Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 221 (2023) 493–500

Procedia Computer Science

www.elsevier.com/locate/procedia

10th International Conference on Information Technology and Quantitative Management

The impact of subjective and objective inconsistencies in scientific and technological innovation attributes on the listing of enterprises on the Science and Technology Innovation Board

Nan Chen[a], Chen Han[a], Lu Wei[a]*

[a]Central University of Finance and Economics, No. 39 Xueyuan South Road, Haidian District, Beijing 100081, China

Abstract

The attribute of high-tech innovation is the most essential characteristic that distinguishes enterprises on the Science and Technology Innovation Board (STAR) Market from other stock markets. In this paper, we aim to examine the impact of consistency in information disclosure of subjective and objective tech-attributes of enterprises on their success of the application on this board. We construct a science and technology innovation dictionary by Word2vec, and calculate the subjective tech-attributes degree of the enterprise registration statements; Based on the official listing standards document for tech-attributes, we select 7 objective scientific and technological innovation indicators and calculate the enterprises' objective tech-attributes degree via the principal component analysis, and construct the consistency indicator for the two tech-attributes. In the empirical analysis, we collected the registration statement of 708 companies that applied for listing from 2019 to 2022 and studied the inconsistencies between the subjective tech-attributes of the text of the enterprise registration statements and their objective tech-attributes innovation evaluation indicators, and found the factors that affected the successful listing. The results show that both subjective and objective scientific innovation are positively correlated with the successful listing, and the consistency indicator is significantly negatively correlated with the successful listing. When the degree of inconsistency between the subjective and objective tech-attributes is greater, the lower the probability of passing the review and listing of the STAR Market is.

Peer-review under responsibility of the scientific committee of the Tenth International Conference on Information Technology and Quantitative Management

Keywords: *STAR Market; science and technology innovation; IPO; text analysis; principal component analysis*

1. Introduction

Since the launch of the Science and Technology Innovation Board (STAR) Market in 2019, science and

* Lu Wei. Tel.:+86 18510628807.
E-mail address: weilu2014ucas@163.com

1877-0509
Peer-review under responsibility of the scientific committee of the Tenth International Conference on Information Technology and Quantitative Management
10.1016/j.procs.2023.08.006

technology enterprises have applied for listing on the Board enhance their competitiveness. The main regulatory concept of the registration system is "disclosure-based", focusing on the tech-attributes of enterprises [1]. Under the registration-based system, the STAR Market registration statement is one of the important ways for regulators to understand basic information such as an enterprise's scientific and technological innovation capabilities, financial status, and risk factors, among which the scientific and technological innovation capabilities have attracted great attention and directly affect the success of enterprises applying for listing on the STAR Market.

In 2022, the Shanghai Securities Exchange (SSE) issued "The Interim Provisions of the Shanghai Stock Exchange on Application and Recommendation for Issuance and Listing of Enterprises on the STAR Market" [2], which measures the science and technology innovation attributes(tech-attributes) of technology enterprises based on four types of core indicators contained in the registration statements: R&D investment, invention patents, scientific research talents, and business performance. The SCRC and SSE finally decide whether the company can list or not through the contents disclosed in the registration statements. By the end of 2022, the Shanghai Stock Exchange (SSE) had accepted a total of 906 listing applications from issuers, of which 198 failed applications, accounted for 21.54% of the total. Among all enterprises that failed to apply, 44.94% of the reasons for listing failure were related to insufficient *tech-attributes*. Therefore, according to the four types of core indicators that measure the attributes of scientific and technological innovation in this paper, we select 7 representative objective indicators for calculating the objective scientific and technological innovation degree (*Tech_obj*) of enterprises.

