WILEY | Hindawi

Research Article

Intelligent Data Mining Based on Market Circulation of Production Factors

Hefang Sun [ORCID]

School of Marxism, Capital Normal University, Beijing 10010, China

Correspondence should be addressed to Hefang Sun; 2140301011@cnu.edu.cn

R&D investment is an important way to improve scientific and technological innovation capabilities. In an increasingly competitive market, the rapid changes in science and technology have brought new opportunities for enterprise development. However, if production factors cannot be rationally allocated, low allocation efficiency or low allocation efficiency is likely to occur. The phenomenon of excessive overflow of production factors makes the input factors unreasonable and causes the problem of lowering the economic output of enterprises. Therefore, this article analyzes the feasibility and timeliness of R&D investment from factors of production and enterprise output performance based on data mining. The problem is to optimize the rational allocation of future factors of production and provide assistance in achieving a combination of existing and new factors of production. For this test, we selected the companies listed in the Growth Enterprise Market and the survey period is from 2018 to 2020. The data is taken from the Guotaian database, some of which is obtained by manually reading the company's annual report, and a multiple regression analysis model is established and tested. The relationship between R&D investment, production factors, and corporate performance is obtained. The group regression method is used to test the impact of production factors on R&D. Whether input and corporate performance have a moderating effect, and the specific moderating and lagging effects of production factors are investigated. Experiments have proved that the nonstandardized coefficient of R&D investment intensity and operating gross profit margin is 0.714, and the T value of 9.296 is positive and significant. Each increase of enterprise R&D investment intensity by 1 will increase operating gross profit margin by 0.714. The coefficient of operating gross profit margin is much smaller than the coefficient of Tobin's Q value. This shows that the factor of production has a great influence on the relationship between R&D investment and corporate performance. It has the importance of being a specific practical guide for guiding GEM companies in my country with different elemental intensities to carry out R&D activities and improve corporate performance.

1. Introduction

In today's global economic integration, the concept of technological innovation has sounded the clarion call of the new industrial revolution one after another, and the scientific and technological capabilities of enterprises have become the touchstone of whether enterprises can survive. In recent years, Chinese companies are growing rapidly as a whole, but their growth point remains at the top of the total economic volume, and their capacity and quality development are not well coordinated. A series of common problems, such as low technical content and low market value of finished products, have become bottlenecks that limit the develop-

ment and growth of many companies, making them difficult to even break through.

In today's world, economic downward pressure is relatively high, companies want to create economic benefits in the downturn, and research and development activities may be another option for companies. As for the economic consequences that R&D investment will bring to enterprises, many scholars at home and abroad have also conducted fruitful research on this, but the research conclusions are inconsistent and there are big differences. As the main body of scientific and technological innovation, enterprises must proceed from their own perspective and demand resource allocation that meets their own conditions, and the ultimate goal of

the enterprise is to make profits and maximize the effectiveness of production factors in order to improve performance more effectively.

Kim and Yang proposed that it is a top priority to realize the transition from traditional agriculture to modern market agriculture as soon as possible in the construction of a new socialist countryside. Realizing market agriculture can improve China's agricultural core competitiveness, increase farmers' income, and contribute a lot. They also discussed the current situation and influencing factors of the flow of production factors in the process of implementing market-oriented agriculture in our country and discussed the means to promote the circulation of agricultural production factors. However, their research did not clearly propose how to realize the transition from traditional agriculture to modern market agriculture, and the overall research lacks data support [1]. Data mining is playing an increasingly important role in politics, economy, transportation, and life. There have been many cases of applying data mining to solve practical problems at home and abroad. Wang et al. proposed that thresholding the ensemble coherence is a common method to identify radar scatterers that are less affected by decorrelated noise. However, thresholding the consistency may result in the loss of information in areas that experience more complex deformed scenes. If differences in moderately coherent regions have similar behaviors, it is important to consider their spatial correlation for correct reasoning. Then, the information on the low-coherence area may be used in a similar way, while the coherence is used for thematic mapping applications, such as change detection. A method based on data mining and statistical procedures is proposed to reduce the influence of outliers in the results. Our method allows minimizing the outliers in the final results, while preserving the spatial and statistical correlation between the observations. The experimental results lack more data support so that data mining can alleviate the impact of outliers in the results is still doubtful [2]. Tao et al. collect data on IoT devices to achieve better decision-making, higher automation, higher efficiency, productivity, accuracy, and wealth creation. Data mining and other artificial intelligence methods will play a key role in creating a smarter IoT, despite the many challenges. The applicability of eight well-known data mining algorithms to IoT data is tested. Including the deep learning artificial neural network, which constructs a feedforward multilayer artificial neural network for modeling advanced data abstraction, their research data is insufficient, making the data mining results inaccurate [3].

