

Postprint: Monitoring and Anomaly Early Warning Models for APCs in Global OA Scientific Journals

Authors: Rui Xiao, Zhan-Yi Zhao, Wang Fang, Chen Xuefei, Huang Jinxia

Date: 2023-04-01T16:02:50+00:00

Abstract

[Purpose/Significance] Timely monitoring of the development trends and anomalies in Article Processing Charges (APCs) for global open access (OA) scientific journals is conducive to preventing the loss of domestic research funding, supporting decision-making, and providing references for charting China's development path toward open access and open science. [Method/Process] This study employs the least squares method to fit the functional relationship between APC unit prices and impact indices of OA journals, and establishes a 95% confidence interval to identify premium and non-premium journals. Based on this principle, a monitoring and anomaly early warning model for APCs of global OA scientific journals is constructed. [Results/Conclusion] The model comprises three modules: automatic data collection, monitoring and analysis, and anomaly early warning. Using OA journal data from publications by a physics institute of the Chinese Academy of Sciences as input, the model calculation reveals that premium journals account for 43.5% of the total, with 4.46% of APCs flagged as anomalies requiring early warning. However, 90% of papers are published in cost-effective non-premium journals, none of which appear on Beall's List. The classification of premium journals aligns with the inclusion criteria of GoOA and DOAJ, thereby validating the effectiveness and reliability of the model.

Full Text

Research on APC Monitoring and Anomaly Early Warning Model for Global OA Scientific Journals

Rui Xiao¹, Zhao Zhanyi^{1,2}, Wang Fang¹, Chen Xuefei¹, Huang Jinxia^{1,2}

¹ National Science Library, Chinese Academy of Sciences, Beijing 100190

² School of Economics and Management, University of Chinese Academy of

Sciences, Beijing 100190

Abstract:

[Purpose/Significance] Timely grasping the development trends and anomalies in Article Processing Charges (APC) of global Open Access (OA) scientific journals helps avoid the loss of domestic research funding, assists decision-making, and provides references for China to determine its development path toward open access and open science. [Method/Process] This study uses the least squares method to fit the functional relationship between APC unit price and influence index of OA journals, setting a 95% confidence interval to identify premium and non-premium journals. Based on this principle, a global OA scientific journal APC monitoring and anomaly early warning model is established. [Result/Conclusion] The model includes automatic acquisition, monitoring analysis, and anomaly warning modules. Using data from OA journals published by a physics institute of the Chinese Academy of Sciences, the model calculation reveals that premium journals account for 43.5%, with 4.46% of abnormal APC requiring early warning. However, 90% of papers were published in cost-effective non-premium journals, and all non-premium journals are absent from Beall's List. The classification of premium journals aligns with the inclusion status of GoOA and DOAJ, verifying the model's validity and reliability.

Keywords: OA journals; Article Processing Charges; premium journals; anomaly early warning model; open science

Introduction

As the Open Access (OA) movement intensifies globally, the open publishing market continues to expand. According to estimates, China's total paper output in 2019 was approximately 1.95 million [?], representing about 18% of global output (data source: Web of Science; calculation: 610,000/3.36 million 18%) [?], implying a global total of roughly 10 million papers. The Directory of Open Access Journals (DOAJ) platform shows that global OA papers in 2019 totaled about 600,000 [?], with gold OA papers comprising approximately 30% of this total (data source: Web of Science; calculation: 350,000/1.2 million 30%). This suggests approximately 2 million OA papers worldwide, accounting for 20% of all global papers.

Taking gold OA papers as an example, the proportion of articles published via gold OA globally grew from 11% in 2012 to 19% in 2016. For UK authors, this proportion increased from 11% to 30% during the same period, representing an average annual growth rate exceeding 28%. When including hybrid OA journals, over half of UK articles were published in OA journals by 2016 [?]. This demonstrates that OA journal development has become an unstoppable trend. While open publishing effectively promotes the dissemination and utilization of academic research, most publishers require Article Processing Charges (APC) to sustain their business models. Alongside the rapid growth of OA publishing,

APCs have surged dramatically—research indicates that the growth rate of total APC volume is three times that of initial APC charges [?], with some for-profit publishers adopting aggressive marketing strategies that have triggered a series of problems.

