

An Audit Risk Early Warning Model Based on Machine Learning and ESG Rating Divergence

Yuhan Ren¹, Wanfeng Zhang^{2*}

¹*School of Economics and Management, Nanchang University, Nanchang, China*

²*School of Management Science and Engineering, Tianjin University of Finance and Economics, Tianjin, China*

**Corresponding Author. Email: WanfengZhang@stu.tjufe.edu.cn*

Abstract. In recent years, the capital market has paid increasing attention to ESG, and many scholars have examined the impact of ESG ratings on market efficiency, audit risk, and audit opinions. However, there remains a relative lack of research on predicting audit risk using ESG rating divergence. This paper focuses on the divergence in ESG ratings of A-share-listed companies and establishes an intelligent early-warning model of audit risk using random forests, XGBoost, and linear regression. The empirical results show that, after controlling for basic financial variables, ESG rating divergence continues to have a significant incremental predictive effect on the identification of audit risks. Combined with the idea of cost-sensitive learning. After adjusting the model's risk warning threshold to 5%, the model successfully identifies the risk of material misstatement with strong concealment at the tail of the data. To a certain extent, this paper enriches the interdisciplinary literature on ESG and auditing and also provides a practical reference for using machine learning for early risk warning.

Keywords: machine learning, ESG rating divergence, rating agency disagreement

1. Introduction

Since China announced its "dual carbon" target in 2020, the ESG concept has attracted extensive attention in capital markets. Rating agencies have different priorities in allocating index weights, measuring qualitative standards, and capturing underlying data, which directly leads to obvious differences in ESG scores of the same enterprise [1]. Existing literature primarily attributes these divergences to differences in rating systems, insufficient disclosure of corporate information, and regional preferences in the rating process [2]. It is worth noting that ESG rating divergences are often mixed with market noise, further increasing audit risk premiums. Some enterprises regard the active fulfillment of social responsibility as a self-interested tool to cover up earnings management and avoid external supervision [3]. The greater the divergence in ESG ratings across enterprises, the more complex the internal and external information environments they face. If auditors fail to identify these non-financial false signals in time, they risk severe misjudgments that can ultimately lead to audit failures [4]. Therefore, predicting enterprise audit risk based on ESG rating divergence in the early stage is particularly important. According to the existing literature, ESG rating

divergence creates market information noise, which raises audit premiums, making the overall audit process more difficult and time-consuming, and ultimately increasing audit costs [5].

When faced with enterprises with large ESG rating divergences, the possibility of auditors issuing non-standard audit opinions will also increase. However, research gaps remain regarding the role of ESG rating divergences in audit risk prediction. This paper proposes to develop a multidimensional fraud identification framework that integrates finance, business, and corporate governance and introduces big data algorithms to enable early identification. Using random forests, XGBoost, linear regression, and cost-sensitive learning theory, the incremental predictive value of ESG rating divergence for audit risk is explored by analyzing its variation coefficient. Theoretically, this paper tests the economic consequences of ESG rating divergence from the audit risk perspective. In practice, it responds to the suggestion of lowering detection thresholds to optimize fraud governance and develops a dynamic early-warning model.

2. Theoretical analysis and hypothesis development

In practice, different rating agencies hold different views on enterprises' non-financial information, which is most directly reflected in the divergence of ESG ratings. In practice, each rating agency emphasizes different evaluation criteria, and market standards are difficult to unify, leading directly to divergences in ESG scoring for the same enterprise [6]. This rating divergence is relatively prominent in China's capital market, especially between state-owned and non-state-owned enterprises [7]. High rating divergence is often accompanied by more serious information asymmetry, reflecting the noise effect of chaotic market information, which increases the cost of information for investors and audit costs and risks [8]. Looking back at the domestic capital market, this deviation in rating is actually mixed with more complex factors. The self-interest tool hypothesis suggests that the management of some companies appears to actively fulfill social responsibility, but in fact, they are likely only using it as a means to whitewash corporate images, and it is likely only a means to package themselves [9]. They deliberately build a positive social image, mainly to build a moral shield for the company and avoid external regulatory oversight. When the divergence in an enterprise's ESG rating is high, it suggests that some governance defects are deliberately covered up. The management's behavior of covering up the problem for self-interest invisibly raises the agency cost of the whole enterprise on the practical level.

