3.2.1 Data Sources** This study uses hot paper data from ESI via the Web of Science (WoS) Core Collection, distinct from Scopus’s “TOP25 Hottest Articles” [60] and Altmeter.com’s “Top100” [61]. Since WoS only records hot papers from the past two years, publication years were limited to 2016-2018, with document types restricted to “Article” and “Review” across all disciplines. Data

retrieval on September 2, 2018, yielded 2,959 hot papers, with 2,798 retained after excluding incomplete records.

**3.2.2 Variable Selection and Definition** Based on internal and external factors, variables were selected (see Table 1 ). The dependent variable is citation frequency. Independent variables include: title word count, title distinctiveness (TF-IDF), abstract word count, abstract distinctiveness, author collaboration scale, country collaboration scale, institution collaboration scale, literature type, funding, reference count, article length, publication year, open access status, usage count, number of disciplines, discipline type, and journal impact factor.

Variable processing details: Title and abstract word counts were obtained via natural language processing (removing stop words). TF-IDF algorithm extracted distinctive terms from titles and abstracts. Institution collaboration counted only primary affiliations (internal collaborations within the same primary institution counted as one). Following Qian Xuesen's science and technology system theory, all sciences can be divided into basic science, technical science, and engineering technology. Using Professor Liu Zeyuan's classification standard [62], 2,637 natural science hot papers were categorized accordingly; 161 social science hot papers were not subdivided.

**3.2.3 Model Construction** To analyze hot paper influencing factors, the following model was constructed:

Citation frequency is a non-negative integer count variable with discrete distribution and non-normal distribution, typically fitted with Poisson regression requiring equal mean and variance. However, since the sample variance exceeds the mean (overdispersion), Poisson regression is inappropriate [63]. Negative binomial regression was selected after passing tests, while zero-inflated negative binomial regression failed tests. Stata 14.0 was used for negative binomial regression analysis.

## 4. Empirical Results Analysis

### 4.1 Descriptive Statistics and Correlation Analysis

Descriptive statistics (Table 2 ) show large standard deviations, non-normal discrete distributions, and zero values for many variables, indicating no general patterns. Large ranges suggest influencing factors may lack universality.

Due to non-linear relationships, Spearman correlation coefficients were used (Table 3 ). Most variables show significant correlations within reasonable ranges, with VIF values below 10, indicating no severe multicollinearity.

Given large sample size, conventional scatter plots are too dense. Binned scatter plots [64] were created by dividing X-axis variables into equal groups and plotting group means to clarify relationships (Figure 2 [Figure 2: see original

paper]), confirming that citation frequency and influencing factors may lack general patterns.

## 4.2 Distribution Characteristics of Hot Papers

Figure 3 [Figure 3: see original paper] shows distribution characteristics (top 10) by country, institution, discipline, and journal. Countries are concentrated in developed nations, with the USA contributing the most hot papers (nearly double China's count, though China ranks second). Universities are primary producers, with Stanford University leading and the Chinese Academy of Sciences ranking second. Disciplines concentrate in multidisciplinary sciences, chemistry (multidisciplinary), and medicine. Top journals are authoritative, confirming that high-impact journals produce more hot papers.

## 4.3 Analysis of Hot Paper Influencing Factors

**4.3.1 Internal Factors** As Table 4 shows, titles using distinctive vocabulary are more important than abstracts for hot papers. Models (1)-(4) indicate that concise, distinctive titles quickly capture attention and gain citations. International collaboration significantly and positively impacts hot papers, while author and institutional collaboration show no significant effect. Research articles demonstrate higher academic value than reviews, indicating innovation. Funding negatively affects hot papers in short-term citation windows. Reference count shows no significant impact. Article length effects are uncertain and non-significant. Time window effects are evident: 2016 and 2017 hot papers receive more citations than 2018 papers, with 2016 > 2017.

**4.3.2 External Factors** Open access significantly drives hot paper citations, with usage counts also positively impacting citations (Models 1-4). Journals covering too many disciplines negatively impact citations, likely due to lack of specialization. Technical science fields receive the most citations, followed by basic science, engineering technology, and social sciences. High-impact journals accelerate citations through “follower effects.”

## 4.4 Time Window Differences

Cross-year comparisons (Table 5 ) reveal time-varying factor impacts: earlier publications benefit more from concise titles and open access, while recent publications gain more from distinctive abstracts. Collaboration effects vary by time, but funding, references, and article length show no clear patterns. Open access contributes more to older hot papers, while usage metrics matter more for recent ones. Multi-disciplinary journals show unstable effects. Papers published in the previous year (year 2) are most likely to gain citations. Journal impact factor consistently shows positive but time-invariant effects.