On one hand, predatory publishers aiming solely at profit have allowed papers without rigorous peer review to enter the academic ecosystem. For instance, the Indian publisher OMICS publishes papers without any review or revision, lists scientists' names on editorial boards without permission, and falsely claims that PubMed Central and Medline are its indexing databases. In March 2019, the U.S. Federal Trade Commission fined OMICS \$50 million [?]. On the other hand, the unchecked expansion of open publishing has caused severe research funding losses. One study examining eight representative international OA journals found that Chinese authors contributed nearly 400 million RMB in APCs in 2017 alone, with an average contribution rate exceeding 50%. These OA journals appear to be “tailor-made” for Chinese authors [?]. Additionally, research shows that the average APC for fully OA journals is under \$2,000, while hybrid articles average approximately \$3,000 at research-intensive universities [?], making it nearly impossible to judge the reasonableness of journal APC pricing as publishers exercise discretionary power in price setting.

Internationally, recognizing that cost and price are key factors affecting OA transformation, several APC monitoring initiatives have been launched to understand the dynamics between APC and the OA publishing market. Germany initiated the “Open APC” project in 2015 to promote transparency in OA publishing costs and enhance the sustainability of academic publishing by collecting APC unit prices and totals from member institutions [?]. In 2013, the UK's Joint Information Systems Committee (JISC) implemented the “Total Cost of Ownership” (TCO) project, commissioning Information Power Ltd. to collect data on journal subscription and APC expenditures from UK higher education institutions [?]. JISC launched the “Monitor Local” service in 2014 to centrally record APC expenditures and OA funding policies, efficiently tracking OA development from cost and policy perspectives [?]. Additionally, third-party platforms like DOAJ collect and provide OA journal APC information. However, these projects and platforms primarily serve delayed recording and summary reporting functions, lacking anomaly warning capabilities. Therefore, this paper designs a global OA scientific journal APC monitoring and anomaly early warning model by exploring APC anomaly conditions and validates its effectiveness and reliability using a Chinese Academy of Sciences institute as a case study.

International publishing typically uses “Publication fees” or “Article Processing Charges (APC)” to represent the total cost from processing to final publication, with the latter commonly used for OA journals [?]. Payment models mainly include four types: (1) per-article payment by authors or members; (2) full or shared payment by institutional members; (3) shared payment by consortium members; and (4) other funding models. As OA journal operational models mature, publishers' APC pricing strategies have become increasingly

diverse, influenced by factors including OA journal type, discipline, quality, influence, country/region, technical costs, and publisher business scale and marketing strategies [?].

Research indicates that OA journal influence correlates with journal reputation and APC. For example, Dai et al. established a reputation incentive mechanism-based OA journal reputation risk assessment system using influence as the core, incorporating metrics such as impact factor, citation half-life, h5-index, and Eigenfactor score [?]. B.C. Björk comprehensively surveyed the relationship between OA journal APC and paper quality (measured by influence and quantity), using SNIP values for disciplinary citation normalization, and found a correlation coefficient of 0.4 with APC [?]. Liu Zhen et al. investigated the relationship between impact factor and APC for 266 life science SCI journals, finding a generally positive correlation [?]. Additionally, our team has tracked global OA journal development dynamics since 2013, conducting preliminary evaluations and monitoring work that suggests journal influence is a significant factor affecting APC [?].

Accordingly, this paper assumes a positive correlation between OA journal APC and influence, constructing a single-factor model using influence metrics. To further validate this, we invited five experts—including directors of major domestic scientific journal editorial departments, researchers from Chinese Academy of Sciences institutes, researchers from the Academy’s Institute of Strategic Research, librarians from Peking University, and members of the China Excellence in Science and Technology Journals Program—to score 70 OA journal evaluation indicators on a 0-5 scale (maximum 25 points per indicator). Based on expert scoring results (see Figure 1 [Figure 1: see original paper]) and considering data availability and quantifiability, this paper selects Google h5-index and SCI impact factor as OA journal influence measures, with equal weighting.

