

Analysis of Key Paths for Big Data Technology to Improve Financial Audit Efficiency

Yongxu Zhao *

International Business College, Dongbei University of Finance & Economics, Dalian, China

*Corresponding Author: zyx19904071714@outlook.com

ABSTRACT

Against the backdrop of digitalization, traditional financial audits face issues such as low efficiency, limited scope, and delayed risk identification. Big data technology offers a potential solution to these dilemmas. This study aims to systematically construct the key paths through which big data technology enhances financial audit efficiency and clarify the mechanism behind this efficiency improvement. Taking audit projects as Decision-Making Units (DMUs), the research first defines an input indicator system encompassing hardware investment, software investment, and human resource investment, as well as an output indicator system covering audit coverage rate, risk identification accuracy rate, and audit cycle reduction rate. Subsequently, the DEA-BCC model is introduced to measure the efficiency of 27 audit projects. The results show that the comprehensive technical efficiency of big data audit projects (0.892) is significantly higher than that of traditional audit projects (0.678), and this improvement is mainly achieved by enhancing pure technical efficiency (with a difference of 0.214). On the input side, problems such as hardware redundancy and mismatched human resource skills exist; on the output side, the most prominent shortcoming is audit coverage rate. The data analysis, data cleaning, and data collection modules have a significant impact on audit efficiency, with impact coefficients of 0.415, 0.352, and 0.287 respectively. The research findings can provide theoretical references and practical guidance for the digital transformation of audit institutions, helping to improve financial audit efficiency.

KEYWORDS

Big Data Technology; Financial Audit Efficiency; DEA-BCC Model; Key Paths

1. INTRODUCTION

In the wave of the digital era, big data technology has permeated various fields at an unprecedented pace, and the field of financial auditing has also been deeply affected. With its massive data processing capability, high-speed computing performance, and support for diverse data types, big data technology has brought new opportunities and transformations to financial audit work [1]. When confronted with the ever-expanding business scale, complex transaction structures, and massive data information, the traditional financial audit model has gradually revealed drawbacks such as low efficiency, limited audit scope, and delayed risk identification. The application of big data technology is expected to break through these dilemmas and achieve a significant improvement in financial audit efficiency [2].

From the perspective of research status, many scholars have conducted discussions on the application of big data technology in financial auditing [1-3]. Some studies focus on how big data technology changes the methods of audit data collection and analysis, pointing out that through automated data collection tools and advanced data mining algorithms, valuable audit clues can be quickly screened

out from multi-source heterogeneous data, thereby improving the efficiency of obtaining audit evidence [4]. Other studies pay attention to the impact of big data technology on audit risk assessment, believing that risk assessment models based on big data can more accurately identify potential risk points and enhance the ability of audit risk management and control [5]. However, existing studies mostly focus on describing technical application scenarios in terms of systematically constructing the key paths for improving financial audit efficiency based on big data technology, and there is a lack of quantitative analysis on the mechanism of audit efficiency improvement.

This study takes the mechanism of audit efficiency improvement as the core research object. It first defines the input and output indicator system of big data technology in financial auditing, including hardware input (server computing power, storage capacity), software input (number of algorithm models, data processing software authorization), human resource input (number of professional personnel in big data auditing), and output indicators (audit coverage rate, risk identification accuracy rate, audit cycle reduction rate). Secondly, the DEA-BCC model is introduced to measure the audit efficiency before and after the application of big data technology, and the effectiveness of technology application is verified through comparative analysis. Finally, based on the empirical results, the key paths for big data to improve audit efficiency are identified, including the real-time transformation of the data collection link, the application of automated algorithms in the data cleaning link, the embedding of machine learning models in the data analysis link, and the application of visualization technology in the audit result presentation link. Corresponding optimization suggestions are put forward to provide theoretical references and practical guidance for the digital transformation of audit institutions.

2. METHODOLOGY

2.1. DEA-BCC Model

The DEA (Data Envelopment Analysis) model is a non-parametric evaluation method based on relative efficiency, which is suitable for efficiency measurement scenarios with multiple inputs and multiple outputs [6]. As a classic extended model of DEA, the BCC model relaxes the assumption of constant returns to scale of the CCR model, allowing Decision-Making Units (DMUs) to have variable returns to scale, and can more accurately measure pure technical efficiency and scale efficiency. In this study, audit projects are taken as DMUs, and the BCC model is constructed as follows.

