

Is AI Usage Relevant to Audit Effectiveness: Correlation Findings from the Big Four Firms

Yihan Zhao

College of Arts and Science, Boston University, Boston, USA
yihanzz@bu.edu

Abstract. The article aims to quantify and describe the relationship between the penetration of Artificial Intelligence (AI) in the auditing industry and audit effectiveness, with a focus on examining the heterogeneity of “whether AI is adopted” among companies in different years. The article is based on panel data from the Big Four accounting firms (2020–2025), with audit effectiveness scores as the dependent variable, AI adoption (binary) as the core independent variable, controlling for the impact of compliance violations and total revenue, and incorporating year effects. Methodologically, a robust OLS model incorporating year effects was first constructed. Subsequently, a firm-fixed effects model was also established. Additionally, an “AI $\times$ year” interaction term was included in the FFE framework, with marginal effects and annual mean plots used to characterize the temporal dimension, and standard errors were robustly adjusted. The results indicate a positive overall association, with periodic fluctuations over time. In conclusion, the paper defines the findings as evidence of a correlation, suggesting that AI's explicit impact on audit effectiveness is time- and context-dependent.

Keywords: Econometric, Artificial Intelligence, Audit

1. Introduction

Auditing is the work of independent institutions or personnel who examine the financial and operational status of audited entities by established standards. This work typically involves issuing an economic report after assessing the authenticity, legality, and efficiency of the entity. Its core function is to provide reliable information to stakeholders through professional review, thereby ensuring the proper conduct of economic activities. At the same time, Artificial Intelligence (AI) has developed at a stupendously rapid pace in the 21st century, penetrating various industries, including finance, healthcare, manufacturing and others. Remarkable technical progress leads to a higher requirement for efficiency in data-related work. Given its reliance on data analysis, auditing is naturally one of the tasks that highly demands an intelligent working pattern.

Recently, some related work showed how AI has penetrated the auditing industry. And most of them present the positive relationship between them. Patel et al. has shown that AI can use machine learning algorithms to identify abnormal transactions and patterns, thereby improving the efficiency and accuracy of the audit process [1]. Also, empirical research has found that the application of AI is positively correlated with audit quality and can reduce audit costs and restatement risk. For example,

Fedyk showed that a one-standard-deviation increase in AI investment can reduce the restatement rate by 5% [2]. However, previous studies have some limitations. For example, a study shows that deep learning anomaly detection models can help AI identify potential risks in audit logs, thereby improving audit efficiency [3]. However, this research primarily relies on simulated or small-scale datasets and lacks empirical testing in real-world business settings. Furthermore, many studies have analyzed the application of AI technology in the audit field from a theoretical perspective, such as Natural Language Processing (NLP) and machine learning [4]. However, their research was overly theoretical and did not use quantitative methods to measure its actual impact on audit quality. Additionally, most studies rely on cross-sectional data from the same year. While Fedyk et al. covered 36 companies [2], the study itself ignored the possibility of company endogeneity and time variability. It suggests that further research should be conducted to clarify the dynamic relationship between the two, as existing research generally focuses on short-term effects.

According to the limitations of existing research in terms of data dimensions, empirical methods, and research subjects, this paper aims to fill the aforementioned gaps by providing a more representative and empirical research perspective. Specifically, this paper constructs a panel data set based on Kaggle's data from the Big 4 accounting firms worldwide between 2020 and 2025 [5]. This paper not only employs traditional OLS models to establish baseline regression results but also utilizes fixed effects models to control for firm heterogeneity, thereby enhancing the credibility of causal identification. By comparing performance differences between companies that have adopted AI and those that have not, and observing the impact of AI technology over time through annual-level changes, this paper seeks to map out the influence pathways of AI technology from pilot applications to full-scale deployment.

2. Method

2.1. Dataset preparation

This study draws on a panel dataset covering the Big Four accounting firms including PwC, Deloitte, KPMG and Ernst & Young over the years 2020 to 2025 from Kaggle [5]. The dataset records multiple indicators related to audit activities, risk assessment and firm-level performance. Key variables include the number of audit engagements, high-risk cases, compliance violations and fraud cases. Importantly, it also indicates whether AI tools were used in the audit process. The primary outcome of interest is a numerical measure of audit effectiveness, which serves as a proxy for audit quality.