This article will apply the data mining classification algorithm to the actual problems of the performance evaluation of the market-oriented enterprises of production factors. The data comes from the Guotaian database, or by consulting the GEM 2018-2020 annual report, applies the algorithm to the actual problems and improves the market. In this article, we add Tobin's Q value as an indicator to measure the market performance of a company's R&D investment and use Tobin's Q value to reflect its impact on the future of the company. At the same time, this article introduces the variables of factors of production, investigates the relationship between them in detail and in detail, and expands the research space for corpo-

rate performance in future scientific research activities and research ideas.

2. Intelligent Data Mining Market Circulation of Production Factors

2.1. Data Mining

(1) Data mining method

Data mining is different from ordinary information retrieval. Ordinary information retrieval is to directly obtain the required content through query commands, while data mining is to obtain effective information from the data through association rules and machine learning algorithms [4, 5]; the information obtained is indirect and abstract, and the hidden patterns used for evaluation are discovered through data mining. The general framework of data mining is shown in Figure 1.

The prediction task predicts the value of the target attribute based on the existing attributes, mainly regression tasks and classification tasks. The target variable predicted by the regression task is a continuous variable, and the target variable predicted by the classification task is a discrete variable. However, both of these are used to train the prediction model through the training set. The target variable of the training set is known; that is, the training set there is labeled data, so the generated model is a mapping of existing attributes and target attributes generated under supervision, which is a supervised learning method. Descriptive tasks generally summarize the potential association patterns in the data, mainly including cluster analysis, association analysis, and anomaly detection [6, 7].

(2) Data preprocessing

(a) Reasons for data preprocessing

The purpose of data preprocessing is to obtain reliable data for data mining tasks. There are many reasons for the low quality of data, such as equipment failure during the data collection stage, human error during data input, technical errors during data transmission, and the user's cover-up of information [8, 9]. Due to technical or confidential reasons, the data is incomplete data, which is reflected in the missing part of the attribute values of the data or the missing values of some important attributes. The imbalance of data is reflected in the imbalance of the data distribution, and the imbalance of the data will lead to the inaccuracy of the trained model [10].

(b) Data preprocessing method

Data cleaning is to remove the "dirty data" from the data. Smooth noise, general data will have random noise, and the data obtained by smooth processing is more realistic; the detection and deletion of outliers, generally speaking, the data of outliers has a lot of noise, which can be considered noise data. The training of the model is unfavorable,
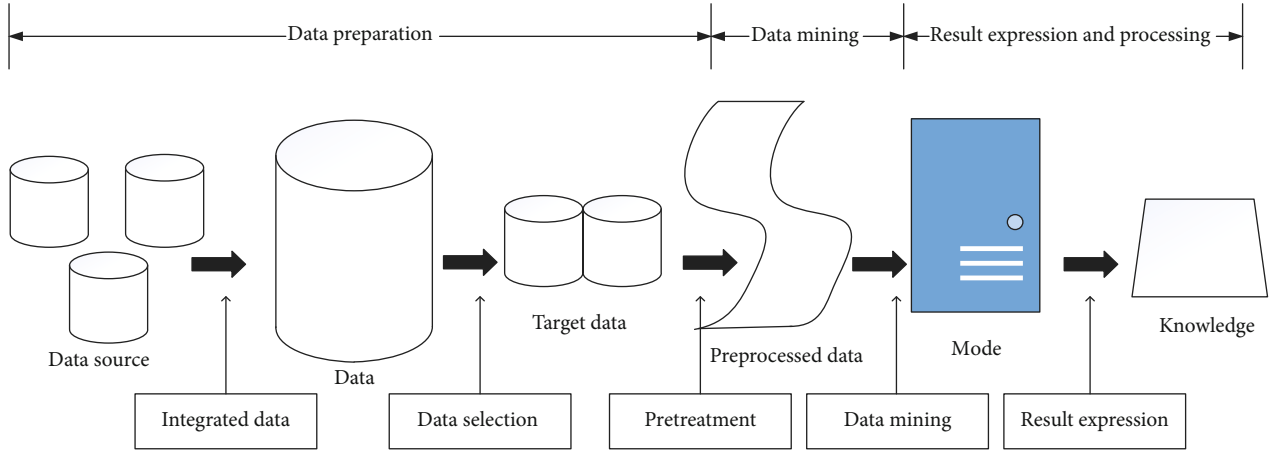
Figure 1: Data mining framework diagram.

Table 1: Confusion matrix of classification results.

Reality forecast result	Positive example	Counterexample
Positive example	TP (real positives)	FN (false counterexample)
Counterexample	FP (false positives)	TN (true counterexample)

and the general processing method is to delete it directly. Since data is generally collected in different systems, and these systems are generally independent, the data is also isolated data from each other [11, 12]. Data reduction is to reduce the scale of data, such as deleting irrelevant attributes, replacing words with numbers in the bag-of-words model, and discretizing continuous values. These processing will greatly reduce the scale of data and save calculation space and time.