The disclosure content of the registration statements is composed of two types of information: digital information and text information, of which digital information reflects the objective R&D investment of the enterprise, and the text information reflects the subjective expression of the enterprise's tech-attributes degree.In the subjective writing process of the registration statements, to specifying the four types of objective scientific and technological innovation indicators, there is also the possibility that enterprises can achieve the impact review through text whitewashing [3]. The text is an important way of information disclosure, which contains incremental information [4], and increasing information disclosure of science and technology enterprises have a positive impact on the market value of enterprises [5]. Based on the theory of information asymmetry, some researches systematically hold that the value of high-tech enterprises depends on R&D activities [6-7]. Although the statements disclose the capital investment in the research and development activities, the details of some sensitive information are often reserved, and information asymmetry still exist [8]. A large amount of unclear and insignificant information is deliberately disclosed, which increases the complexity of text reading and the cost of information search [9]. The establishment of specialized dictionaries based on specific fields can improve the text analysis ability [10]. The quality of information disclosure of enterprises applying for listing on the STAR Market affects the results of the registration review [11], and a high degree of redundancy will lead to the failure of listing [12]. Therefore, this paper scientifically calculates the subjective scientific and technological innovation degree (*Tech_sub*) of enterprises by constructing a proprietary scientific and technological innovation dictionary.

At present, existing studies on information disclosure mostly focus on quantitative financial data and textual features of documents such as financial reports [13-14]. However, the objective data of the enterprise influences noise, and the subjective statement may be manipulated [15]. Few studies have comprehensively considered whether the subjective expression in the registration statement is consistent with the actual level of objective tech-attributes. Therefore, this paper constructs a subjective and objective scientific and technological innovation consistency index to measure the subjective and objective differences.

In this paper, we collect the application forms of 708 enterprises that applied for listing from 2019 to 2022. Combined with the listing application rules documents of the STAR Market, we use the Word2vec word vector model to generate a proprietary science and technology innovation dictionary and calculate the subjective science and technology innovation score of enterprises. Based on the evaluation criteria of scientific and technological innovation attributes in the Provisions, we select 7 representative objective scientific and technological innovation indicators, including the number of patents, the amount of investment in scientific and technological R&D, the proportion of scientific and technological R&D investment, the amount of operating income, the compound growth rate of operating income, the number of scientific research personnel and the proportion of scientific research personnel, and calculate the objective scientific and technological innovation score of enterprises by the method of principal component analysis, and construct the consistency index of subjective and objective scientific and technological innovation. We use the regression model to study the relationship between subjective science and technology innovation, objective science innovation degree, and subjective and objective science and technology innovation degree consistency indicators and the success of the listing on the STAR Market.

The structure of this article is as follows. In section 2, we introduce the calculation method and regression model of tech-attributes. Section 3 provides empirical results. Finally, we give some concluding observations in section 4.

2. Methodology

In this section, we introduce the calculation of subjective and objective scientific innovation degree. In addition, we establish a logistic regression model to study the impact of subjective and objective technological innovation degree, and the consistency in subjective and objective technological degrees on the IPO passing rate.

2.1 The construction of subjective scientific innovation degree - Dictionary

In this paper, the Word2vec word vector model is used to train the relevant science and technology creation dictionary. The algorithm uses a neural network algorithm to find words in the corpus that are semantically similar to a given target word. We download the business rules documents of the STAR Market released by the SSE from 2019 to 2022 as an initial training corpus.

This study follows these steps to construct an innovative dictionary:

- Text preprocessing. After the clause segmentation of the corporate registration statement and the listing rules document, sentences are segmented, and stop words are removed through the Jieba words segmentation package.
- Neologism discovery. Use the Phrase algorithm in the gensim toolkit to construct a new phrase that often appears together in the text and add it to a custom thesaurus.
- Initial dictionary construction. The Word2vec model requires an initial vocabulary list as its input sample. We create an initial list of 132 science and technology words by manually reading all the business rule documents and registration statements.
- Model training. The Word2vec model is trained using the registration statement text after word segmentation and phrase construction, and the word vectors of all words and phrases are obtained by training for 5 iterations.
- Based on cosine similarity, we use the model to output 20 words that are closest to each word in the initial dictionary. After deduplication, there are a total of 1583 science and technology words in the vocabulary list.
- Manual screening. For the generated words, we manually deleted words or phrases that obviously did not conform to the attributes of science and technology to form the final innovation dictionary.