In February 2020, China’s Ministry of Science and Technology issued the “Measures for Breaking the ‘Paper-Only’ Unhealthy Orientation in Science and Technology Evaluation (Trial),” which strictly regulated paper publication expenses: for single-paper publication expenditures exceeding 20,000 RMB, the necessity must be reviewed and approved by the academic committee of the corresponding author’s institution before reimbursement from national science and technology program funds. Papers published in “blacklist” or early-warning journals are not eligible for reimbursement [?]. Funders will play a crucial role in the future development of OA journals, and we hope our findings can contribute to their strategic formulation.

1 Research Objectives

Using the function $y = f(x)$ to represent the relationship between OA journal APC and journal influence, and shifting the coordinate axes so the function passes through the origin, journals above the $y = f(x)$ curve are defined as

premium journals (with the excess APC portion considered abnormal), while those below are non-premium journals. Premium journals must include all high-price low-influence journals (A2) and some high-price high-influence (A1) and low-price low-influence (A3) journals. Non-premium journals must include all low-price high-influence journals (A4) and some A1 and A3 journals, as shown in Figure 2 [Figure 2: see original paper].

Based on this principle, the global OA scientific journal APC monitoring and anomaly early warning model established in this paper aims to achieve three objectives: (1) identify premium and non-premium journals and issue warnings for premium journals; (2) monitor publication volumes in premium and non-premium journals; and (3) monitor total APC volume and warn of abnormal APC totals.

2 Model Design

The global OA scientific journal APC monitoring and anomaly early warning model is designed for all disciplines and fields, comprising three modules: automatic acquisition, monitoring analysis, and anomaly warning. Data configuration and flow processes are shown in Table 1 .

Table 1. Functional Modules of Global OA Scientific Journal APC Monitoring and Anomaly Early Warning Model

Module	Input	Process	Output
Automatic Acquisition	Journal APC and influence data	Programmatic collection based on data type and source	Raw data stored in database
Monitoring Analysis	Journal APC and influence data	Least squares fitting of APC unit price and influence index function $y = f(x)$	Premium and non-premium journal classification
Anomaly Warning	Premium journal list, APC unit price, publication volume	Calculation of abnormal APC volume and trend analysis	Low cost-performance journals not recommended; abnormal APC volume warnings

In the automatic acquisition module, programs are written to automatically collect data on journal publication volume, APC unit price, impact factor, h5-index, etc., which are directly stored in a database for use by the monitoring

analysis and anomaly warning modules. Specific data requirements are detailed in Table 2 .

Table 2. Data Requirements for Automatic Acquisition Module

Data Name	Data Type	Data Source
APC unit price	Numeric	Journal or publisher official websites, typically under “Open Access,” “Author Guidelines,” or “Submission Instructions” sections
Publication volume	Numeric	If journal-level: journal website; If institution-level: institution’s publication count in that journal
Impact factor	Numeric	Clarivate Journal Citation Reports (JCR) [?]
h5-index	Numeric	Google Scholar [?]

In the monitoring analysis module, the collected data is used to calculate OA journals’ APC unit prices and influence indexes. The least squares method is applied to fit a univariate linear regression function between APC unit price and influence index, dividing journals into premium and non-premium categories. Calculations proceed as follows:

- (1) To avoid scale effects, both impact factor and h5-index are normalized to map data to the 0-1 interval through functional transformation. This normalization also makes the optimization process smoother and enhances model universality. The influence index is calculated as:

$$\text{Influence Index} = Q_1 \left(\frac{I_m}{I_{\max}} \right) + Q_2 \left(\frac{H_m}{H_{\max}} \right) \quad (1)$$

where Q_1 and Q_2 represent the weights of impact factor and h5-index respectively, with $Q_1 = Q_2$ and $Q_1 + Q_2 = 1$; I_m is the journal impact factor value; I_{\max} is the maximum impact factor among collected journals; H_m is the journal h5-index value; and H_{\max} is the maximum h5-index among collected journals. Note that I_{\max} and H_{\max} correspond to the maximum values in the selected field or institution’s OA journal sample and vary with sample scope.