2.2. Classification Algorithm Performance Evaluation. For the performance evaluation of the algorithm, in addition to the test set, the performance evaluation index of the algorithm is also required. For different tasks, there are different algorithm performance evaluation indicators to compare the effects of different algorithms or the same algorithm with different parameters [13, 14]. For the two-class classification problem, the category predicted by the classification algorithm and the true category of the sample can be combined to obtain the confusion matrix of the classification result, as shown in Table 1. Among them, TP represents the number of samples predicted by the classification algorithm to be positive and are actually positive examples, called true examples; FP represents the number of samples predicted by the classification algorithm to be positive but actually negative examples, called false positives; FN represents the classification algorithm the number of samples predicted to be negative examples but actually positive examples is called false negative examples; TN represents the number of samples predicted by the classification algorithm to be negative examples and actually is also a negative example, which becomes a true negative example [15, 16].

(1) Accuration (refers to the proportion of correctly classified samples to the total number of samples)

$$\text{accuration} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

(2) Precision (refers to the proportion of the number of positive samples predicted to be correct to the total number of positive samples predicted)

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

(3) Recall (refers to the proportion of the number of positive samples that are correctly predicted to the total number of positive samples)

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

(4) F1 value

F1 value is proposed on the basis of precision rate and recall rate, and is defined as

$$\text{F1} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} \times \text{recall}}. \tag{4}$$

For a classification problem, the prediction accuracy and recall are usually mutually restricted, and the F1 value balances the influence of these two indicators

(5) ROC and AUC

For the predicted value, the larger the probability that the sample belongs to the positive sample, the same, the smaller the value, the greater the probability that the sample belongs to the negative sample. For practical applications, if we pay more attention to the accuracy rate, we can increase the

threshold; if we pay more attention to the recall rate, we can lower the threshold [17, 18]. The proportion of real positive cases predicted to be positive is consistent with the recall rate. The ordinate is the true rate, indicating the proportion of real negative cases predicted to be positive. The definitions are as follows:

$$FPR = \frac{FP}{FP + TN}, \quad (5)$$

$$TPR = \frac{TP}{TP + FN}. \quad (6)$$

AUC is the area included under ROC. When AUC is greater than 0.5, the classification algorithm is effective, and the larger the AUC, the stronger the generalization ability of the classification algorithm. When the AUC is less than or equal to 0.5, the classification algorithm is invalid [19].

2.3. Data Mining Classification Algorithm

(1) Naive Bayes

Bayes' theorem is defined as follows:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \quad (7)$$

where $P(B|A)$ refers to the probability of event B occurring under the condition of event A and $P(B)$ and $P(A)$, respectively, represent the probability of event B occurring in event A, and event B is two independent events in event A [20].

Assume that there are t classified samples in the training sample T, and the sample attribute $X = \{x_1, x_2, \cdots, x_k\}$ belongs to the c class, where x_i represents the i attribute in the sample sign. According to Bayes' theorem,

$$P(c|X) = \frac{P(X|c)P(c)}{P(X)}, \quad (8)$$

where $P(c|X)$ is the conditional probability that the sample attribute X is the class label c; $P(X|c)$ is the conditional probability that the sample attribute is X under the c class; $P(C)$ is the proportion of each type of sample, which is obtained by counting the frequency of each type of sample; and $P(X)$ is a normalized evidence factor and has nothing to do with category [21, 22].

For the unknown sample X, we calculate the conditional probability of this sample for each category separately. The calculation of $P(X|c)$ is more difficult because the attributes of the unknown sample may not appear in the training set. According to the assumption that the attributes are independent of each other, there are

$$P(X|c) = \prod_{i=1}^{k} P(x_i|c). \quad (9)$$

So, the expression of the Bayesian classification algorithm can be written as

$$h(X) = \arg \max P(c) \prod_{i=1}^{k} P(x_i|c) \quad (10)$$

(2) Logistic regression

The Sigmoid function is similar to a step function. At the jumping point, it can be regarded as jumping from 0 to 1 in myopia, which meets the requirements of classification. At the same time, the differentiability of the function ensures that the solution is more convenient [23, 24]. The calculation formula of the Sigmoid function is

$$y = \frac{1}{1 - e^{-z}}. \quad (11)$$

Among them, z is a regression function; set the regression coefficient ω, and the input is X, then $z = \omega^T X$, into the above formula, we can get

$$y = \frac{1}{1 - e^{-(\omega^T X)}}. \quad (12)$$

Ultrasound images have high requirements for edge detail and are nonstationary signals that cannot be met by traditional Fourier transform-based signal denoising methods. Ultrasonic speckle suppression and denoising methods can be broadly divided into spatial area local statistical filtering, anisotropic diffusion filtering, and wavelet transform-based filtering.

$$\ln \frac{y}{1 - y} = \omega^T X. \quad (13)$$

If y is the probability that sample X belongs to a positive sample, $1 - y$ is the probability of a negative example, and $\ln (y/(1 - y))$ is the relative probability that sample X is a positive sample. Considering y as the posterior probability, estimate $P(y = 1|X)$, then Equation (13) can be something like

$$\ln \frac{P(y = 1|X)}{P(y = 0|X)} = \omega^T X. \quad (14)$$

Obviously,

$$P(y = 1|X) = \frac{e^{\omega^T X}}{1 + e^{\omega^T X}}, \quad (15)$$

$$P(y = 0|X) = \frac{1}{1 + e^{\omega^T X}}. \quad (16)$$

To estimate ω through the maximum likelihood method, we can get

$$L(\omega) = \sum_{i=1}^{m} y_i \ln P(y = 1|x_1) + (1 - y_1) \ln P(y = 0|x_1). \quad (17)$$

The loss function $J(\omega)$ is

$$J(\omega) = -\frac{1}{m} L(\omega). \tag{18}$$

This formula cannot be solved analytically, and the gradient descent algorithm can be used to approximate the solution to obtain the best regression parameter ω [25].