In the end, our science and technology innovation dictionary contains 632 science and technology innovation words, and the *Tech_sub* refers to the percentage of words in the science and technology dictionary contained in the registration statement, which reflect the subjective scientific innovation degree.

2.2 The construction of the objective scientific innovation degree - Principal component analysis

Principal component analysis (PCA) is a widely used objective weighting method. In this paper, we use the principal component analysis method to calculate index weight and use the idea of data dimension reduction to convert multiple indicators into a few comprehensive indicators on the premise of losing little information.

The steps of the principal component analysis used in this article are as follows：

- For the four evaluation criteria of invention patent, R&D investment, business income, and scientific research talents, we selected a total of seven objective indicators to represent the scientific innovation ability of enterprises, which are described in Table 1.
- The principal component is obtained from the correlation matrix, and the corresponding eigenvalues and standard eigenvectors are obtained.
- KMO test and Bartlett test are used to determine whether there is obvious multicollinearity. Test whether the KMO value is greater than 0.6 and $P<0.05$ of the Bartlett test, that is, at the significance level of 5%, there is a strong correlation between sample variables, which is suitable for factor analysis.
- Get the expression of the principal component from the factor score coefficients, determine the number of principal components according to the cumulative variance contribution rate, and select the first several principal components whose cumulative variance contribution rate reaches 70%.

- Take the contribution ratio of each principal component factor as the weight, and finally get the objective scientific innovation index score, which can reflect the objective scientific innovation degree.

Table 1. Description of objective innovation indicators.

Evaluation criteria	Index
Invention patent	Number of patents
R&D investment	Investment in scientific and technological research and development
	Proportion of scientific and technological research and development investment
Business income	Operating revenue
	Compound growth rate of operating income
Scientific research talents	Number of scientific researchers
	Proportion of scientific research personnel

2.3 Construction of subjective and objective consistency indicators

Based on the dictionary method and the entropy weight method, we construct *Tech_sub* and *Tech_obj* of the enterprises on the STAR Market, respectively. In order to further investigate the differences and consider the data dimension, we first map both to the range of (0,1) through normalization as follows:

$$Tech_{sub} = \frac{Tech\_sub - Min_{Tech\_sub}}{Max_{Tech\_sub} - Min_{Tech\_sub}}$$

$$Tech_{obj} = \frac{Tech\_obj - Min_{Tech\_obj}}{Max_{Tech\_obj} - Min_{Tech\_obj}}$$

By normalizing, we convert the indicator values into normalized values in the range from 0 to 1. Therefore, we define the subjective and objective consistency indicators as follows:

$$Tech_{diff} = Tech_{obj} - Tech_{sub}$$

The value of $Tech_{diff}$ can reflect the degree of difference between subjective and objective indicators. If the value of $Tech_{diff}$ is close to 0, it means that the difference between the $Tech_{sub}$ and $Tech_{obj}$ is smaller. If the value of $Tech_{diff}$ is larger, it means that the objective tech-attributes of the enterprise are better than its subjective tech-attributes, and the greater the difference between the two subjective and objective attributes

2.4 Logistic regression

We use the binary logistic regression model to study the relationship between the subjective and objective consistency of corporate science and technology innovation and the IPO passing rate. The logistic regression model is a probability-based binary classification model that is used to solve classification problems. Its basic principle is to classify a given input variable by establishing a relationship between input variables and output variables.

The logistic regression model in this article is represented as follows:

$$P = \frac{1}{1 + e^{-s}}$$

$$s = \alpha + \sum_{i=1}^{n} \beta_i x_i$$

where *P* is the probability value of the explained variable; α is the constant term; β_i is the coefficient to be estimated; x_i is the explanatory variable, including the tech-attributes indicators and control variables, which are explained in Table 2.

Table 2. Indicator definition and specific description.

