# Application of Multivariate Logistic Model in Prospective Identification of Initial Public Offering Risk from the Perspective of Investment Banking

Xingcheng Kong

Lee Shau Kee School of Business and Administration, Hong Kong Metropolitan University, Hong Kong, China
1342319@live.hkmu.edu.hk

**Abstract.** This paper explores Initial Public Offering (IPO) risk identification from the perspective of investment banking, focusing on the misreporting of R&D expenses among companies listed on the Science and Technology Innovation Board between 2019 and 2024. Firms penalized or rejected due to false Research and Development (R&D) disclosures were selected as the experimental group, while companies successfully listed without disputes served as the control group. A logistic regression model incorporating six core variables—such as project anomaly index and R&D personnel salary dispersion—was constructed. Parameters were estimated using the maximum likelihood method, with Lasso regression applied to refine variable selection and improve model clarity. The analysis shows that project anomaly index and R&D salary dispersion significantly raise the likelihood of misreporting, whereas auditor industry expertise reduces this risk. The model achieves strong predictive performance, with an AUC of 0.89 and an overall accuracy of 85%. The findings support a forward-looking framework for investment banks to assess IPO risks and strengthen audit quality.

**Keywords:** IPO risk; Multivariate logistic model; Prospective identification; Investment banking; Science and technology innovation board.

## 1 Introduction

The market of IPO of the companies that are already listed in Science and Technology Innovation Board (STAR Market) is highly volatile with more scrutiny being seen on corporate disclosures. Discussing those companies where technological innovation tends to be stressed by them, it becomes important that Research and Development (R&D) expenses may be properly evaluated by investors and underwriters. The false reporting of R&D expenditure weakens the market confidence, invalidates the valuation and places the capacity of investment banks to make far-reaching due diligence under pressure [1]. This study deals with such issues since its aim is to consider how it is possible to reveal a possible misrepresentation in IPO filings based on quantitative approaches.


A. J. Moshayedi (ed.), *Proceedings of the 2025 International Conference on Hybrid Commerce, Human Capital, and Economic Dynamics (ICHCH 2025)*, Advances in Economics, Business and Management Research 374,
https://doi.org/10.2991/978-2-38476-585-0_57

Science and Technology Innovation Board will access the enterprises strongly with the core technology and innovation capability by strict listing criteria and aggravate the area of science and technology innovation industry in China [2]. The financial fraud and information disclosure of some IPO companies in science and technology innovation board has recently being punished and the problem of defects in internal control has a large extent in this, the phenomenon of false reporting on R&D expenses is more serious which harms the interests of the investors and lowers the principles of the market [3]. The motivation to select science and technology innovation board to be the object of study is because; it targets innovation enterprise of science and technology whose core competitiveness is in research and development activities [4]. Being the experimental field of capital market reform, the supervision and audit standard of science and technology innovation board are prospective and leading; The high attention and information disclosure of markets in this platform also offer a good data support to the research. The analysis of the false reports of R&D expenditures of the IPO companies in the science and technology innovation board is very important so as to enhance the quality of the supervision system and quality of the audit.

This study addresses the limitations of traditional risk assessment methods which tend to be subject to the qualitative opinion and are not able to pinpoint the latent patterns of financial anomalies in the IPO filings. In order to conquer these deficiencies, a multivariate logistic model is formulated based on six determinants specifically chosen which include project anomaly index, R&D personnel salary dispersion, hyper growth rate of R&D costs as well as the level of material consumption. These variables encompass the important dimensions of R&D misreporting which traditional models might be blind too. The data-driven methods provided by the model can properly classify the level of risk so that investment banks can detect the possibility of fraud at the early stages and spend much more efficiently. In addition to enhancing the quality of audit, the results would offer a real-world guidance to the regulators in additional to improving supervision, market transparency, and protecting the interest of investors.

## 2 Research on IPO Risk Identification

### 2.1 The Core Issues and Progress of Existing Research

Studies of IPO risk identification have also discussed different motive as to why there is misreporting of R&D expenses such as to access financing benefits in capital markets, wish to comply with policy or to improve the corporate image. According to several studies, the most typical patterns of misreporting include fictious projects in R&D, the overstatement of staff numbers or staff expenses, the improper statement of expenditure [5]. The studies provide the contribution to the comprehension of the mechanisms of misreporting, being rather qualitative with little quantitative versions adapted to industry specificities.