If a company's financial and non-financial indicators deviate significantly from the growth rate, this is usually a dangerous signal of potential fraud [10]. Audit risk can be decomposed into material misstatement risk and detection risk. In earlier studies, the modified Jones model was used to calculate the degree of earnings management and directly measure the risk of material misstatement, while audit opinions can represent detection risk. From the auditor's perspective, under the principle of prudence, auditors tend to be more conservative about market noise and the possible opacity of accounting information in enterprises with particularly different ESG ratings. They usually prefer to issue nonstandard audit opinions directly to these enterprises to improve audit risk. ESG divergence increases audit premiums and audit costs, suggesting that asymmetries in market information led to greater earnings management or a higher risk of material misstatement.

H1: Holding other conditions constant, ESG rating divergence is significantly positively correlated with enterprise audit risk, thereby improving the prediction accuracy of the audit risk early warning model.

3. Research design

This paper selects Chinese A-share listed companies from 2019 to 2023 as the research sample. Financial data, including audit opinions, audit firms, total assets, and net cash flow from operating activities, are obtained from the CSMAR and RESSET databases, while ESG rating data are derived from Huazheng, SynTao Green Finance, and FTSE Russell, which are included in the Wind database. After removing missing values, 13054 effective observation samples were obtained. According to the 8:2 ratio, it is randomly split into a training set and a test set for model training and evaluation. There are about 10443 training sets and 2611 test sets.

This paper follows the classical audit and governance theory in selecting variables. In view of the audit risk associated with the explained variable, in line with existing research, non-standard audit opinions are used to represent the audit inspection risk. If the auditor issued an unqualified opinion with an emphasis of matter, a qualified opinion, an adverse opinion, or a disclaimer of opinion, they are uniformly recorded as 1 in the model. If the enterprise receives a standard unqualified opinion, the observation is coded as 0. In view of the divergence in ESG ratings of core explanatory variables across Huazheng, Wind, SynTao Green Finance, and other institutions, this paper adopts a standardized mapping method to convert each institution's rating to a 1-9 scale. Using the coefficient of variation to measure the rating divergence within the same enterprise and year can eliminate the dimensional impact of differences in average scores across institutions. This paper selects control variables from the perspectives of financial characteristics and corporate governance characteristics, and the specific definitions are shown in Table 1.

Table 1. Variable definition

Variable type	Variable name	Symbol	Variable meaning
Explained variable	Audit Risk	MAO	If an unqualified opinion, reservation, denial, or inability to express an opinion with emphasis is issued, it is taken as 1, and the standard unqualified opinion is taken as 0.
Explanatory variable	ESG Rating Divergence	ESG_Div	To measure the rating divergence level of the same enterprise in the same year, the coefficient of variation after standardized scoring by rating agencies is used to calculate.
	Enterprise Size	Size	The natural logarithm of the total assets of the enterprise at the end of the period.
	Asset Liability Ratio	Lev	Total liabilities/total assets, which measures the incentive for fraud brought about by debt repayment pressure.
Controlled variable	Profitability	Roa	The ratio of net profit to total assets at the end of the period reflects the impact of operating performance on audit opinions.
	Cash Flow Ratio	Cashflow	Net cash flow from operating activities/total assets, which measures the authenticity of profits
	Dual Role	Dual	If the chairman and the general manager are the same person, take 1. Otherwise, take 0.

In this paper, the random forest algorithm is used to build the model. Considering that the proportion of samples with non-standard opinions in the whole market is very low, only about 2.1%, the data show a serious imbalance. Therefore, this paper aims to improve the model's predictive performance by adjusting the decision boundary. If the default 0.5 threshold in machine learning is used, the model is prone to bias towards the majority class to maximize overall accuracy. That is, even if the model classifies all samples as risk-free, it can still achieve an overall accuracy of 97.9%.

However, this simple pursuit of high accuracy will lead to the direct omission of a very small number of genuinely illegal enterprises. Combined with the real audit environment, the cost of miscalculation is actually asymmetric. If enterprises with major misstatements are not identified, that is, type II errors, it often leads to serious audit failures and litigation risks. If a normal enterprise is misjudged as a high risk, that is, a type I error, it usually only increases the cost of auditors' subsequent verification procedures. Given the difference in misjudgment costs, the classification model should be appropriately biased towards the minority class. In the prediction stage, the high-risk judgment threshold is explicitly lowered from 0.5 to 0.05, which not only corrects the defect in the model's tendency to ignore minority samples but also improves the ability to capture hidden illegal enterprises. From a practical perspective, this is also more in line with the practice logic that auditors usually need to focus on preventing tail risks.