3. Experimental Design of Market-Oriented Circulation of Production Factors

3.1. Variable Selection

(1) Dependent variable: corporate performance

There are various indicators that reflect the performance of enterprise R&D input and output, including financial and nonfinancial. Nonfinancial indicators include upgrade of production process, number of patent applications, and update of knowledge and skills, considering that the measurability and availability of this type of data is difficult, so when combined with assumptions and samples, Choose financial indicators to measure R&D input and output. Tobin's Q value is also added as an indicator to measure the market performance of enterprise R&D investment. Regarding R&D investment as a dynamic and long-term process, it has given high weight to the development prospects of the company, and the Tobin Q value is used to reflect its impact on the future of the company.

(2) Independent variable: R&D investment

The R&D expenditure is divided into two parts: expense and capitalization. It includes the actual R&D investment of the enterprise in the year, the investment of personnel and equipment, and the investment of information and creativity. On a quantifiable basis, the current R&D investment amount is obtained. In order to avoid a large gap in the value of the sample's R&D investment due to the difference in company size, the R&D investment ratio is selected as an independent variable.

(3) Categorical variable: intensity of production factors

The factors of production are divided into three dimensions: labor, capital, and technology. Different measurement standards are used for the dimensional division of production factors, and the intensity division is completed according to corresponding indicators and methods. Capital-labor ratio and fixed asset ratio are selected to measure capital-labor intensity. The ratio of capital to labor can reflect the quotas between the two most basic factors of production; the ratio of fixed assets can indicate whether the company has idle fixed assets and the importance of fixed assets, and the two can be directly measured and complemented and analyzed, and the ratio obtained the larger the value, the more important the capital. The ratio of R&D expenses to product costs is cited to reflect how much R&D expenses are condensed in each unit cost; the ratio of R&D personnel to the number of employees reflects the personnel input of the company in conducting R&D activities, combining the above two indicators to distinguish between technology and nontechnology type enterprise.

(4) Control variables

(a) Enterprise scale

Since the scale of listed companies has an impact on output effects, large-scale companies often have certain accumulated advantages, so company scale is a control variable that needs to be established. Since R&D investment not only has an impact on the value of fixed assets but also has a more important impact on the value of intangible assets. In addition, in the R&D investment intensity index, the operating income has been calculated as the denominator, so the total assets are taken as the natural logarithm.

(b) Asset-liability ratio

Since R&D activities are a corporate activity with high capital investment, and the sample selected companies listed on the Growth Enterprise Market, the sample has the characteristics of short business cycle, fast replacement, and large growth inertia. The specific situation of corporate debt management will eventually be reflected in business performance. Therefore, from the perspective of the relationship between corporate capital stock and technological innovation capabilities, the asset-liability ratio is used as a control variable to indicate the abundance of corporate funds. The relationship and meaning between the variables are shown in Table 2.

3.2. Test Subject.
This test selects companies listed on the Growth Enterprise Market, and the research period is from 2018 to 2020. The data comes from the Guotaian database, partly obtained by manually reading the company's annual report. Since the China Securities Regulatory Commission did not include the R&D investment data in the scope of mandatory disclosure by listed companies, the R&D investment data is not available in databases such as Guotaian. Therefore, such data is manually collected by reading the company's annual report. After strict screening of the sample data through the screening criteria, a total of 2942 sample data were obtained.

(i) Screening criteria

Financial data for the period 2018-2020 must be complete. If something is missing, the sample will not be used. If the data is complete, the subject should be representative. If your company's industry data is less than 5 (including 5), it will be deleted. Due to the peculiarities of the financial industry, data on listed companies in the financial industry have also been deleted.

3.3. Moderating Effect Test Method.
For companies, the relationship between R&D activities and corporate performance will also be affected by the company's own "personality." The exact same R&D management model is effective for some

TABLE 2: List of variable relations and meanings.