#### 4.5 Disciplinary Differences

Cross-disciplinary comparisons (Table 6 ) show: basic and technical sciences benefit more from titles than abstracts, while social sciences benefit from abstracts; basic science collaboration has no significant effect, technical science shows less author/institutional but more country collaboration, engineering shows more institutional but less author collaboration, and social sciences show less international collaboration; research articles are more cited in basic and technical sciences but not in engineering/social sciences; funding effects are discipline-invariant; references have no significant effect across disciplines; article length only matters in technical sciences; time window effects are universal; open access and impact factor positively affect basic, technical, and social sciences but not engineering.

### 5. Conclusions and Discussion

- (1) Hot papers feature concise titles, international collaboration, research-type output, open access, and high-impact journals. Developed countries contribute most hot papers, with China as a major player. Hot papers focus on interdisciplinary fields like multidisciplinary sciences.
- (2) Hot papers exhibit time window effects that moderate how factors like titles, abstracts, and open access influence citations. Earlier hot papers benefit more from titles and open access, while recent ones gain from abstracts.
- (3) Disciplinary differences significantly impact hot papers. Technical science hot papers are most cited, followed by basic science, engineering technology, and social sciences. Disciplines differ in title/abstract effects, collaboration patterns, literature types, open access, and journal impact.
- (4) This study analyzes hot paper characteristics without comparing them to other literature types. Whether hot papers include “flash-in-the-pan” articles remains unknown and requires future research comparing hot papers with other literature types.

### References

- [1] GROSS PL, GROSSE M. College libraries and chemical education[J]. *Science*, 1927, 66(1713): 385-389.
- [2] VIRGO JA. A statistical procedure for evaluating the importance of scientific papers[J]. *Library quarterly*, 1977, 47(4): 415-430.
- [3] MOED HF. The impact-factors debate: the ISI's uses and limits[J]. *Nature*, 2002, 415(6873): 731-732.
- [4] BORNMANN L, ANEGÓN FM, LEYDESDORFF L. Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega Hypothesis[J]. *Plos one*, 2010, 5(10): e13327.
- [5] BORNMANN L, DANIEL H. What do citation counts measure? A review

- of studies on citing behavior[J]. *Journal of documentation*, 2008, 64(1): 45-80.
- [6] MARX W, BORNMANN L. Change of perspective: bibliometrics from the point of view of cited references: a literature overview on approaches to the evaluation of cited references in bibliometrics[J]. *Scientometrics*, 2016, 109(2): 1397-1415.
- [7] PERITZ BC. On the objectives of citation analysis: problems of theory and method[J]. *Journal of the Association for Information Science & Technology*, 1992, 43(6): 448-451.
- [8] HIRSCH JE. An index to quantify an individual's scientific research output[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(46): 16569-16572.
- [9] RAAN AFJ. Sleeping beauties in science[J]. *Scientometrics*, 2004, 59(3): 467-472.
- [10] DAVIS JB. Problems in using the social science citation index to rank economics journals[J]. *American economist*, 1998, 42(2): 59-64.
- [11] CLARIVATE. Analyze top research output and research fronts[EB/OL]. [2018-08-01]. <https://clarivate.com/products/essential-science-indicators/>.
- [12] SONG M, DU S, LIU D, et al. Analysis of hot papers funded by the National Natural Science Foundation of China—Based on ESI data from 2013-2015[J]. *China Science Foundation*, 2017, 31(5): 489-493.
- [13] THOR A, BORNMANN L, MARX W, et al. Identifying single influential publications in a research field: new analysis opportunities of the CR Explorer[J]. *Scientometrics*, 2018, 116(1): 591-608.
- [14] EOM YH, FORTUNATO S. Characterizing and modeling citation dynamics[J]. *Plos one*, 2011, 6(9): e24926.
- [15] WANG D, BARABÁSI AL. Quantifying long-term scientific impact[J]. *Science*, 2013, 342(6154): 127-132.
- [16] ZHENG X, OUYANG Z, ZHANG P, et al. Modeling the citation network by network cosmology[J]. *Plos one*, 2015, 10(3): e0120687.
- [17] WANG X, LÜ C, SUN X. Analysis of hot paper output in international pharmacology and toxicology research based on ESI[J]. *Journal of Pharmaceutical Practice*, 2016, 34(5): 437-440.
- [18] WANG J. Citation time window choice for research impact evaluation[J]. *Scientometrics*, 2013, 94(3): 851-872.
- [19] SCHOLLES GD, KAMAT PV. Hot papers in physical chemistry[J]. *The journal of physical chemistry letters*, 2016, 7(2): 339-342.
- [20] BORNMANN L, YEA Y, YEF Y. Identifying “hot papers” and papers with “delayed recognition” in large-scale datasets by using dynamically normalized citation impact scores[J]. *Scientometrics*, 2018, 116(2): 655-674.
- [21] SUN X. What is the new “article-level usage metric” feature in SCI?[EB/OL]. [2018-08-01]. <http://blog.sciencenet.cn/blog-41174-926981.html>.
- [22] LETCHFORD A, MOAT HS, PREIS T. The advantage of short paper titles[J]. *Royal society open science*, 2015, 2(8): 150266.
- [23] HABIBZADEH F, YADOLLAHI M. Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals[J]. *Croatian medical journal*, 2010, 51(2): 165-170.