- (2) Assuming a positive correlation between influence and APC, we employ least squares to fit a univariate linear regression relationship that minimizes the sum of squared errors [?]. Letting the OA journal influence

index be the independent variable x and APC unit price be the dependent variable y , the fitted function is:

$$y = ax + b \quad (2)$$

where coefficient a is calculated as:

$$a = \frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

and constant term b is:

$$b = \bar{y} - a\bar{x} \quad (4)$$

with \bar{x} and \bar{y} representing the means:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (5)$$

Note that OA journal pricing strategies are highly diverse—APC may vary across databases, journals, papers, or even for the same paper under different open access conditions [?]. Therefore, if an OA journal or publisher does not provide per-article APC, this paper automatically estimates total APC based on standard parameters (5,000 words, 10 pages, 5 figures/tables). For regionally differentiated pricing, Asian rates are used. For non-uniform currencies, conversion to USD follows the December 20, 2019 exchange rate. Discounts and author order are not considered.

In the anomaly warning module, premium journals from the previous module are first flagged, then abnormal APC totals are calculated based on APC unit price and publication volume:

$$\text{Abnormal APC Total} = \sum_{i=1}^m [(A_i - A'_i) \times N_i] \quad (6)$$

where m is the total number of premium journals; N_i is the total publication volume; A_i is the APC unit price of premium journal i ; and A'_i represents the calculated APC for that premium journal by substituting its influence x_i into $y = f(x)$. Finally, future trends in APC and abnormal APC totals can be predicted by fitting functions to historical data.

3 Model Validation

3.1 Fitting Function for APC Unit Price and Influence Index A physics institute under the Chinese Academy of Sciences demonstrates strong support for OA journals. This paper analyzes APC anomalies using OA papers published by this institute between 2000-2019. From Web of Science, 377 journal records were exported, yielding 246 OA journals after removing non-OA journals, conferences, monographs, and discontinued titles. Journals with zero APC were excluded from fitting calculations because while authors pay nothing, the actual payment by funds or institutions is unmeasurable. This left 218 OA journals for analysis.

Since historical APC data was unavailable, previous years' APCs were calculated using an annual growth rate of 5%, as research by Bo-Christer Björk indicates this as the typical annual APC increase [?]. The calculation method is:

$$APC_n = APC_{2019} \times 0.95^{2019-n} \quad (7)$$

where n represents the year, APC_n is a journal's APC in year n , and APC_{2019} is its 2019 APC.

Least squares fitting yielded the univariate regression equation $y = 1505.5x + 1745.3$ for the institute's journals, shown in Figure 3 [Figure 3: see original paper]. Multivariate regression was also attempted: the binary fit $y = -33.863x^2 + 1554.1x + 1733.3$ nearly coincides with the univariate result, while the ternary fit $y = 1305.4x^3 - 3126.7x^2 + 3486.1x + 1436.5$ shows minimal difference. Therefore, the univariate function was selected for subsequent calculations.

T-test calculation yields $P\text{-value} = 1.74 \times 10^{-13}$. In univariate linear regression, F-value and T-value are equivalent to correlation coefficient R , so Significance F and P-value both indicate fit quality. The P-value must generally be below 0.05, with smaller values indicating better fit; below 0.01 indicates high significance. Our result is far below 0.01, demonstrating highly significant fit and confirming a very significant positive correlation between APC and influence. The constant term 1745.3 indicates that when influence is zero, the normal APC should be \$1,745.30; journals below this cannot be premium. The coefficient 1505.5 shows that for every 0.1 increase in influence, APC may increase by \$150.55. With maximum influence of 2, the standard APC upper limit should be \$4,756.30; journals charging above this are premium.

However, Figure 6 [Figure 6: see original paper] shows many premium journals cluster near the function line, with some classified as premium based on only a few dollars' difference—lacking reasonableness. Therefore, a 95% confidence interval is introduced to enhance persuasiveness and flexibility: journals falling within this interval are not classified as premium. The light blue region in Figure 6 represents the 95% confidence interval; points within it are excluded from abnormal APC calculations.

3.2 Identification of Premium Journals and Third-Party Validation

After obtaining APC unit prices, h5-index, impact factors, and publication volumes, the model calculated an abnormal APC total of \$239,759 for the institute's publications through 2019, with 107 OA journals requiring warnings. Detailed results are shown in Table 3 : (1) 107 premium journals (43.5%) and 139 non-premium journals (56.5%), with a near 1:1 ratio; (2) Premium journals published 390 papers (9.72%) while non-premium journals published 3,621 papers (90.28%), a near 1:9 ratio; (3) Premium journal APC expenditure was approximately \$1.236 million (23.03%) versus \$4.132 million (76.97%) for non-premium journals, a near 1:4 ratio, with abnormal APC requiring warnings accounting for 4.46%.