The dataset contains 12 variables and 216 firm-year observations. Several preprocessing steps were applied to prepare the data for analysis. First, categorical fields such as AI adoption were converted into binary indicators (e.g., `ai_used` equals 1 if AI was used for auditing, and 0 otherwise). Then, numerical variables were reviewed for outliers and missing values. Observations with extreme values in variables like `total_revenue_impact` and `employee_workload` were winsorized at the 1st and 99th percentiles. No major missing data issues were found, and the panel is strongly balanced across firms and years. Industry classification and year are kept as identifiers for later use in fixed effects estimation.

2.2. Regression model

To quantitatively examine the relationship between key factors and audit effectiveness, this study employs two widely used econometric models: the Ordinary Least Squares (OLS) model and the

Fixed Effects (FE) panel data model.

The OLS regression model is a fundamental econometric method for estimating the linear relationship between independent variables and a dependent variable [6]. The model is:

$$\text{Audit effectiveness score} = \beta_0 + \beta_1 \text{AI used} + \beta_2 \text{Compliance violations} + \beta_3 \text{Total revenue impact} + \beta_4 \text{year2021} + \beta_5 \text{year2022} + \beta_6 \text{year2023} + \beta_7 \text{year2024} + \beta_8 \text{year2025} + \varepsilon \quad (1)$$

where, audit effectiveness score is dependent variable and AI used is the main independent variable with Compliance violations, Total revenue impact, and years being the control variables. In addition, β_0 is intercept and β_1 to β_8 are the coefficients to be estimated. Also, ε means the error term.

OLS is simple to interpret and widely used in empirical research, but it relies on strict assumptions such as no omitted variable bias, homoskedasticity, and exogeneity. Violations of these assumptions may lead to biased or inconsistent estimates.

Given that the used dataset is structured as a panel—observations across firms and years—OLS may not fully account for unobserved heterogeneity. For example, firm-specific characteristics such as internal management structure or organizational culture might remain constant over time but still influence audit outcomes. To address this, the article extends analysis using a FE model, which controls for time-invariant characteristics by effectively "demeaning" the data within each entity.

The FE model is written as:

$$\text{Audit effectiveness score}_{it} = \alpha_i + \beta \text{AI used}_{it} + \delta_1 \text{Compliance violations}_{it} + \delta_2 \text{Total revenue impact}_{it} + \tau_t + \varepsilon_{it} \quad (2)$$

where, Audit Effectiveness score is the dependent variable. AI used is the main independent variable of interest. Compliance violations and Total revenue impact are control variables, and τ_t denotes year fixed effects (one year omitted as the baseline). α_i is the firm fixed effect, which acts as a firm-specific intercept capturing all time-invariant characteristics of firm i . The parameters β , δ_1 , δ_2 are the coefficients to be estimated, and ε_{it} is the idiosyncratic error term. With this specification, β is interpreted as the within-firm association between adopting AI and the audit effectiveness score, holding constant firm invariants and common year shocks.

By focusing only on within-firm variation over time, FE reduces bias caused by unobserved fixed traits. However, one limitation of FE is that it cannot estimate the effect of variables that do not vary over time within entities. In this study, OLS serves as a benchmark for assessing the general linear relationship, while the FE model helps to obtain more robust causal interpretations by accounting for heterogeneity across firms. Together, these models allow to triangulate the results and evaluate the robustness of the association between audit practices and observed outcomes.

At the same time, the article tests the year heterogeneity extending from the normal fixed effect. The formula for the model is:

$$\text{Audit effectiveness score}_{it} = \alpha_i + \beta \text{AI used}_{it} + \sum_{t \neq 2020} \gamma_t (\text{AI used}_{it} \times D_t) + \delta_1 \text{Compliance violations}_{it} + \delta_2 \text{Total revenue impact}_{it} + \tau_t + \varepsilon_{it} \quad (3)$$

where, based on model 2, $AI_{usedit} \times Dt$ was added as the interaction term between AI and year. This allows the impact of AI to vary by year, in order to report the annual marginal effects for each year.

3. Results and discussion

3.1. Descriptive analysis

3.1.1. Statistics summary

Table 1 presents the descriptive statistics for the key variables in this study. The sample covers audit engagement records from the Big Four accounting firms (KPMG, PwC, EY, and Deloitte) globally from 2020 to 2025 [5]. Approximately 45% of the observations in the sample reported the use of AI in the audit process. In addition, the sample shows an average Audit Effectiveness Score of 7.49 (out of 10) and a standard deviation of 1.52. In terms of audit volume, the Big Four firms conduct an average of approximately 2,784 audits annually, with around 278 classified as high-risk cases. The risk rate averaged 10.3% and showed significant variations across different companies and years.