- [24] JACQUES TS, SEBIRE NJ. The impact of article titles on citation hits: an analysis of general and specialist medical journals[J]. *Journal of the Royal Society of Medicine short reports*, 2010, 1(1): 1-5.
- [25] NAIR LB, GIBBERT M. What makes a 'good' title and (how) does it matter for citations?[J]. *Scientometrics*, 2016, 107(3): 1331-1359.
- [26] PAIVA CE, PAIVA BSR. Articles with short titles describing the results are cited more often[J]. *Clinics*, 2012, 67(5): 509-513.
- [27] SUBOTICS, MUKHERJEE B. Short and amusing: the relationship between title characteristics, downloads, and citations in psychology articles[J]. *Journal of information science*, 2014, 40(1): 115-124.
- [28] GNEWUCH M, WOHLRABE K. Title characteristics and citations in economics[J]. *Scientometrics*, 2016, 110(3): 1-6.
- [29] DIDEGAH F, THELWALL M. Which factors help authors produce the highest impact research? A review and general model of article title attributes in management science[J]. *Journal of informetrics*, 2014, 8(2): 195.
- [30] XU Q, KANG X, ZHANG C. Study on several influencing factors of paper citation frequency from a multi-journal comparative perspective[J]. *Journal of Intelligence*, 2018, 37(2): 147-153.
- [31] LETCHFORD A, PREIS T, MOAT HS. The advantage of simple paper abstracts[J]. *Journal of informetrics*, 2015, 10(1): 1-8.
- [32] QIU J, ZENG Q. Can international cooperation improve research impact? A case study of computer science[J]. *Information Studies: Theory & Application*, 2013, 36(10): 1-5.
- [33] BEAVER DD. Reflections on scientific collaboration (and its study): past, present, and future[J]. *Scientometrics*, 2001, 52(3): 365-377.
- [34] FIGG WD, DUNN L, LIEWEHR DJ, et al. Scientific collaboration results in higher citation rates of published articles[J]. *Pharmacotherapy the journal of human pharmacology & drug therapy*, 2006, 26(6): 759-767.
- [35] GLÄNZEL W, SCHUBERT A. Double effort = double impact? A critical view at international co-authorship in chemistry[J]. *Scientometrics*, 2001, 50(2): 199-214.
- [36] GLÄNZEL W. Seven myths in bibliometrics about facts and fiction in quantitative science studies[J]. *Collnet journal of scientometrics & information management*, 2008, 2(1): 9-17.
- [37] ZHONG Z. Correlation analysis between co-authorship scale and performance of Web of Science journal papers in agricultural economics and policy[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2014, 25(12): 1513-1518.
- [38] BORNMANN L, SCHIER H, MARX W, et al. What factors determine citation counts of publications in chemistry besides their quality?[J]. *Journal of informetrics*, 2012, 6(1): 11-18.
- [39] WEBSTER GD, JONASON PK, SCHEMBERTO. Hot topics and popular papers in evolutionary psychology: analyses of title words and citation counts in evolution and human behavior, 1979-2008[J]. *Evolutionary psychology*, 2009, 7(3): 348-362.
- [40] VANCLAY JK. Factors affecting citation rates in environmental science[J].

Journal of informetrics, 2013, 7(2): 265-271.