Assume there are n decision units ($DMU_1, DMU_2, \dots, DMU_n$). Each DMU includes m types of input factors (x_1, x_2, \dots, x_m) and s types of output factors (y_1, y_2, \dots, y_s). For the j_0 -th DMU , its efficiency evaluation model is:

$$\left\{ \begin{array}{l} \min \theta - \varepsilon \left(\sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right) \\ \text{s.t.} \sum_{j=1}^n \lambda_j x_{ij} + s_i^- = \theta x_{ij_0}, i = 1, 2, \dots, m \\ \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ = y_{rj_0}, r = 1, 2, \dots, s \\ \sum_{j=1}^n \lambda_j = 1 \\ \lambda_j \geq 0, j = 1, 2, \dots, n \\ s_i^- \geq 0, s_r^+ \geq 0 \end{array} \right.$$

Among them, θ is the comprehensive technical efficiency value of the j_0 -th DMU ($0 \leq \theta \leq 1$), λ_j is the weight coefficient of the DMU, s_i^- is the input slack variable, s_r^+ is the output slack variable, and ε is a non-Archimedean infinitesimal (usually taking 10^{-6}). When $\theta=1$ and $s_i^- = s_r^+ = 0$, the DMU is in a DEA-efficient state; when $\theta=1$ but $s_i^- \neq 0$ or $s_r^+ \neq 0$, the DMU is in a DEA-weakly efficient state; when $\theta < 1$, the DMU is in a DEA-inefficient state.

The BCC model decomposes the comprehensive technical efficiency (TE) into pure technical efficiency (PTE) and scale efficiency (SE).

$$TE = PTE \times SE$$

Among them, pure technical efficiency reflects the efficiency of the decision unit in utilizing input factors under the existing technical level, and scale efficiency reflects whether the decision unit is in the optimal production scale. If the scale efficiency < 1 , the state of returns to scale can be further judged by $\sum_{j=1}^n \lambda_j$. When $\sum_{j=1}^n \lambda_j < 1$, it is in the state of increasing returns to scale; when $\sum_{j=1}^n \lambda_j > 1$, it is in the state of decreasing returns to scale.

2.2. Construction of Indicator System

In the research scenario of big data-driven financial audit transformation, the scientific construction of an input-output indicator system is the basis for accurately measuring audit efficiency and identifying the path of technical empowerment. This study focuses on audit projects after the application of big data technology, conducts an in-depth analysis of resource input and value output in each link of the audit process, and constructs a multi-dimensional and refined input-output indicator system to provide a clear measurement framework for subsequent efficiency evaluation.

2.2.1. Input Indicators

Hardware Input (x_1): As the scale of audit data expands from the GB level to the PB level, high-performance hardware has become the basic guarantee for data processing. In this study, hardware input is defined as the purchase and operation and maintenance costs of servers, storage devices, etc., required for big data auditing, with ten thousand yuan as the measurement unit. It specifically includes the cost allocation of enterprise-level servers, distributed storage devices, supporting network equipment, and computer room environment facilities, comprehensively reflecting the physical resource input supporting big data auditing.

Software Input (x_2): Big data auditing relies on professional software tools to realize data collection, cleaning, analysis, and visualization, so software input is one of the core elements of technical

application. This indicator is measured in ten thousand yuan, including data processing software, algorithm model authorization, and the expansion of big data function modules of audit software, accurately measuring the resource consumption at the software level.

Human Resource Input (x_3): Although big data technology improves the level of audit automation, professional human resources are still the key to promoting the implementation of technology and ensuring audit quality. With person-days as the measurement unit, it focuses on the input of professional personnel in big data auditing, including compound talents with big data analysis skills and professional audit knowledge. Specifically, it includes data engineers, algorithm engineers, and big data auditors. The number of person-days of human resource input is obtained by summing up the working hours of personnel in different roles in the audit project, reflecting the characteristics of resource input combining human resources and technology.

2.2.2. Output Indicators

Audit Coverage Rate (y_1): Traditional auditing is limited by sampling methods, making it difficult to conduct verification on the entire amount of data, resulting in audit blind spots. With the support of big data technology, audit coverage rate has become a key indicator to measure the expansion of the audit scope, with % as the unit, defined as the proportion of the entire amount of data verified. That is, through big data tools, the enterprise's financial data, business data, and external related data are comprehensively collected and analyzed, and the ratio of the actual verified data volume to the total amount of data that should be verified is calculated. It reflects the ability of big data technology to break through the limitations of sampling and realize full-scale auditing. The higher the coverage rate, the lower the risk of audit omissions.

Risk Identification Accuracy Rate (y_2): Accurately identifying risks is the core goal of financial auditing. Big data technology improves the efficiency of risk identification through multi-dimensional data analysis. This indicator is measured in %, referring to the proportion of truly identified risks. That is, using big data analysis models, the number of risk clues screened out from massive data that are confirmed as real risks through subsequent audit verification, compared with the total number of identified risk clues. Compared with the traditional audit method of risk identification relying on experience judgment, this indicator reflects the advantage of big data technology in the accuracy of risk identification. The higher the accuracy rate, the better the input-output ratio of audit resources in risk verification.