The variable type	Symbol	Variable definition and description
The variable being explained	Success	Successful listing is assigned a value of 1, otherwise 0
Explanatory variables	$\mathrm{Tech}_{\mathrm{obj}}$	Weighted integration score of enterprise objective science and technology innovation indicators
	$\mathrm{Tech}_{\mathrm{sub}}$	The percentage of the registration statement that contains science and technology coins
	$\mathrm{Tech}_{\mathrm{diff}}$	The difference between objective and subjective scientific innovation after standardization
Control variables	LEV	Gearing ratio for the year before the IPO
	ROA	Net asset margin for the year before the IPO
	CR	Quick ratio for the year before the IPO
	GNPR	Net profit growth rate in the year before the IPO
	A	Whether the accounting firm is in the top ten
	U	Whether the underwriters are in the top ten

3. Empirical analysis

3.1 Data description

Our initial sample includes 533 listed companies and 175 unlisted companies from 2019 to 2022 in the STAR Market. The STAR Market was officially launched on June 13, 2019, so the sample includes all companies from the beginning of the board in this period. For successful listed companies, we collected relevant financial data from the Wind database and the China Economic and Financial Research Database (CSMAR). For failed companies, we manually collected relevant financial data from their registration statement documents. To avoid the influence of extreme values, we performed a 1% winsorize treatment for continuous variables.

3.2 The regression results of subjective scientific innovation attribute

Consider that $Tech_{diff}$ may have different meanings at different intervals : When both $Tech_{sub}$ and $Tech_{obj}$ take larger or smaller values, the difference $Tech_{diff}$ may still be the same. The k-means algorithm can cluster only one indicator, and we can treat the dataset as one-dimensional data. Each data point has only one value, so the k-means algorithm can be considered as a way to find the best-split point. Therefore, we use the k-means clustering algorithm to cluster into three categories, and divide them into three categories of subjective scientific innovation high/medium/low according to the numerical size, and the clustering results and the statistics of subjective and objective scientific innovation and consistency indicators are shown in Table 3.

Table 3. Descriptive statistics.

Category	Overall (N=692)		High(N=75)		Medium(N=242)		Low (N=375)	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
$\mathrm{Tech}_{\mathrm{sub}}$	0.2779	0.016	0.5286	0.008	0.3392	0.002	0.1882	0.003
$\mathrm{Tech}_{\mathrm{obj}}$	0.1149	0.024	0.1079	0.017	0.1117	0.021	0.1283	0.028
$\mathrm{Tech}_{\mathrm{diff}}$	-0.1631	0.043	-0.4207	0.030	-0.2276	0.024	-0.0699	0.032

It can be seen that the mean value of $Tech_{sub}$ (0.2779) of the overall sample is significantly higher than $Tech_{obj}$ (0.1149). For the high level, the mean $Tech_{sub}$ degree (0.5286) was much higher than the $Tech_{obj}$(0.1079), and the gap between $Tech_{sub}$ and $Tech_{obj}$ attributes is the largest. For the medium level, the mean $Tech_{sub}$degree (0.3392) is higher than the objective (0.1117). For the lower level, the mean $Tech_{sub}$ degree (0.1882) is slightly higher than the objective (0.1283). With the decrease of the $Tech_{sub}$ degree of the sample data, the mean $Tech_{obj}$ degree continues to increase, resulting in the gradual reduction of the subjective and objective gap.

3.3 The regression results of objective scientific innovation attribute

Based on "The Interim Provisions of the Shanghai Stock Exchange on Application and Recommendation for Issuance and Listing of Enterprises on the STAR Market", we select 7 representative objective scientific and technological innovation indicators accordingly. This part uses the method of PCA to calculate the *Tech_obj* and *Tech_sub* through 7 objective scientific and technological innovation attribute indicators. The PCA method is objective empowerment when calculating the weight of indicators, and uses the idea of data dimensionality reduction to transform multiple indicators into a few comprehensive indicators.

To eliminate the influence of dimensions, the data are standardized before principal component analysis. KMO and Bartlett tests are performed on the data, wherein KMO=0.63, the P value of Bartlett test < 0.05, that is, at the significance level of 5%, there is a strong correlation between sample variables, which is suitable for factor analysis. The top four principal components with a cumulative contribution rate of 78.05% are selected. The contribution rates of the four principal components after variance maximization rotation are 31.22%, 19.04%, 14.98%, and 12.83%, respectively. The eigenvalues and contribution results of each principal component are shown in Table 4.