In order to cope with the fraud detection some scholars have adopted statistical models such as logistic regression, and Probit model, which implies that financial indicators could be connected with the potential risks of fraud [6]. Although these models have produced useful information, they are more inclined on studying large-scale financial

fraud in general and not always considering variables in high-tech and R&D-intensive companies. The distinctive characteristic of firms on Science and Technology Innovation Board that is based on a unique hard technology nature demands that industry-specific risk factors should be chosen in order to predict them correctly. Furthermore, the previous studies also emphasize the impact of the auditor expertise in minimizing fraud risk and the effect of subsidies provided by the government on the conduct of corporations [7]. Nevertheless, current evidence implies the necessity of detailed models, which incorporate both non-financial and financial measures to assess the IPO risks equally [8].

### 2.2 Limitations of Existing Research

The existing model variables mostly focus on financial indicators, ignoring the unique non-financial indicators of R&D activities, resulting in limited ability to identify R&D expenses. No special model is designed for enterprises in the science and technology innovation board, which are characterized by "high R&D and light assets", and the selection of variables lacks the adaptability of the science and technology industry. Some studies only pass theoretical deduction or small sample case analysis, lacking a large sample empirical test and model prediction efficiency [9].

The existing research has made some progress in the motivation of R&D expense misrepresentation, model construction, and external supervision, but there are some problems, such as an incomplete variable system, insufficient industry pertinence, and weak empirical verification. In this paper, through the construction of science and technology innovation board special variable system, the design of customized industry model and the empirical test of large samples, the forward-looking identification process of IPO risk of investment banks is innovatively put forward, which provides theoretical basis and practical guidance for improving the supervision system and improving the audit quality.

## 3 Research Method

### 3.1 Research Object and Data Source

The subjects were divided into an experimental group and a control group. The experimental group included the listed companies that were punished or inquired by the CSRC or the Exchange during the IPO in science and technology innovation board from 2019 to 2024, and the sample number was not less than 50 cases. The control group is composed of science and technology innovation board companies that have successfully listed in the same period and there is no dispute about R&D expenses, and they are paired with the experimental group in a ratio of 1:1 according to the industry and scale to ensure the effectiveness and scientific of the comparison. In terms of data sources, the core data comes from the regulatory inquiry letter of the Shanghai Stock Exchange and the company's reply, the R&D expenses chapter in the prospectus and the notes to the audit report, which provides direct evidence support for the research [10].

Auxiliary data are obtained from multiple sources, including patent database to verify the authenticity of R&D results and a talent recruitment platform to verify the real situation of the R&D team. This multi-source data is helpful to comprehensively evaluate the authenticity of the company's R&D activities. In addition, external indicators such as industry R&D investment intensity (from the Yearbook of Statistics Bureau) and regional science and technology subsidy policy (based on the public documents of local governments) are also included in the analysis framework, to more accurately understand the relative position of companies in their industries and the policy environment in their regions. When the sample size is insufficient, the Bootstrap sampling method is used to expand the sample to ensure the stability and reliability of the model.

### 3.2 Variable Selection and Definition

**Table 1.** Selection and definition of variables (based on three paths of false report)

| Variable type | Variable name | Calculation | Economic meaning |
|---|---|---|---|
| Dependent variable | Y | Binary variable (false =1, normal =0) | Supervision conclusion identification |
| R&D project dimension | X1: Project anomaly index | (Number of sudden new projects/total projects) × patent conversion rate | The project was launched suddenly and there was no output |
| | X2: Sudden change of capitalization rate | Capitalization rate of R&D in this period-industry average | Abnormal capitalization adjustment profit |
| Personnel compensation dimension | X3: R&D personnel salary dispersion | R&D personnel compensation variance/management compensation variance | Unreasonable salary distribution |
| | X4: Deviation of R&D personnel proportion | Capitalization rate of R&D in this period-industry average | R&D personnel proportion-industry quantile value |
| Cost classification dimension | X5: Abnormal growth rate of R&D expenses | (R&D expenditure growth rate-revenue growth rate)/σ (industry) | Cost growth is out of touch with business |
| | X6: Proportion of material consumption | Total consumption of R&D materials/R&D expenses | Confuse production cost risk |
| Governance and environment | X7: Auditor's professional expertise | Whether it is the science and technology industry group of Top10 firm (Yes =1) | Audit quality restriction |
| | X8: Dependence of government subsidies | Science and technology subsidy/net profit | Policy arbitrage motivation |

In order to help eliminate the misrepresentation of R&D expenses, six main explanatory variables are distinguished, including project anomalies, the change to capitalization rate, the dispersion of salaries, the proportion of R&D personnel, abnormal increase in expenses and material consumption ratio. Audit quality and government subsidy dependency are also added in order to account external control and optional impact. The

dependent variable is the Y that determines whether the R&D costs were misreported. Table 1 provides an entire definition of all variables.