Audit Cycle Reduction Rate (y_3): The audit cycle directly affects the timeliness of audit services and the degree of interference with enterprise operations. Big data technology accelerates the audit process through automated and intelligent processes. Measured in %, it calculates the proportion of time reduced compared with the traditional audit model. Traditional auditing requires manual verification and data analysis one by one, with cumbersome processes; big data auditing realizes rapid data collection, automatic analysis, and risk early warning with the help of data batch processing, real-time monitoring, and other technologies, effectively compressing the time of the audit preparation, implementation, and reporting stages. This indicator intuitively shows the value of big data technology in improving audit efficiency and accelerating audit turnover.

2.2.3. Data Sources and Processing

In this study, 27 audit projects from domestic accounting firms during 2023-2024 are selected as samples (including 20 projects applying big data technology and 7 traditional audit projects). The Min-Max standardization method is used to preprocess the original data to eliminate dimensional differences, and the formula is as follows:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

Among them, x'_{ij} is the standardized data, x_{ij} is the original data, and $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of the j -th indicator respectively.

3. RESULTS

It can be seen from Table 1 that the comprehensive technical efficiency of big data audit projects (0.892) is significantly higher than that of traditional audit projects (0.678), with a difference of 0.214, indicating that big data technology can effectively improve audit efficiency. From the perspective of efficiency decomposition, the difference in pure technical efficiency (0.214) is much larger than that in scale efficiency (0.013), which shows that big data technology improves audit efficiency mainly by optimizing the efficiency of technical application (such as algorithm iteration and data processing automation) rather than by scale expansion. The DEA efficiency rate of big data audit projects (65.0%) is 2.37 times that of traditional audit projects (25.0%), and there are more projects with increasing returns to scale (12 projects), reflecting that there is still room for scale expansion in the application of big data technology.

Table 1. Comparison of Overall Efficiency between Big Data and Traditional Audit Projects

Audit Type	Big Data Audit	Traditional Audit
Comprehensive Technical Efficiency (TE)	0.892	0.678
Pure Technical Efficiency (PTE)	0.935	0.721
Scale Efficiency (SE)	0.954	0.941
DEA Efficiency Rate (%)	65	25
Returns to Scale State (Increasing/Decreasing/Constant)	12/3/5	8/7/5

From the statistical results of input-output slack variables of big data audit projects in Table 2, there are obvious problems of unbalanced resource allocation on the input side. The hardware input has the highest slack value (0.087 ± 0.023), and 45.2% of the projects cause resource waste due to server computing power redundancy and idle storage devices. The human resource input also has a slack value of 0.059 ± 0.015 , and 25.8% of the projects are limited by the mismatch between the skills of traditional audit personnel and technology. However, the software input has a relatively low slack value (0.012 ± 0.008), and only 30.6% of the projects have the problem of repeated authorization of basic software. On the output side, the most prominent shortcoming is the audit coverage rate, with a slack value of 0.036 ± 0.011 . 60.3% of the projects are affected by incomplete cross-border data collection and missing unstructured data. The slack values of risk identification accuracy rate (0.029 ± 0.009) and audit cycle reduction rate (0.008 ± 0.005) are relatively low, and only 30.1% and 8.7% of the projects have corresponding problems. This indicates that to improve the efficiency of big data auditing, priority should be given to the dynamic allocation of hardware resources, the training of compound human resources, and the optimization of data collection processes, while supplementing with the improvement of algorithm accuracy and the upgrading of report automation.

Table 2. Statistics of Input-Output Slack Variables in Big Data Audit Projects

Indicator	Mean±std.	Main Sources of Slack	Proportion of Problematic Projects (%)
Hardware Input	0.087±0.023	Server computing power redundancy, idle storage devices	45.2
Software Input	0.012±0.008	Repeated authorization of basic data processing software	30.6
Human Resource Input	0.059±0.015	Mismatch between traditional audit personnel's skills and technology	25.8
Audit Coverage Rate	0.036±0.011	Incomplete cross-border business data collection, missing unstructured data	60.3
Risk Identification Accuracy Rate	0.029±0.009	Insufficient accuracy of abnormal transaction identification algorithms, single feature dimension	30.1
Audit Cycle Reduction Rate	0.008±0.005	Low automation level in the report generation link	8.7

It can be seen from Table 3 regarding the impact coefficients of big data technology modules on audit efficiency that different technology modules have significant differences in their effects on audit efficiency.