Table 4. Principal component eigenvalues and contribution rates.

Ingredient	Initial eigenvalue			Extract the sum of squared loads		
	Total	Percentage variance	Cumulative %	Total	Percentage variance	Cumulative %
1	2.184	31.207	31.207	2.184	31.207	31.207
2	1.333	19.037	50.244	1.333	19.037	50.244
3	1.049	14.980	65.224	1.049	14.980	65.224
4	0.898	12.829	78.053	0.898	12.829	78.053
5	0.627	8.963	87.016			
6	0.472	6.747	93.763			
7	0.437	6.237	100.000			

The linear expressions of the four principal component financial indicators are obtained by the factor loading matrix, and the contribution rate of each principal component is taken as the weight to get the score of *Tech_obj*. We define the *Tech_obj* indicators as follows:

$$Tech\_obj = \sum_{i=1}^{n} \gamma_i F_i$$

where γ_i is the weight of the contribution rate of the principal component, and F_i is the value of the principal component of the financial index.

3.4 The regression results of inconsistency between subjective and objective scientific innovation attributes

We study the influence of $Tech_{obj}$ and consistency of enterprise owners on the IPO passing rate of the STAR Market through the binary logistic regression model, and the experimental results are shown in Table 5. Among them, models 1 to 3 study the effects of $Tech_{sub}$ and $Tech_{obj}$ and consistency indicators, and the combined effects of the three are studied in model 4.

Table 5. The impact of subjective and objective consistency on IPO passing rate.

Notation	Model 1	Model 2	Model 3	Model 4
LEV	0.005	0.007	0.005	0.006
	(0.571)	(0.406)	(0.549)	(0.534)
ROA	0.005	0.010	0.002	0.008
	(0.456)	(0.166)	(0.801)	(0.287)
CR	-0.013	-0.014	-0.003	-0.020
	(0.815)	(0.801)	(0.964)	(0.730)
lnTA	0.238*	0.215*	0.232*	0.235**
	(0.052)	(0.077)	(0.054)	(0.046)
GNPR	-0.001	-0.001	-0.001	-0.001
	(0.243)	(0.280)	(0.130)	(0.337)
U	0.668***	0.658***	0.699***	0.652***
	(0.005)	(0.005)	(0.003)	(0.006)
A	0.154	0.119	0.136	0.150
	(0.556)	(0.646)	(0.600)	(0.565)
$\mathrm{Tech}_{\mathrm{sub}}$	2.952***			2.801***
	(0.002)			(0.001)
$\mathrm{Tech}_{\mathrm{obj}}$		1.965		1.744
		(0.175)		(0.198)
$\mathrm{Tech}_{\mathrm{diff}}$			-1.115*	-0.804**
			(0.083)	(0.060)
Constant	-0.700	0.105	0.079	0.006
	(0.283)	(0.858)	(0.891)	(0.225)

Note:*** p<0.01, ** p<0.05, * p<0.1, * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

According to the results of Model 1, the $Tech_{sub}$ is significantly positively correlated with the successful listing at the level of 1%. According to Model 2, indicators are also positively correlated with successful listing, but the result is not significant. The results of Model 3 show that the consistency index is negatively correlated with successful listing at the level of 10%. According to the comprehensive effect result of Model 4, $Tech_{sub}$ at 1% is significantly positively correlated with the success of listing, and the index is also positively correlated with the successful listing, but the result is not significant, and the consistency index at 5% is negatively correlated with the success of declaration.

The above results show that the greater the difference between the objective index and the subjective index, the lower the success rate of listing. Enterprises with both highly subjective and objective degrees of scientific innovation are the easiest to be audited for listing. That is, when enterprises have strong hard strength of scientific innovation and can fully express it in the registration statements, the probability of passing the audit for listing is the greatest. If an enterprise does not have the corresponding scientific innovation strength, but still whitewash and package the scientific innovation content with words, it often fails to pass the examination of the SSE and CSRC.