4. Analysis of empirical results

To test the traditional linear impact of ESG rating divergence on audit risk, this paper first establishes an OLS regression model as a benchmark. According to Table 2, the regression coefficient of the core independent variable ESG_Div is 0.0221, which is significantly positive at the 1% level with a t-statistic of 7.455, supporting hypothesis H1 in this paper. This indicates that the greater the difference between institutions in the company's ESG rating, the higher the probability of being issued a non-standard audit opinion, thereby verifying the incremental risk of material misstatement associated with the rating difference. The correlation of the coefficient is relatively strong, the standard error is small, the T value is large, the regression error is small, the reliability is high, and the significance is still valid at the 1% level.

Table 2. Test results of OLS model

	Regression coefficient	Standard error	T value	P value	Significance
const	0.374	0.0232	16.091	0.0000	***
ESG_Div	0.0221	0.003	7.455	0.0000	***
Size	-0.0176	0.0011	-15.6812	0.0000	***
Lev	0.0013	0.0001	14.6068	0.0000	***
ROA	-0.0054	0.0002	-24.2713	0.0000	***
Cashflow	0.1695	0.022	7.7134	0.0000	***
Dual	-0.0063	0.0032	-1.9596	0.0501	*

To further explore the nonlinear characteristics and improve the accuracy of early warning, this paper introduces machine learning algorithms and compares the traditional multiple regression, random forest, and XGBoost models at the same stage on the same test set. In addition to the classification index, this paper introduces the mean square error and the determination coefficient to measure comprehensively. From Table 3 and the multi-model ROC comparison chart, it is evident that the random forest model performs best on the test set. Its AUC value reaches its peak, fully demonstrating that the random forest algorithm shows greater robustness and prediction accuracy in the face of unbalanced audit risk data.

Table 3. Comparison of early warning model results

	Accuracy	Precision	Recall	F1-Score	AUC Value	Mean square error	Coefficient of determination
Multiple linear regression	0.9724	0.3846	0.0725	0.122	0.8697	0.0233	0.0934
Random forest	0.9736	0.5	0.0435	0.08	0.858	0.0229	0.1108
XG-Boost	0.9717	0.3529	0.087	0.1395	0.8475	0.0247	0.0415

The three models achieve about 97% accuracy. The comprehensive F-score is relatively high. Risk identification accuracy is strong. The mean square error is small, and the fitting degree is relatively high, which can better predict audit risk. To verify the impact relationship between the two variables, the three models' predictive results are highly interpretable and reliable, demonstrating incremental predictive power for ESG divergence in audit risk. As shown in Figure 1, the AUC curves for the three algorithms are relatively high. This shows that the model distinguishes between high- and low-risk enterprises and that it is reasonable to use this algorithm to process data from the domestic capital market.

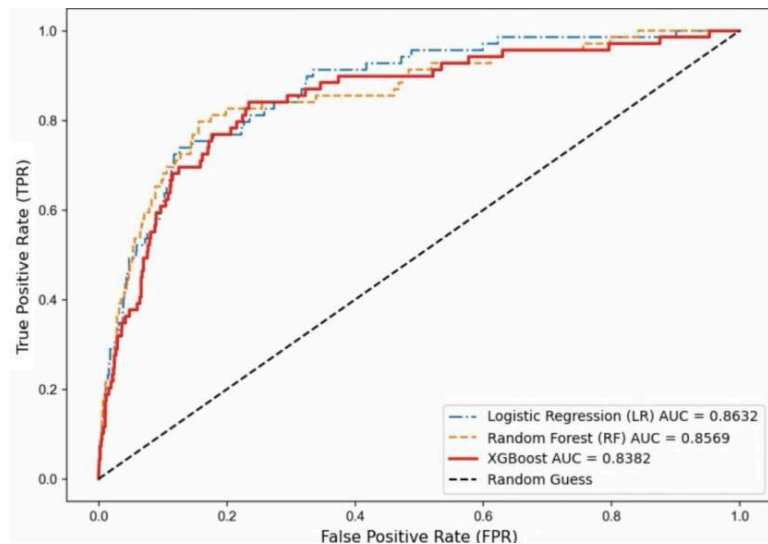


Figure 1. ROC curve comparison (picture credit: original)

According to the SHAP visualization, among the control variables, the asset-liability ratio, which represents the company's financial risk, and the net interest rate on total assets, which represents profitability, play a fundamental role in identifying the risk of material misstatement in the model. As seen in Figure 2, it is noteworthy that ESG_Div, as the core explanatory variable for external non-financial information, occupies a very central position in the characteristic importance matrix, indicating that it provides incremental predictive value beyond financial data.