### 3.3 Model Construction and Solution

In this study, the logistic regression model is used to identify the risk of R&D expense misrepresentation of IPO companies in science and technology innovation board. The model formula is:

$$ln\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i \quad (1)$$

Where P(Y=1 ⊢|X⊣ ) represents the probability of false reporting in a given combination of variables, β _i represents the regression coefficient of all early warning indicators, and positive values represent an increase in the risk of false reporting, while negative values play a restraining role, such as X5 (abnormal growth rate of R&D expenses) as a positive risk indicator and X7 (auditor's industry expertise) as a negative risk indicator.

When finding the solution of the model, first of all, multicollinearity of variables was tested. Ridge regression was used when the variance inflation fact (VIF) was more than 5 since they wanted to correct the problem of collinearity without eliminating other important indicators. Contrastingly, LASSO regression was used to ensure the automatic removal of redundant variables and make it less complex and enhance its explanatory power in the model. The reason why this combined process was selected is that it is a middle ground between robustness and simplicity since Ridge brings robustness whereas LASSO does feature selection. Estimation of parameters then followed through the maximum likelihood method. Ultimately, both discrimination and calibration tests were carried to validate the model, as AUC was greater than 0.85, and the Hosmer-Lemeshow p-value was larger than 0.05, indicating the strong predictive value and good model fit.

### 3.4 Design of IPO Risk Forward-looking Identification Process for Investment Banks

As shown in Figure 1, the forward-looking identification process of IPO risk of investment bank is as follows: firstly, the best-effort team starts the investigation and inputs enterprise data into the risk prediction model, and the model outputs three types of responses according to the probability of falsely reporting R&D expenses: when the probability is higher than 0.7, it triggers a red alarm and enters the in-depth verification stage, including R&D project penetration test, personnel authenticity verification and expense reclassification audit, and finally forms a risk mitigation report for the decision of the core committee; If the probability is between 0.4 and 0.7, a yellow warning will be issued, and supplementary inquiry is needed; If it is lower than 0.4, it will enter the green channel to accelerate. This process realizes risk classification management and precise intervention based on data analysis.