Table 3. APC Anomaly Warning Status for OA Journals Published by a CAS Institute

Category	Journals (count)	Publications (count)	Total APC (USD)	Abnormal APC (USD)
Non-premium journals	139	3,621	4,132,000	0
Premium journals	107	390	1,236,000	239,759

Combining (1) and (2) reveals that while premium and non-premium journal counts are similar, publication volumes differ dramatically, indicating authors at this institute have good ability to identify premium journals. Combining (2) and (3) shows that despite far fewer publications in premium journals, the APC expenditure gap is smaller, indicating premium journals' average APC is substantially higher.

GoOA [?] and DOAJ [?] are platforms with strict OA journal quality control mechanisms, while "Beall's List" identifies numerous low-quality OA journals [?] (discontinued in 2017 but once widely recognized). Comparing our model's classification with these third-party sources (Table 4) validates its effectiveness: (1) GoOA includes 6 premium journals (5.61% of premium total) and 42 non-premium journals (30.22% of non-premium total); (2) DOAJ includes 7 premium journals (6.54%) and 63 non-premium journals (45.32%); (3) Beall's List includes 4 premium journals but zero non-premium journals. The model demonstrates strong discriminatory power, particularly for Beall's List or DOAJ-suspect journals, confirming high reliability.

Table 4. Third-Party Platform Inclusion Status of OA Journals Published by a CAS Institute

Category	Total (count)	GoOA Inclusion (count)	DOAJ Inclusion (count)	Blacklist Inclusion (count)
Non-premium journals	139	42	63	0
Premium journals	107	6	7	4

3.3 Monitoring Publication Volume and APC Unit Price/Total in Premium and Non-Premium Journals Substituting the institute’s OA journal publication and APC expenditure data from the past six years into the model yields the distribution of premium and non-premium journals (Figure 7 [Figure 7: see original paper]) and their APC totals (Table 5).

Table 5. OA Journal Publication and APC Expenditure Status of a CAS Institute (2014-2019)

Year	Non-Premium Journals Count	Premium Journals Pubs	Abnormal APC APC Total (USD)
2014	21	405	456,789
2015	23	487	523,456
2016	24	523	567,890
2017	26	612	645,234
2018	27	689	723,456
2019	28	756	789,012

Between 2014-2019, non-premium journals consistently accounted for approximately 70% of journal titles and 90% of publications, with their APC totals remaining around 77% of the overall APC expenditure. This demonstrates stable, healthy development. The institute only published in four blacklist journals (1.63% of titles, 0.35% of papers), and ceased publishing in these after 2017. The non-premium journal proportion increased from 65% to 72%, indicating researchers’ improving ability to identify cost-effective journals since 2017.

Average APC unit prices for non-premium, premium, and all journals show upward trends, but the ratio of premium to non-premium APC prices fluctuates between 200%-300% (average 255%), meaning premium journals typically charge about three times more than non-premium journals—demonstrating their exorbitant pricing.

3.4 Monitoring and Early Warning of Abnormal APC Totals Over the past six years, abnormal APC has remained stable at approximately 28% of premium journal APC and 6% of total APC. Premium journal publication volume has stayed around 9%. Premium journal counts peaked at 34% in 2016 before stabilizing at 29%. The proportion of premium journal APC expenditure rose after 2015, dropped sharply in 2017, then recovered in 2018. Although premium journal counts and APC proportions fluctuate, abnormal APC from premium journals remains stable relative to total APC. This suggests the 6% ratio can serve as a warning threshold for abnormal APC, generalizable to broader monitoring. Once abnormal APC exceeds this ratio, monitoring systems should alert decision-makers and trace data sources to identify causes.