Variable type	Variable meaning	Variable value and method description
Dependent variable	Enterprise market performance	Market value/total assets at the end of the year
	Corporate financial performance	Main business income-main business cost/main business income
Independent variable	R&D investment intensity	R&D investment/operating income
	Capital labor ratio	Fixed assets/labor force
Categorical variables	Proportion of fixed assets	Net fixed assets/total assets
	Proportion of R&D expenses	R&D expenses/product production costs
	Proportion of R&D personnel	Number of R&D personnel/total number of employees
Control variable	Enterprise size	Natural logarithm of the company's total assets
	Assets and liabilities	Total liabilities/total assets

companies, but invalid for some companies. The "personality of an enterprise" can be expressed as the "foundation" of the enterprise, that is, the factors of production, specifically the number of laborers, the richness of assets, or the level of technology. When the independent variable is a continuous variable and the adjustment variable is a categorical variable, the adjustment effect can be tested by the group regression method; that is, after the adjustment variable M is grouped, X and Y are linear regressions; if there is a difference between the group regression coefficients, it proves that there is a moderating effect.

3.4. Model Building. We establish a multiple regression analysis model to test the relationship between R&D investment, production factors, and corporate performance. The establishment steps are as follows: put independent variables and control variables into the model and perform regression analysis to explore the relationship between R&D investment intensity and corporate performance and introduce adjustment variables, divide the samples into three groups in the three-dimensionality groups of production factors, and build the model as follows:

$$Y_i = a + bX_{1i} + cX_2 + \varepsilon. \tag{19}$$

Among them, Y_i represents the performance of different groups of enterprises, X_{1i} represents the R&D investment of different groups, X_2 represents the control variable, and ε is introduced as a random variable to represent other influencing factors not involved in this study to modify the hypothetical model established. If there are significant differences in the grouping regression under each dimension of production factors, it can be concluded that each dimension of production factors has a moderating effect on the relationship between R&D investment and enterprise performance. On this basis, the subsamples classified according to the various dimensions of the production factors are then used for regression comparison. By observing the significance of b in different groups, there are significant differences. It can further reflect the concrete manifestation of the relationship between production factors and R&D investment and enterprise performance.

3.5. Statistical Processing. Statistical analysis was carried out with SPSS 13.0 statistical software. The significance test of the difference was performed by one-way analysis of variance, the difference between the two groups was tested by LSD-t, and the statistics of intelligent data mining analysis results of the market-oriented circulation of production factors were performed by the group t-test. $P < 0.05$ is considered to be significant and statistically significant.

4. Market Circulation of Production Factors

4.1. Descriptive Statistical

(1) Overall situation analysis

Here, we first analyze the overall situation of GEM listed companies from 2018 to 2020 according to the intensity of R&D investment. The results are shown in Figures 2–4.

It can be seen from Figure 2 that during the three years from 2018 to 2020, the average R&D investment intensity of listed companies on the ChiNext is 6.9%, and the standard deviation is only 0.0675, indicating that this value can well represent the overall level. The R&D intensity of 2% is only the level at which the company can barely survive, and the R&D intensity of 5% is the level at which the company has a competitive advantage. This shows that the Growth Enterprise Market as a whole has innovative vitality. This is mainly because most of the companies listed on the ChiNext are high-tech companies or entrepreneurial companies, and they regard R&D activities as a necessary condition for maintaining innovation and competitiveness.

It can be seen from Figures 3 and 4 that the overall Tobin Q value of the sample varies between 0.6149 and 13.267, with an average of 3.5792 and a standard deviation of 2.67, indicating that there is a large gap in the market value of listed companies on the GEM. The operating gross profit margin values are relatively the same. The standard deviation is 0.167, the minimum is -0.03, the maximum is as high as 0.92, and the overall average is 0.366, maintaining a gross profit margin of about 38%. This shows that the GEM listed companies have continuous competitive advantages.