Table 1. Statistics summary

Variable	obs	Mean	Std.dev.	Min	Max
Total audit engagements	100	2784.52	1281.863	603	4946
High risk cases	100	277.73	135.7352	51	500
Compliance violations	100	105.48	55.3701	10	200
Total revenue impact	100	272.539	139.1514	33.46	497.06
AI used	100	.45	.5	0	1

The adoption rate of AI at 45% suggests that, within the 2020–2025 period, AI technology has not yet achieved full penetration in the Big Four auditing firms. The variation in audit effectiveness scores indicates that firm-level and contextual factors likely play an important role, consistent with prior literature that highlights heterogeneous productivity effects of AI adoption [2]. However, the relatively small proportion of high-risk audits implies that extreme cases may have limited influence on the mean effectiveness score.

3.1.2. Two-sample t-test on AI usage and audit effectiveness

In order to examine the relationship between AI usage and audit effectiveness, the article conducted a preliminary two-sample t-test. The results in Table 2 show that firms using AI scored 0.19 points higher on average. However, such data was not statistically significant with p value equals to 0.532.

Table 2. Two-sample t-test

Group	obs	mean
Not use AI	55	7.403636
Use AI	45	7.595556

$\Pr(|T| > |t|) = 0.5316$

While the descriptive comparison hints at a positive relationship between AI usage and effectiveness of audit, the lack of statistical significance indicates that other variables must be controlled to isolate the AI effect. This justifies the use of multivariate regression models in subsequent sections.

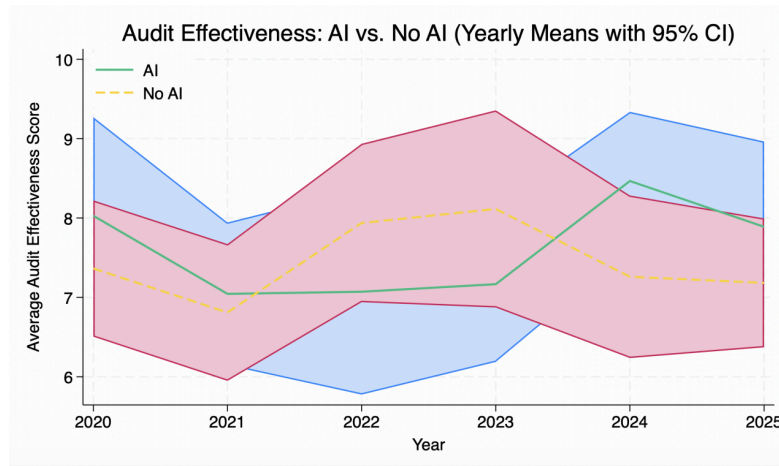


Figure 1. Audit effectiveness: annual mean values for the AI group and the non-AI group (95%CI)
(picture credit : original)

Figure 1 shows the average audit effectiveness scores for audit samples using AI and those not using AI within their 95% confidence intervals for each year between 2020 and 2025.

From the annual trends, the mean difference between the two groups was small in 2020–2021, with a synchronized decline in 2022–2023. By 2024, the average score for the AI-adopting group shows a significant rebound, and by 2025, the gap between the two groups narrows but remains higher for the AI group. Although this figure is used solely for descriptive comparison and does not control firm-specific characteristics or other covariates, the overall trend is clearly non-monotonic. This also underscores the need for other regression models.

3.2. Regression results

3.2.1. OLS

Table 3. OLS model outcome

variables	Audit effectiveness score
Ai used	0.2838 (0.92)
Compliance violations	0.0015 (0.53)
Total revenue impact	0.0008 (0.62)
year	
2021	-0.7400 (-1.46)
2022	-0.077 (-0.14)
2023	0.0280 (0.05)
2024	0.0850 (0.16)
2025	-0.1552 (-0.29)

$R^2=0.0477$, $F(8,91)=0.72$ $p<0.05$

Table 3 shows the OLS model, with the model formula as follows formula 1 in method.

The core explanatory variable is whether AI is adopted. In addition, `compliance_violation`, `total_revenue_impact`, and year dummy variables (with 2020 as the base year) serve as control variables. The sample size is $N=100$. Furthermore, the key data metrics are $F(8,91)=0.72$, $R^2=0.0477$, and Root MSE=1.5432. The coefficients and t-values in parentheses for each variable are presented in Table 3.