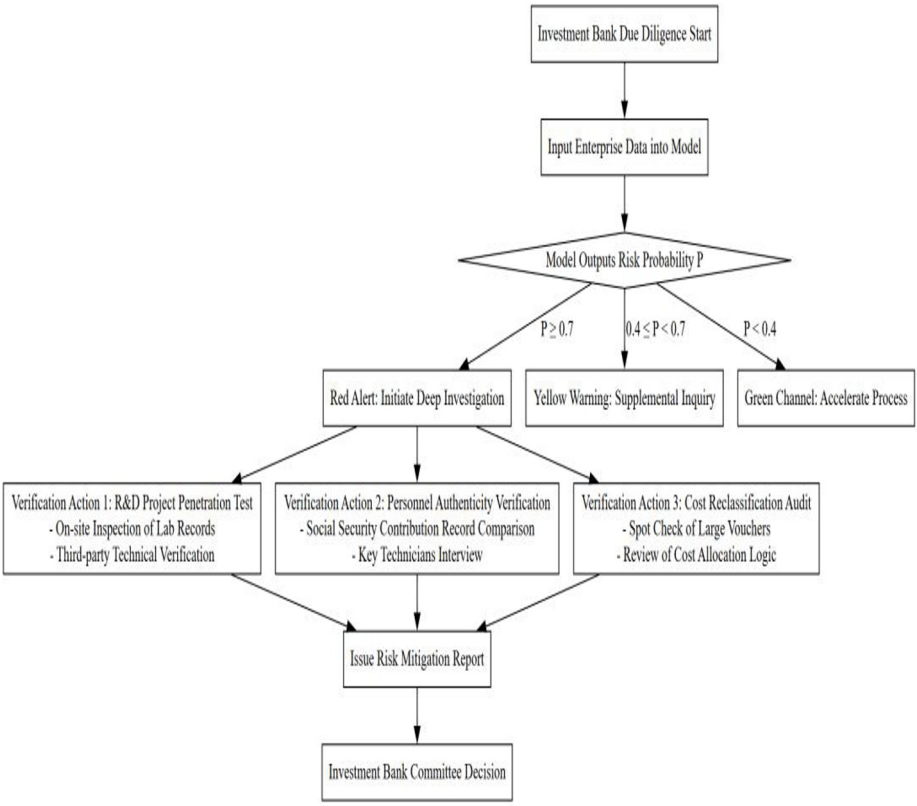


**Fig. 1.** Investment bank IPO risk forward-looking identification process

# 4 Empirical Analysis

## 4.1 Sample Characteristics and Descriptive Statistics

120 samples of IPO enterprises in the science and technology innovation board from 2019 to 2024 were collected (60 in the experimental group and 60 in the control group), and the statistics of key variables are shown in Table 2 below. The project anomaly index (X1) of the enterprises that falsely reported (the experimental group) is significantly higher, showing the characteristics of a surprise project establishment. The abnormal growth rate of R&D expenses (X5) reached 2.5 times the industry standard deviation in the experimental group, far exceeding the actual demand of business growth. The proportion of material consumption (X6) is as high as 42.1% in enterprises with false reports, which exposes the transfer of production costs.

**Table 2.** Descriptive statistics of main variables

| variable | Full sample mean | Experimental group mean | Control group mean | T-test difference |
|---|---|---|---|---|
| X1:project anomaly index | 0.21 | 0.38 | 0.04 | 12.35*** |
| X3: R&D personnel salary dispersion | 1.82 | 2.87 | 0.77 | 8.92*** |
| X5: Abnormal growth rate of R&D expenses ($\sigma$) | 1.20 | 2.50 | -0.10 | 15.67*** |
| X6: Material consumption ratio (%) | 28.3 | 42.1 | 14.5 | 9.84*** |
| X7: auditor's industry expertise (1/0) | 0.48 | 0.25 | 0.71 | -5.33*** |
| X8: Dependence on government subsidies (%) | 22.1 | 35.2 | 9.0 | 7.45*** |

## 4.2 Multiple Logistic Regression Results

Table 3 shows the results of multivariate Logistic regression analysis show that multiple variables have a significant impact on the probability of falsely reporting R&D expenses of science and technology innovation board IPO companies. Among them, the project anomaly index (X1), R&D personnel salary dispersion (X3), abnormal growth rate of R&D expenses (X5) and material consumption ratio (X6) positively promote the risk of false reporting, and their regression coefficients are 1.82, 0.94, 1.05 and 0.03 respectively, which are statistically significant. Auditor's industry expertise (X7) and government subsidy dependence (X8) showed inhibition and slight promotion with coefficients of -1.20 and 0.02, respectively. The marginal effect shows that if X1 increases by one unit, the false report probability will increase by 18.2%, while X7 will decrease by 11%. The model fits well as a whole, with a pseudo $R^2$ of 0.42 and a sample size of N=120, which has strong explanatory power.

**Table 3.** Analysis of risk driving factors of R&D expense misrepresentation

| Variable | Regression coefficient ($\beta$) | Standard error | Advantage ratio (OR) | Marginal effect |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| X1: Project anomaly index | 1.82*** | 0.43 | 6.17 | +18.2% |
| X3: R&D personnel salary dispersion | 0.94** | 0.38 | 2.56 | +9.4% |
| X5: Abnormal growth rate of R&D expenses | 1.05*** | 0.29 | 2.86 | +10.5% |
| X6: Material consumption ratio (%) | 0.03** | 0.01 | 1.03 | +0.3%/ unit |
| X7: Auditor's professional expertise | -1.20*** | 0.34 | 0.30 | -11.0% |
| X8: Dependence on government subsidies (%) | 0.02* | 0.01 | 1.02 | +0.2%/ unit |
| constant | -3.15*** | 0.51 | - | - |

### 4.3 Model Prediction Efficiency Verification

The validation results of model prediction efficiency show that the risk identification model has strong discrimination and calibration. The area under the ROC curve (AUC) is 0.89, which indicates that the model is excellent in distinguishing false reports from normal enterprises. The p-value of the Hosmer-Lemeshow test is 0.51, which shows that the model is well calibrated. The sensitivity, specificity, and accuracy of the model are 86.7%, 83.3% and 85.0% respectively, in the evaluation of confusion matrix with a probability threshold of 0.5, which can effectively identify most enterprises with the risk of false reporting and avoid excessive misjudgment of normal enterprises, and has high practical value, as shown in Table 4.