As can be seen from Figures 2–4, from the perspective of time span, the average value of R&D investment intensity has increased from 7.3% in 2018 to 6.73% in 2019 and 6.67% in
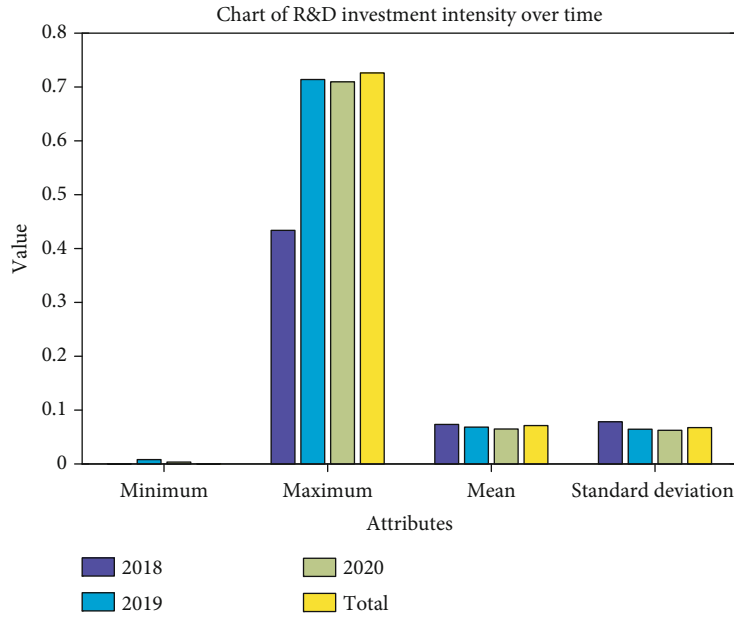
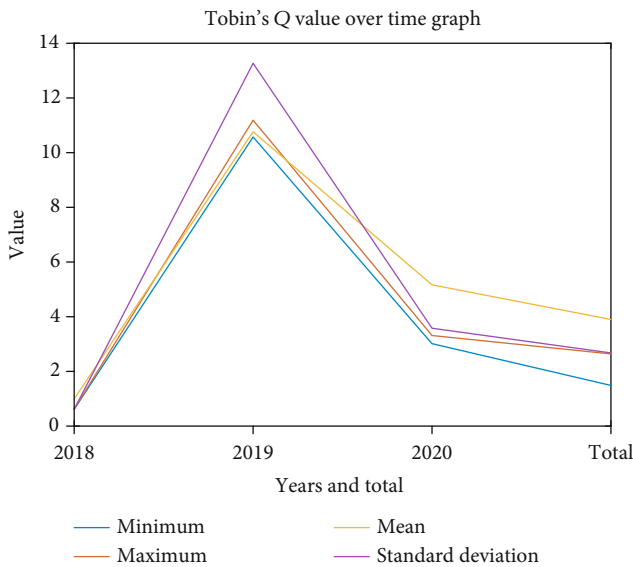Figure 2: Chart of R&D investment intensity over time.

Figure 3: Tobin's Q value over time graph.

It can be seen from Table 3 that the R&D investment intensity of labor-, capital-, and technology-intensive industries are 4.32%, 4.49%, and 8.96%, respectively; Tobin's Q values are 3.613, 3.623, and 4.346, respectively; operating gross profit margins are 0.304, 0.347, and 0.358, respectively, indicating that under different production factor intensities, there is a difference between enterprise R&D investment and enterprise performance. The specific difference is reflected in a stepped difference. Technology-intensive industries far surpass other groups in terms of R&D investment and corporate performance. At the same time, capital-intensive industries are slightly higher than labor-intensive industries. The R&D intensity of technology-intensive samples, Tobin's Q value, and operating gross profit margin are the largest; capital-intensive industries have the smallest standard deviation of R&D intensity, indicating that the R&D intensity of enterprises in this type of industry remains at a relatively stable level; labor-intensive industries, the industry's operating gross profit margin standard deviation, is relatively lowest, indicating that the gross profit margin of this type of industry tends to be stable, but the overall profitability is weak.

4.2. Correlation. This paper uses SPSS 13.0 software to analyze the data and conducts correlation analysis before regression analysis. Correlation analysis is used to describe the degree of interdependence between variables. It can detect whether there is autocorrelation between explanatory variables in the model. The results of correlation analysis are shown in Table 4. The table lists the coefficients of the correlation between R&D investment intensity, corporate performance, and control variables. Based on the correlation analysis results in Table 4, the relationship between the explanatory variables and the explained variables in the model will be described separately.

2020. Although there is a slight downward trend, compared with the 3.5% R&D investment intensity of main board listed companies, it still has a big advantage. Tobin's Q value has been significantly improved in three years. While growing, the gap in the market value has gradually widened, and the standard deviation has increased by 1.26 compared with the previous year. In comparison, the company's operating gross profit margin data has basically stabilized at 0.34 during 2018-2020, indicating that the company's short-term profit level has changed less.

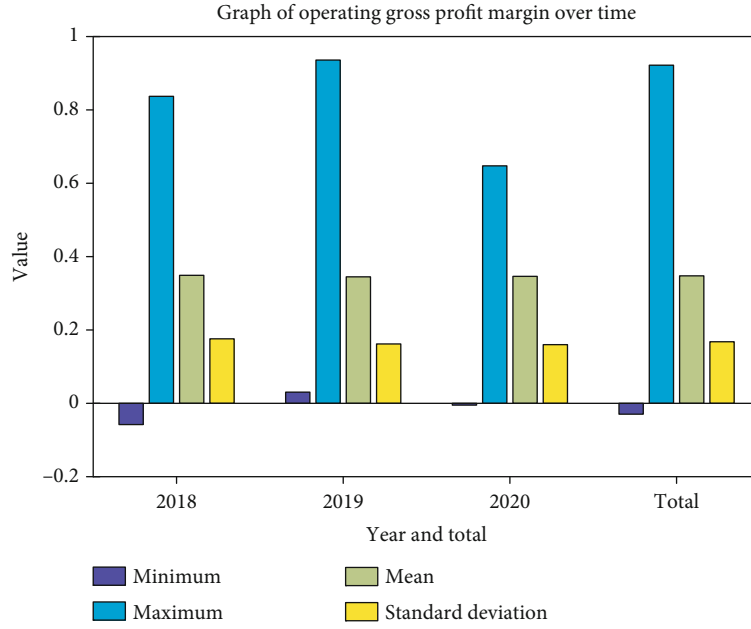(2) Analysis of each grouping situation

FIGURE 4: Graph of operating gross profit margin over time.