In terms of magnitude, the coefficient of `ai_used` (0.2838) indicates that, after controlling for compliance, revenue impact, and year differences, observations with “AI adoption” have an audit effectiveness score approximately 0.28 points higher (on a scale of 1–10). The two-tailed p-value calculated from $t=0.92$ and $df=91$ is approximately 0.36. The p-value of the results shows bigger than 0.05, which means that it is statistically insignificant. Meanwhile, the coefficient estimates for the control variables are positive (`compliance_violations` = 0.0015; `total_revenue_impact` = 0.0008). The year term shows slight fluctuations relative to 2020. This pattern of “positive direction but effect size influenced by context” aligns with the main findings of previous studies. Industry and academic evidence generally suggest that AI currently has a more direct impact on process efficiency in areas such as anomaly detection, sample coverage, and evidence collection, while its explicit improvement in quantifiable “audit quality” is context-dependent and subject to time lags [7]. Additionally, reviews of ED-500 emphasize that technological interventions have increased the

complexity and governance requirements of evidence evaluation and professional skepticism, but the presentation of quality metrics must be aligned with evolving standards and organizational practices [8]. Based on this, the next step is to use firm-level fixed effects and heterogeneous grouping to conduct a more detailed analysis of firm differences.

3.2.2. Fixed effect model

Table 4. Fixed effect model outcome

variables	Audit effectiveness score
Ai used	0.3341 (1.05)
Compliance violations	0.0017 (0.59)
Total revenue impact	0.0008 (0.60)
year	
2021	-0.8110 (-1.61)
2022	-0.1866 (-0.33)
2023	0.0078 (-0.01)
2024	0.0728 (0.13)
2025	-0.2057 (-0.37)

$R^2=0.0559$, $F(8,88)=0.87$, $p<0.05$

Table 4 (FE) builds upon the OLS specification in the table by incorporating firm-level fixed effects to control for unobserved differences across firms that do not vary over time with the model2 in method. Other control variables, including compliance violations, total revenue, and year dummy variables (with 2020 as the base year), remain unchanged. The sample size is $N=100$. The key results are $F(8,88) = 0.87$, $R^2 = 0.0559$, and Root MSE = 1.5626. The coefficient estimates for each variable (with t-values in parentheses) are shown in Table 4.

As shown in Table 4, the coefficient estimates for ai_used is 0.334. This implies that, after controlling for compliance, revenue impact, and year differences within the same firm, the use of AI is associated with an approximately 0.33-point increase in audit effectiveness scores (on a scale of 1–10). Compared to the OLS estimate of 0.2838, the FE coefficient is slightly larger. This indicates that the positive association between ai_used and effectiveness within firms remains after controlling firm-specific differences, and is not diluted by fixed effects. In summary, the conclusions of FE and OLS are consistent. Both indicate that AI functions more like a general-purpose technology, with its actual performance dependent on complementary investments such as process

restructuring, skill retraining, and data governance. Therefore, the statistical effects observed at different stages within an organization may vary [9].

3.2.3. Year heterogeneity fixed effect

Next, the study tested for year heterogeneity with model 3 in method. It examines the impact of whether AI usage changes over time on audit effectiveness within a firm-fixed effects framework. By including the interaction term “AI usage $\times$ year,” the marginal effect of AI on audit effectiveness scores can be estimated separately by year, holding all other conditions constant. The model formula follows formula 3 in method. By presenting the marginal effect diagram of annual heterogeneity (Figure 2), the article reflects the average change in each year corresponding to “adopting AI” relative to “not adopting AI” within the same firm, after controlling for compliance violations and total revenue impact and incorporating the year effect.

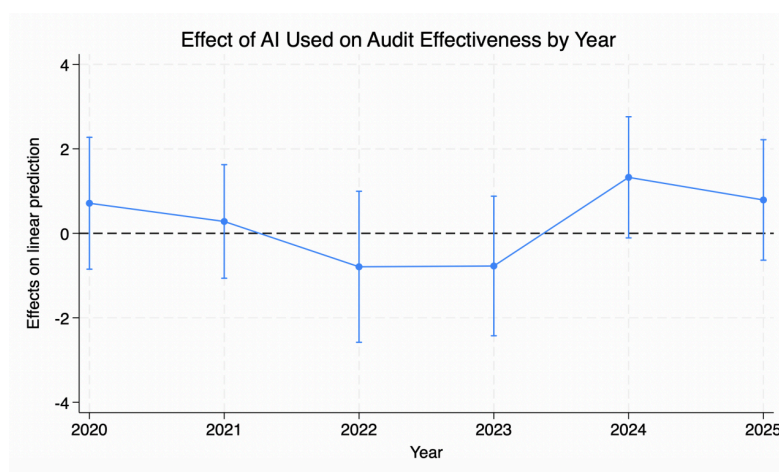


Figure 2. Marginal effect of AI used on audit effectiveness by year (picture credit : original)