**Table 4.** Performance evaluation of risk identification model

| Evaluating indicator | Result | Criterion |
|---|---|---|
| Area under ROC curve (AUC) | 0.89 | > 0.85 (excellent) |
| Hosmer-Lemeshow test | $\chi^2$=7.21 (P=0.51) | P>0.05 (well calibrated) |
| **Confusion matrix (probability threshold 0.5)** | | |
| Sensitivity (recall rate) | 86.7% | Falsely reporting the correct recognition rate of enterprises |
| Specificity | 83.3% | Correct recognition rate of normal enterprises |
| Accuracy rate | 85.0% | Overall prediction accuracy |

## 5 Conclusion

The paper tries to explore the misrepresentation of R&D costs in IPO firms that are traded on the Science and Technology Innovation Board during the period 2019-2024. The study introduces essential drivers that ought to be considered when mitigating the risk of fraud when reporting IPOs, through a multivariate logistic model that acts as a guideline and basis to the analysis. A closer examination of the results of the logistic regression supports the statistical finding as it signifies that index of project anomalies, dispersion of salary rates of R&D staff, abnormal increases in R&D costs, and ratio of material consumption are the definite drivers of misreporting risk. On the contrary, the industry experience of auditors minimizes the risk, and the dependency on governmental subsidies indicates a minor inhibitory trend. The model exhibits a high level of predictive performance, with the AUC being 0.89, which is comparable with the previously existing models of financial risk but adds an extra level of precision because it uses R&D specific indicators.

The results are in line with the mainstream literature that appreciate the role of auditor competency and financial anomalies as the main factors in the detection of fraud. Nevertheless, this study brings new knowledge with the incorporation of non-financial factors, including measures of project anomalies that would represent better the nature of companies that issue high-tech IPOs. Practically, investment banks ought to improve their due diligence procedures as based on primary warning signals as determined by this research paper. The regulators are recommended to enhance better supervision structures and impose heavier fines on fraudulent disclosures, but not at the cost of real innovative endeavors. On the whole, the findings recommend practicable measures that can be used to this end to increase transparency, safeguard investors and keeping the capital market integrity thereon.

## References

1. Shi, X., Zhang, X.: Are key clients resources or risks? Evidence from IPO pricing scenarios. Journal of Shanxi University of Finance and Economics **43**(8), 112-126 (2021)
2. Zhang, M., Xu, H.: Case analysis of the implementation effects of audit inquiry and underwriter co-investment systems. Accounting Monthly **46**(1), 102-108 (2025)
3. Hu, Y.: Financing dependence, IPO, and innovation investment of listed companies on the Growth Enterprise Board. Research Management **43**(3), 117-124 (2022)
4. Huang, Y., Luo, D., Zhang, S.: IPO issuance under the comprehensive registration system: easier or harder? Evidence from the capital market. Accounting Monthly **44**(10), 132-139 (2023)
5. He, X., Gu, X., Liu, H.: The impact of risk information disclosure on IPO failure under the registration system. Nankai Business Review **26**(3), 86-98 (2023)
6. Li, X., Bai, Y.: Risk information disclosure of IPO companies and its impact on IPO underpricing: evidence from Chinese listed companies in the U.S. and domestic A-share market. Management Review **33**(7), 29-42 (2021)
7. Liu, N., Li, P.: Securities firm-backed venture capital, institutional reputation, and IPO underpricing. Journal of Beijing Technology and Business University (Social Sciences) **38**(6), 61-72 (2023)
8. Zhang, G., Qin, L., Wang, Y.: Information value analysis of IPO risk factor audit inquiries. Contemporary Economic Science **45**(6), 110-122 (2023)
9. Temkin, Aaron. Polarization and Exchange Effects in the Scattering of Electrons from Atoms with Application to Oxygen. Physical Review, **107**(4), 1004-1012 (2024)
10. Anastassiou G A. Symmetrized, Perturbed Hyperbolic Tangent-Based Complex-Valued Trigonometric and Hyperbolic Neural Network Accelerated Approximation. Mathematics **13**(10), 2227-7390 (2025)