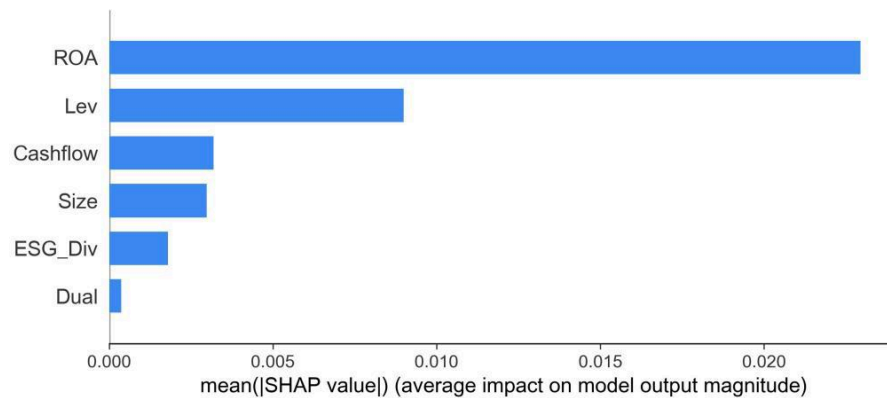


Figure 2. Feature importance analysis (picture credit: original)

In the real audit business and capital market supervision, failing to detect financial fraud or material misstatement in an enterprise may result in huge civil compensation, regulatory penalties, and reputation damage, which is much higher than the cost of misrepresenting a company that is not a problem. This extremely asymmetric cost structure requires that the early warning model not only pursue overall statistical accuracy. Because non-standard audit opinions are very rare across the entire sample, adopting the default 50% classification threshold in machine learning would make the model overly conservative, thereby resulting in a lower recall rate. Therefore, combined with the principle of prudence in audit practice, this paper introduces cost-sensitive learning and greatly reduces the risk identification and early warning threshold of the random forest model to 5%, that is, as long as the model captures the potential misstatement characteristics of the enterprise of 5%, it will sound the alarm of non-standard opinions. The comparison results are shown in Table 4. Although lowering the early warning threshold to 5% has reduced overall accuracy due to false alarms from some healthy companies, the model's ability to capture high-risk and non-standard opinion companies has improved to 0.7101. In practice, by lowering the early-warning threshold, the random forest model enables accounting firms to identify high-risk enterprises with greater sensitivity during large-scale project screening, thereby effectively mitigating the risk of material misstatements. This conclusion not only fully meets the core requirements for maintaining professional skepticism in modern risk-oriented audit, but also demonstrates the value of this early-warning model.

Table 4. Results of dynamic threshold adjustment

	Accuracy	Precision	Recall	F1-Score
Default Threshold (50%)	0.974	0.5714	0.058	0.1053
Strict Warning Threshold (5%)	0.879	0.142	0.7101	0.2367

5. The logical relationship

Combined with the model's characteristic importance and specific classification performance, this paper further examines the internal relationship between ESG rating divergence and the underlying audit risk. First, the perspective on the data-mapping relationship. Combined with the characteristic importance distribution, ESG_Div occupies a higher weight in the prediction model. From the perspective of the data-mapping mechanism, the scoring differences across rating agencies are not simple calculation errors but a substantial positive mapping relationship with the potential risk of

enterprise violations. Specifically, when an enterprise's ESG rating divergence is high, it often indicates a significant increase in the probability of major misstatements in its underlying data.

Second, the perspective of the information environment. The model can use ESG rating divergence to better identify non-standard opinion samples, as high rating divergence often indicates low transparency in non-financial information. In this way, the market will produce complex information noise. These noises will not only distort external investors' judgment of enterprises' actual conditions but also significantly increase the inherent risks for auditors when they take over such enterprises.

Third, the perspective of management motivation. The empirical results of this paper also provide some support for the self-interest tool hypothesis in the field of social responsibility. The model can still maintain a high recall rate at the 5% threshold, indicating that enterprises that are eventually found to be in violation of regulations or to receive non-standard opinions are more likely to engage in greenwashing in advance. Some managers will deliberately whitewash ESG indicators that are relatively easy to quantify to cover up actual earnings management or internal control loopholes within the company. It directly leads to huge differences in the scoring of external rating agencies. Therefore, the high divergence in ESG ratings can serve as an early warning signal of management's opportunistic motives.