Table 3. Impact Coefficients of Big Data Technology Modules on Audit Efficiency

Module	Impact Coefficient on TE	Impact Coefficient on PTE	Impact Coefficient on SE	P-Value
Data Collection Module	0.287	0.312	0.054	0.002
Data Cleaning Module	0.352	0.389	0.061	0
Data Analysis Module	0.415	0.453	0.072	0
Result Visualization Module	0.098	0.015	0.048	0.315
Constant Term	0.321	0.287	0.812	0.001
R ²	0.823	0.856	0.314	

The data analysis module has the highest impact coefficients on comprehensive technical efficiency (0.415) and pure technical efficiency (0.453), with a P-value of 0, indicating that it can accurately identify risks and optimize the audit process through algorithm iteration and multi-dimensional data mining, and is the core driving force for improving audit efficiency. The data cleaning module (comprehensive technical efficiency coefficient of 0.352, pure technical efficiency coefficient of 0.389) can eliminate noise data and unify data formats through automated algorithms, reducing manual intervention costs and making a significant contribution to efficiency improvement. The data collection module (0.287/0.312) can also effectively promote efficiency improvement by virtue of its advantages of real-time and multi-source collection, expanding the coverage of audit data. In contrast, the result visualization module has a relatively low impact coefficient (0.098/0.015), which may be because visualization is mostly used for auxiliary presentation in current audit scenarios and has a limited direct effect on core efficiency. The constant term is positively correlated with all efficiency dimensions (0.321/0.287/0.812), reflecting that the basic guarantees outside the technology modules (such as organizational management and institutional processes) also support audit efficiency. The R² values (0.823 for comprehensive efficiency, 0.856 for pure technical efficiency) indicate that the model has a strong explanatory power for efficiency differences, verifying the significant correlation

between technology modules and audit efficiency, and providing empirical evidence for subsequent focus on optimizing core modules to improve audit efficiency.

Table 4. Comparison of Audit Efficiency Before and After Key Path Optimization

Indicator	Before Optimization	After Optimization	Indicator	Before Optimization	After Optimization
TE	0.815	0.948	Audit Coverage Rate	82.3	96.8
PTE	0.852	0.976	Risk Identification Accuracy Rate	85.6	94.2
SE	0.957	0.971	Audit Cycle Reduction Rate	32.1	48.5

Table 4 selects 10 pilot big data audit projects, optimizes the key paths based on the results in Table 3 (strengthening the data analysis module, improving the data cleaning module, and optimizing the data collection module), and compares the efficiency changes before and after optimization. The results show that after optimization, the comprehensive technical efficiency increases by 16.3%, the pure technical efficiency increases by 14.6%, and among the core output indicators, the audit cycle reduction rate increases the most significantly (51.1%), while the audit coverage rate and risk identification accuracy rate increase by 17.6% and 10.0% respectively, verifying the effectiveness of the key paths. At the same time, the scale efficiency increases by a small margin (1.5%), which further indicates that the core driving force for audit efficiency improvement is the optimization of technology application rather than scale expansion.

4. CONCLUSIONS

Through in-depth research on the application of big data technology in financial auditing, this study constructs an input-output indicator system and uses the DEA-BCC model to measure and compare the audit efficiency before and after the application of big data technology. The research results show that big data technology can significantly improve the efficiency of financial auditing, mainly through optimizing the efficiency of technology application rather than relying on scale expansion. The data analysis module, data cleaning module, and data collection module are the key technology modules for improving audit efficiency, among which the data analysis module has the most significant effect on improving audit efficiency. By optimizing these key paths, the audit efficiency of the pilot projects has been significantly improved, verifying the effectiveness of the key paths.

The research findings of this study provide theoretical support and practical guidance for the digital transformation of audit institutions. Audit institutions should attach importance to the application of big data technology, optimize resource allocation, strengthen the training of compound talents, and improve the automation and intelligence level of data collection, cleaning, and analysis. At the same time, they should further improve the data processing process and enhance algorithm accuracy to maximize audit efficiency. Future research can further explore the integrated application of big data technology and other emerging technologies to further improve the efficiency and quality of financial auditing.

REFERENCES

- [1] Sun Y, Li J, Lu M, et al. Study of the impact of the big data era on accounting and auditing [J]. arXiv preprint arXiv:2403.07180, 2024.
- [2] Abdelwahed A S, Abu-Musa A A E S, Badawy H A E S, et al. Investigating the impact of adopting big data and data analytics on enhancing audit quality [J]. Journal of Financial Reporting and Accounting, 2025, 23(2): 472-495.

- [3] Gepp A, Linnenluecke M K, O'neill T J, et al. Big data techniques in auditing research and practice: Current trends and future opportunities [J]. *Journal of Accounting Literature*, 2018, 40(1): 102-115.
- [4] Xing Z, Zhu L, Lijun Z. A study on the application of the technology of big data and artificial intelligence to audit [C]//2020 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2020: 797-800.
- [5] Alles M, Gray G L. Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors [J]. *International Journal of Accounting Information Systems*, 2016, 22: 44-59.
- [6] Asmild M, Paradi J C, Pastor J T. Centralized resource allocation BCC models [J]. *Omega*, 2009, 37(1): 40-49.