TABLE 3: Descriptive statistics grouped by the intensity of production factors from 2018 to 2020.

Intensity of production factors		N	Minimum	Maximum	Mean	Standard deviation
	R&D investment intensity	64	0.003	0.450	0.045	0.061
Labor intensity	Tobin's Q	64	0.976	9.736	3.613	1.838
	Operating gross profit margin	64	0.056	0.663	0.304	0.141
	R&D investment intensity	462	0.003	0.166	0.051	0.028
Capital intensity	Tobin's Q	462	0.619	19.544	3.623	2.514
	Operating gross profit margin	462	0.030	0.959	0.347	0.169
	R&D investment intensity	597	0.009	0.728	0.091	0.084
Technology intensity	Tobin's Q	597	0.844	18.105	4.346	2.298
	Operating gross profit margin	597	-0.060	0.980	0.358	0.182

TABLE 4: Correlation analysis data sheet.

	R&D investment intensity	Assets and liabilities	Enterprise size	Tobin's Q	Operating gross profit margin
R&D investment intensity	1	—	—	—	—
Assets and liabilities	-0.293	1	—	—	—
Enterprise size	-0.132	0.461	1	—	—
Tobin's Q	0.165	-0.216	-0.273	1	—
Operating gross profit margin	0.378	-0.413	-0.119	0.419	1

It can be seen from Table 4 that the correlation coefficients between the independent variable R&D investment intensity and the dependent variable Tobin's Q value and operating gross profit margin are 0.165 and 0.378, respectively, and the associated probability P value is less than the significance level of 0.01. There is a significant positive correlation between. Moreover, the operating margin coefficient is greater than Tobin's Q coefficient, indicating that the correlation between R&D investment and operating gross

profit margin is stronger than its correlation with Tobin's Q value, which indirectly proves that it is suitable for further analysis of different groups and variables. The difference exists. The two control variables of asset-liability ratio and enterprise scale are also analyzed here. It can be seen that the correlation coefficients between asset-liability ratio and Tobin's Q value and operating gross profit margin are, respectively -0.216 and -0.413, with a significance level of 0.001; enterprise size and the correlation coefficients between
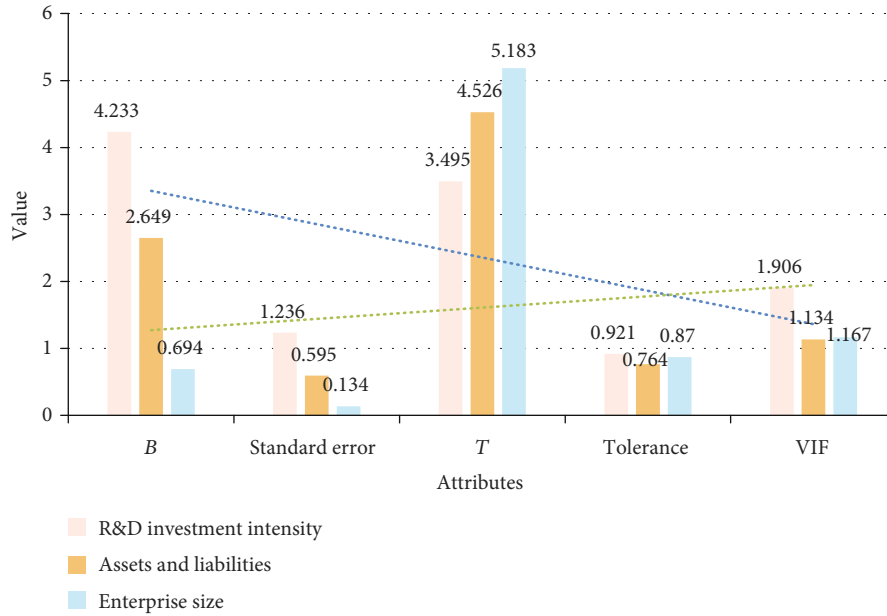
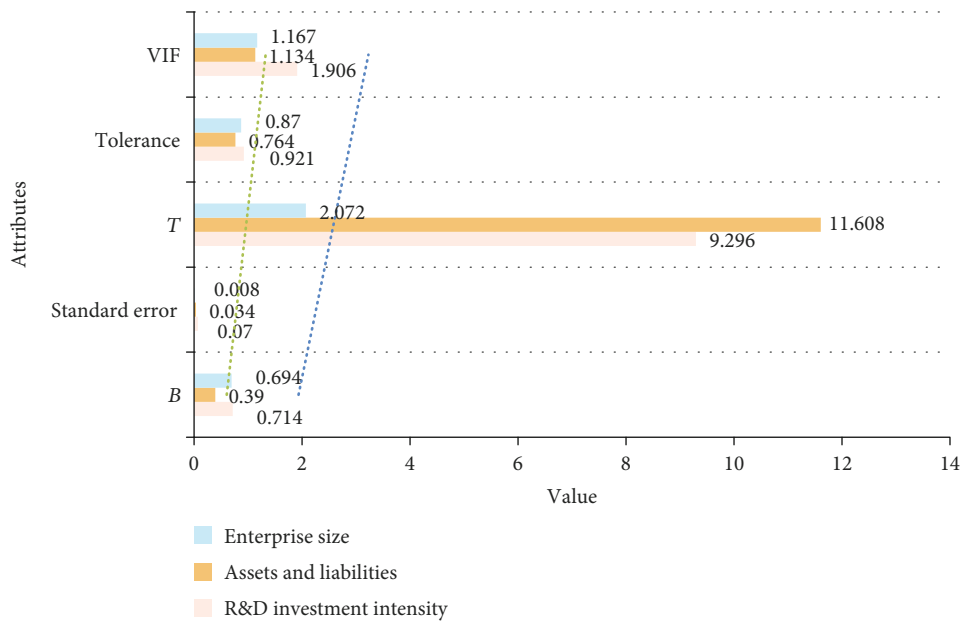FIGURE 5: The regression result of R&D investment on Tobin's Q value.