4. Conclusions

This paper examines the impact of the consistency of objective tech-attributes of enterprise owners on the IPO passing rate of the STAR Market. We constructed a science and technology innovation dictionary through the Word2vec model to calculate the subjective scientific innovation degree of the enterprise's registration statements. According to the evaluation criteria document of tech-attributes, we select 6 indicators to calculate the objective scientific innovation degree of enterprises by using the principal component analysis. Then we calculate the consistency indicator for the subjective and objective scientific innovation degrees. In the empirical analysis section, we collected the registration statements of 708 companies that applied for listing from 2019 to 2022, and established the logistic regression model to study the impact of the above-mentioned subjective and objective scientific and technological innovation consistency on the IPO passing rate.

The results show that both subjective and objective scientific innovation are positively correlated with the successful listing, and the subjective and objective consistency indicators are significantly negatively correlated. When the degree of inconsistency between the subjective and objective tech-attributes of the enterprise is greater, the lower the probability of passing the review and listing of the STAR Market, and the enterprise with both high subjective and objective scientific innovation degree is most likely to pass the review.

Acknowledgments

This research was funded by grants from the National Natural Science Foundation of China (NSFC Grant Numbers: 72001223）

References

[1] Lu G., Han H., and Chen Y. (2020). Non-punitive Supervision of Accounting Firms and IPO Review Inquiries: Evidence Based on the Registration System of the Science and Technology Innovation Board. Auditing Research, 218(06),43-50.

[2] Shanghai Stock Exchange (SSE)."The Interim Provisions of the Shanghai Stock Exchange on Application and Recommendation for Issuance and Listing of Enterprises on the STAR Market". 2022. Available at: http://www.sse.com.cn/lawandrules/sselawsrules/stocks/review /firstepisode/c/c_20230209_5715889.shtml.

[3] Wu H., Deng Y., Wu, C., Han C., and Wei L. (2022). The impact of disclosure characteristics of application texts on the Science and Technology Innovation Board listed companies. Procedia Computer Science, 199, 246-253.

[4] Loughran T., and McDonald B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.

[5] Hope O., Hu D., and Lu H. (2016). The benefits of specific risk-factor disclosures. Review of Accounting Studies, 21, 1005-1045.

[6] Cho J., and Lee J. (2013). The venture capital certification role in R&D: Evidence from IPO underpricing in Korea. Pacific-Basin Finance Journal, 23, 83-108.

[7] Lu C., Kao L., and Chen, A. (2012). The effects of R&D, venture capital, and technology on the underpricing of IPOs in Taiwan. Review of Quantitative Finance and Accounting, 39, 423-445.

[8] Aboody D., and Lev B. (2000). Information asymmetry, R&D, and insider gains. The journal of Finance, 55(6), 2747-2766.

[9] Loughran T., and McDonald B. (2014). Measuring readability in financial disclosures. Journal of Finance, 69(4), 1643-1671.

[10] Henry E., and Leone A. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. The Accounting Review, 91(1), 153-178.

[11] Li F. (2008). Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics, 45(2-3), 221-247.

[12] Chen S., Li A., Chu Y., Yang S., and Yao, S. (2022). Research on the quality of information disclosure and the registration prediction of the Sci-Tech Innovation Board enterprises based on text classification. Procedia Computer Science, 214, 321-328.

[13] Han C., Wu C., and Wei L. (2023). The impact of the disclosure characteristics of the application material on the successful listing of companies on China's Science and Technology Innovation Board. Journal of Behavioral and Experimental Finance, 37, 100733.

[14] Aasmund E., Erin L.H., and William F.M. (2021). The importance of quantifying uncertainty: Examining the effects of quantitative sensitivity analysis and audit materiality disclosures on investors' judgments and decisions. Accounting, Organizations and Society, 90, 101169.

[15] Kelly K., and Tan T. (2017). Mandatory management disclosure and mandatory independent audit of internal controls: Evidence of configural information processing by investors. Accounting Organizations and Society, 56: 1-20.

[16] Huang X., Teoh S., and Zhang Y. (2014). Tone management. The Accounting Review, 89(3), 1083-1113.