Fourthly, the auditor response mechanism. Combined with the early warning model's recall performance, the significant positive correlation between ESG rating differences and non-standard audit opinions essentially reflects the defensive logic of auditors in practice. When auditors implement the early risk assessment procedures, if they find that an enterprise has large rating differences on non-financial information, to avoid follow-up litigation and maintain their own professional reputation, they often raise the level of control risk assessment of the enterprise accordingly, take the initiative to lower the level of audit importance, and Consequently, they are more likely to issue non-standard audit opinions in the final report.

6. Conclusion

This paper empirically investigates the early-warning value of ESG rating divergence for audit risk using a machine learning model. It is found that, after controlling for traditional financial characteristics, ESG rating divergence is effective in identifying non-standard audit opinions and predicting an increase in earnings management. Given the extremely sparse sample of audit violations, this paper dynamically reduces the model's early-warning threshold to 5% and achieves efficient, accurate recall of hidden risks. This result confirms the hypothesis and theoretical logic of this paper, namely, that the divergence in ESG ratings deeply reflects information asymmetry in the capital market and the opacity of enterprises' internal information disclosure. Its essence is that managers use social responsibility as a self-interested tool to cover up defects, thereby aggravating agency conflicts and increasing audit risks. Against the backdrop of the big data era, market information is chaotic, and the quality and reliability of accounting information are insufficient. It is suggested that audit institutions can utilize big data and intelligent algorithms for prediction and analysis, and that differences in non-financial evaluation should be included in the red line for wind control. Regulators should also introduce a machine learning model with dynamic parameters to enhance their ability to crack down on hidden fraud and to provide risk warnings for audit work.

The main contribution of this paper is to examine the economic consequences of ESG rating divergence from an audit risk perspective. A dynamic early warning model with great practical value is constructed, which enables audit institutions and enterprises to predict financial risks, reduce audit workload and control costs, promote transparency of ESG evaluation information in the market,

reduce information asymmetry, and make up for the gap in previous research on ESG rating divergence and prediction analysis based on machine learning. Although the model constructed in this paper shows strong early warning performance on the test set, there is still room for improvement. Due to the limited sample period, this paper uses a random division of pooled cross-sectional data for modeling. In future research, with the further enrichment of ESG disclosure data and the extension of the time span, it can try to use a time series rolling window or extended window to carry out a forward prediction test to further verify the dynamic adaptability and actual early warning accuracy of the model under different macroeconomic fluctuation cycles.

Author contribution

All the authors contributed equally, and their names were listed in alphabetical order.

References

- [1] Gibson, B.R., Krueger, P. and Schmidt, P.S. (2021) ESG Rating Disagreement and Stock Returns. *Financial Analysts Journal*, 77(4), 104-127.
- [2] Berg, F., Kölbl, J.F. and Rigobon, R. (2022) Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, 26(6), 1315-1344.
- [3] Stratopoulos, T.C., Wang, V.X. and Ye, H. (2022) Use of Corporate Disclosures to Identify the Stage of Blockchain Adoption. *Accounting Horizons*, 36(1), 197-220.
- [4] Su, K. and Zhou, Z. (2023) Does Corporate Social Responsibility Reduce Crash Risk of Stock Price: Evidence from China. *Chinese Management Studies*, 17(2), 251-273.
- [5] Zhou, Z., Ding, X. and San, Z. (2023) ESG Rating Divergence and Audit Risk Premium. *Auditing Research*, (6), 72-83.
- [6] Yao, J. and Song, Y. (2026) Corporate ESG Rating Divergence and Auditor Response Strategies. *Communication of Finance and Accounting*, (1), 55-59.
- [7] Li, P. and Wang, J. (2025) The Impact of ESG Rating Divergence on Audit Opinion Decisions. *Statistics & Decision*, 41(8), 177-182.
- [8] Zhou, W., Wang, S., Xu, Y. and Shen, F. (2024) Machine Learning and Accounting Fraud Governance: A Prediction Perspective Based on Non-Selected Factors. *The Journal of World Economy*, 47(11), 116-149.
- [9] Ma, W. and Yu, B. (2023) Corporate Ownership and Divergence Between Chinese and Foreign ESG Ratings. *Journal of Finance and Economics*, 49(6), 124-136.
- [10] Le, T.T., Tran, P.Q., Lam, N.P., Tra, M.N.L. and Uyen, P.H.P. (2024) Corporate Social Responsibility, Green Innovation, Environmental Strategy and Corporate Sustainable Development. *Operations Management Research*, 17(1), 114-134.