- [41] FU Z. Citation characteristics of non-citable documents in international authoritative journals and their contribution to impact factor[J]. Chinese Journal of Scientific and Technical Periodicals, 2016, 27(3): 324-329.
- [42] KULKARNI AV, BUSSE JW, SHAMS I. Characteristics associated with citation rate of the medical literature[J]. Plos one, 2007, 2(5): e403.
- [43] ROBSON BJ, MOUSQUES A. Can we predict citation counts of environmental modeling papers? Fourteen bibliographical and categorical variables predict less than 30% of the variability in citation counts[J]. Environmental modelling & software, 2016, 75(1): 94-104.
- [44] HASLAM N, KOWAL P. Predicting long-term citation impact of articles in social and personality psychology[J]. Psychol rep, 2010, 106(3): 891-900.
- [45] SCHLOEGL C, GORRAIZ J. Comparison of citation and usage indicators: the case of oncology journals[J]. Scientometrics, 2010, 82(3): 567-580.
- [46] FU Z. Empirical study on the relationship between paper volume and citations under different citation time windows—From the perspective of papers and journals[J]. Journal of Intelligence, 2017, 36(7): 128-133.
- [47] DIDEGAH F, THELWALL M. Determinants of research citation impact in nanoscience and nanotechnology[J]. Journal of the Association for Information Science & Technology, 2013, 64(5): 1055-1064.
- [48] FINARDI U. Correlation between journal impact factor and citation performance: an experimental study[J]. Journal of informetrics, 2013, 7(2): 357-370.
- [49] SEGLE PO. Why the impact factor of journals should not be used for evaluating research[J]. British medical journal, 1997, 314(7079): 498-502.
- [50] WANG X, LIU C, MAO W, et al. The open access advantage considering citation, article usage and social media attention[J]. Scientometrics, 2015, 103(2): 555-564.
- [51] VIEIRA ES, GOMES JANF. Citations to scientific articles: its distribution and dependence on the article features[J]. Journal of informetrics, 2010, 4(1): 1-13.
- [52] MARTÍNEZ MA, HERRERA M, CONTRERAS E, et al. Characterizing highly cited papers in social work through H-classics[J]. Scientometrics, 2015, 102(2): 1713-1729.
- [53] JAHANDIDEH S, ABDOLMALEKI P, ASADABADI EB. Prediction of future citations of a research paper from number of its internet downloads[J]. Medical hypotheses, 2007, 69(2): 458-459.
- [54] DING Z. Correlation analysis between usage counts and citation frequency based on Web of Science[J]. Chinese Journal of Scientific and Technical Periodicals, 2017, 28(12): 1166-1170.
- [55] MOED HF, HALEVI G. On full text download and citation distributions in scientific-scholarly journals[J]. Journal of the Association for Information Science & Technology, 2016, 67(2): 412-1065.
- [56] BRODY T, HARNAD S, CARR L. Earlier web usage statistics as predictors of later citation impact[J]. Journal of the Association for Information Science & Technology, 2006, 57(8): 1060-1072.

- [57] GUERRERO-BOTE VP, MOYA-ANEGÓN F. Relationship between downloads and citations at journal and paper levels, and the influence of language[J]. *Scientometrics*, 2014, 101(2): 1043-1295.
- [58] LIPPI G, FAVALORO EJ. Article downloads and citations: is there any relationship?[J]. *Clinica chimica acta*, 2013, 415: 107-195.
- [59] WANG X, FANG Z, SUN X. Usage patterns of scholarly articles on Web of Science: a study on Web of Science usage count[J]. *Scientometrics*, 2016, 109(2): 917-926.
- [60] SHENG L. Analysis of hot papers in ophthalmology journals in Scopus database[J]. *Chinese Journal of Medical Library and Information Science*, 2014, 23(2): 60-62, 67.
- [61] KUANG D. Examining the influence of scientific journals through Altmetrics hot papers—Taking Altmetric.com top100 papers as an example[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2016, 27(11): 1188-1194.
- [62] LIU Z, CHEN Y, HOU H. Atlas of Frontiers in Technical Science and Strategy for a Powerful Nation[M]. Beijing: People's Publishing House, 2012.
- [63] XIE F, WEI B, LIN J. Statistical Analysis and Application of Zero-Inflated Data[M]. Beijing: Science Press, 2013.
- [64] CHETTY R, FRIEDMAN JN, LETH-PETERSEN S, et al. Active vs. passive decisions and crowd-out in retirement savings accounts: evidence from Denmark[J]. *Quarterly journal of economics*, 2014, 129(3): 1141-1219.

## Author Contributions

**Song Chao:** Conceptualization, data processing, writing, revision

**Chen Yue:** Conceptualization, revision (Corresponding author: chenyueldut@163.com)

**Wang Ling:** Data collection

**Zuo Jia:** Data collection

**Liu Zeyuan:** Revision suggestions

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*