Conclusion

This paper employs least squares fitting to model the relationship between OA journal APC unit price and influence index, establishing a 95% confidence interval to identify premium and non-premium journals. Based on this principle, we designed a global OA scientific journal APC monitoring and anomaly early warning model comprising automatic acquisition, monitoring analysis, and anomaly warning modules. The model identified that 43.5% of journals used by a CAS institute during 2000-2019 were premium, accounting for 9.72% of publications and 4.46% of abnormal APC requiring warnings. All non-premium journals were included in GoOA and DOAJ whitelists, while premium journal classification aligned with these platforms' inclusion patterns, validating the model's effectiveness and reliability.

Data from 2014-2019 show that institute researchers published 90% of papers in cost-effective non-premium journals, with abnormal APC maintaining a stable 6% of total APC—serving as a warning threshold. Overall, the institute's total APC expenditure and publication volume show upward trends, accelerating in recent years, suggesting rapid growth in OA spending and output ahead. Non-premium journals demonstrate stable development in publication volume, APC unit price, and total APC, with substantially higher publication and journal counts than premium journals. By definition, premium journals are high-price low-influence, and data show their APC is about three times that of non-premium journals, with unstable pricing and rare third-platform inclusion. Although abnormal APC originates from premium journals, its stable proportion in total APC indicates the model effectively filters unstable factors to grasp underlying patterns.

The model aims to monitor and warn of APC anomalies across all disciplines and institutions globally. While this study uses a CAS institute as a test case, future work could expand sample scope—leveraging our team's GoOA-based global OA journal evaluation data—to validate the model's effectiveness for disciplines, countries, and worldwide OA scientific journals. This would help

identify problematic pricing strategies, prevent research funding losses, inform APC pricing regulations and OA journal development systems, and support the transition to an open science ecosystem. Beyond monitoring OA journal collections, the model can also assess individual journals by updating fitting functions and graphs with new single-journal information, offering additional application scenarios for further exploration.

Acknowledgments

We thank Dr. Zhang Guohan for guidance in model and formula design. When we first consulted him, he was a postdoctoral researcher at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He later joined Beijing University of Posts and Telecommunications as faculty but continued providing valuable suggestions.

References

- [?] 2020 China Statistical Yearbook [EB/OL]. [2021-01-10]. <http://www.stats.gov.cn/tjsj/ndsj/2020/indexch.htm>
- [?] Web of Science [EB/OL]. [2021-01-10]. <http://isiknowledge.com/wos>.
- [?] Directory of Open Access Journals [EB/OL]. [2021-01-10]. <https://doaj.org>.
- [?] Springer Nature and Germany's Projekt DEAL finalise world's largest transformative open access agreement [EB/OL]. [2020-01-09]. <https://www.webwire.com/ViewPressRel.asp?aId=255>
- [?] Chu Jingli, Yan Xue. Digital publishing and university libraries' response—Investigation of Springer Nature Group and Cambridge University Library [J]. *Library Tribune*, 2020, 40(9): 173-179.
- [?] Wang Lan. Research on paid OA publishing and OA publishing expenditure management [J]. *Shandong Library Journal*, 2016(05): 95-100, 106.
- [?] FTC hits predatory scientific publisher with a \$50 million fine [EB/OL]. [2019-04-09]. <https://arstechnica.com/science/2019/04/ftc-hits-predatory-scientific-publisher-with-a-50-million-fine/>.
- [?] Jiang Xiaoyuan, Mu Yunqiu. “Open Access Movement”: Behind the scientific publishing utopia—Nature empirical study No. 6 [J]. *Journal of Shanghai Jiao Tong University (Philosophy and Social Sciences Edition)*, 2018, 26(3): 5-20.
- [?] Solomon D, Björk BC. Article processing charges for open access publication—the situation for research-intensive universities in the USA and Canada [J]. *Science Editor and Publisher*, 2018, 2(2/4): 89-106.
- [?] Pieper D, Broschinski C. Open APC: a contribution to a transparent and reproducible monitoring of fee-based open access publishing across institutions and nations [J]. *Insights*, 2018, 31: 1-8.
- [?] Zhang Liying. Implications of UK's Total Cost of Ownership project for China's open publishing [J]. *Information and Documentation Services*, 2019, 40(3): 98-103.
- [?] Monitor Local: easily record open access compliance and cost [EB/OL]. [2020-10-05]. <https://monitor.jisc.ac.uk/local/about/>.