The numerical values follow a time trajectory of positive—negative—positive: positive in 2020–2021, negative in 2022–2023, clearly positive in 2024, and still positive in 2025. In terms of direction and magnitude, the decline in 2022–2023 and the rebound in 2024 constitute the primary annual heterogeneity characteristics. This aligns with the “2022–2023 trough, 2024 rebound, and 2025 sustained positive trend” pattern described in the previous descriptive chart and the FE main regression. It also confirms that the association between AI and audit effectiveness changes over time, resembling a learning cycle of adoption and absorption. Mid-term effects may be influenced by factors such as implementation adjustments and process revisions [10]. Subsequently, as tools and organizational support become fully established, the positive effects become more evident. This annual heterogeneity is mutually corroborated by the descriptive trends and fixed-effects main results discussed earlier. It also serves as one of the pieces of evidence supporting the paper’s overall argument that “AI penetration exhibits time-path dependence/phased release.”

4. Conclusion

This paper examines the correlation between AI usage and audit effectiveness by dealing with data from the Big Four accounting firms. Both the benchmark OLS and fixed effects results show a positive direction, with a trough in 2022–2023, a rebound in 2024, and a continued positive trend in 2025 in the “year-by-year heterogeneity” analysis. The joint test for annual differences did not

reveal significant variations, so the article positions this study as correlational evidence rather than causal inference. Its direction aligns with the industry-academic consensus that “AI is more likely to indirectly impact quality through expanded coverage, anomaly detection, and improved efficiency in evidence collection”. This result is also corroborated by external evidence from larger samples. Research from multiple large audit firms has found that AI investment is associated with higher audit quality, lower audit rates, and long-term human replacement effects, but visible changes in labor and performance exhibit time lags. In other words, technological gains are not released linearly, and their manifestation is constrained by organizational and temporal pathways. The “mid-term decline—late-term rebound” annual trajectory observed in the Big Four sample aligns with this logic.

The main limitation of this paper is that it only discusses data from the Big Four accounting firms. The sample size is small, the time dimension is short, and the AI variable is binary data, which limits statistical power and results in wide confidence intervals for the coefficients. In future developments, the following points could be explored. First, expand the data to cover more firms or longer time periods. Under such conditions, more cyclical causal changes could be observed. Second, expand research into more fields. The impact of AI is not limited to the accounting industry. Comparisons with other industries could be made to conduct a horizontal comparative analysis and derive the broader relationship between AI and industry impact.

References

- [1] Patel, R., Khan, F., Silva, B., & Shaturaev, J. (2023). Unleashing the potential of artificial intelligence in auditing: A comprehensive exploration of its multifaceted impact. Munich Personal RePEc Archive. <https://mpra.ub.uni-muenchen.de/119616/>
- [2] Fedyk, A., Hodson, J., Khimich, N., & others. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies*, 27, 938–985.
- [3] Hemati, H., Schreyer, M., & Borth, D. (2022). Continual learning for unsupervised anomaly detection in continuous auditing of financial accounting data. *arXiv*. <https://arxiv.org/abs/2203.16705>
- [4] Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory*, 36(4), 1–27.
- [5] Soundankar, A. (2025). Big 4 financial risk insights (2020–2025) (Version 1). Kaggle. Retrieved August 19, 2025, from <https://www.kaggle.com/datasets/atharvasoundankar/big-4-financial-risk-insights-2020-2025>
- [6] Pavelescu, F. M. (2010). An analysis model for the disturbances generated by collinearity in the context of the OLS method. *Romanian Journal of Economic Forecasting*, 2, 245–264.
- [7] Almufadda, G., & Almezeini, N. A. (2022). Artificial intelligence applications in the auditing profession: A literature review. *Journal of Emerging Technologies in Accounting*, 19(2), 29–42.
- [8] Barr-Pulliam, D., Calvin, C., Eulerich, M., & Maghakyan, A. (2023). Audit evidence, technology and judgement: A review of the literature in response to ED-500. *Journal of International Financial Management & Accounting*, 35(1), 36–67.
- [9] Law, K. K. F., & Shen, M. (2024). How does artificial intelligence shape audit firms? *Management Science*, 71(5), 3641–3666.
- [10] Leocádio, D., Malheiro, L., & Reis, J. (2024). Artificial intelligence in auditing: A conceptual framework for auditing practices. *Administrative Sciences*, 14(6), 238.