FIGURE 6: The regression result of R&D investment on operating gross profit margin.

Tobin's Q value and operating gross profit margin are -0.273 and -0.119, respectively, and the significance level is 0.001, indicating that in addition to the impact of R&D investment intensity on corporate performance, there are other variables that act in it.

4.3. Regression R&D Investment on Enterprise Performance

(1) The influence of R&D investment intensity on Tobin's Q value

The statistical performance of this model is significant, reaching 19.118, indicating that the regression model of

R&D investment intensity and Tobin's Q value is overall significant in the regression analysis based on the full sample. 8.8% of the model is adjusted, indicating that the regression equation explains the 8.8% square difference of Tobin's Q value. According to the collinearity diagnosis result, it is found that the tolerances are 0.921, 0.764, and 0.870, respectively, and there are no small values; the variance expansion factor (VIF) is 1.906, 1.134, and 1.167, respectively, and there are no very large values, further explanation there is no problem of collinearity between variables in the equation. The nonstandardized coefficient of R&D investment intensity and Tobin's Q value is 4.233, T value is 3.459, the sign is positive, and it is lower than the significance level of 0.01, which means that for

each increase in R&D investment intensity of enterprises, Tobin's Q value will increase accordingly 4.233 shows that R&D investment is positively correlated with the company's market performance. The results are shown in Figure 5.

(2) The impact of R&D investment intensity on operating gross profit margin

The statistical performance of this model is significant, reaching 117.457, indicating that in the regression analysis of operating gross profit margin, the regression model of R&D investment intensity and operating gross profit margin is overall significant. The adjusted model is 22.5%, indicating that the regression equation explains 22.5% of the variation of operating gross profit margin. At the same time, the collinearity diagnosis result also shows that there is no collinearity problem among the explanatory variables in the equation. The nonstandardized coefficient of R&D investment intensity and operating gross profit margin is 0.714, and the T value of 9.296 is positive and significant, indicating that the R&D investment of an enterprise can directly promote the growth of operating gross profit margin and thus bring business performance to the enterprise. An increase of 1 will increase the operating gross profit margin by 0.714 accordingly. The coefficient of the operating gross profit margin is much smaller than the coefficient of Tobin's Q value, indicating that R&D investment has a greater effect on the market value of the company, which proves that R&D investment and corporate finance performance are positively correlated, and the results are shown in Figure 6.

5. .Conclusions

This article uses the 2018-2020 GEM listed companies as a sample, based on the perspective of production factors, to study the impact of R&D investment and corporate performance. The article divides the entire industry on the Growth Enterprise Market into labor-, capital-, and technology-intensive industries by calculating the corresponding production factor intensity indicators. Through the use of statistical software SPSS 13.0, descriptive statistical analysis, correlation analysis, and regression analysis were performed on the full sample and subsamples. At the same time, the group regression method was used to test whether production factors have a moderating effect on R&D investment and corporate performance. It also examines the specific adjustment and hysteresis effects of production factors in it. In the research of this article, the factors of production are divided into three parts: labor factors, capital factors, and technology factors. It is found that the factors of production have a significant impact on the relationship between R&D investment and corporate performance. This conclusion will guide my country's GEM in the future. Companies with different factor intensives carry out R&D activities to improve corporate performance, which has certain practical guiding significance.

Data Availability

This article is not supported by data.

Conflicts of Interest

The author declares that he/she has no conflicts of interest.

References

[1] W.-Y. Kim and K.-M. Yang, "regional upgrade plan for smart circulation of pedestrian - case study of district circulation including pedestrian -," *Journal of the Architectural Institute of Korea Planning & Design*, vol. 32, no. 11, pp. 39–46, 2016.

[2] L. Wang, R. Huang, and R. Wu, "Tropical cyclogenesis associated with four types of winter monsoon circulation over the South China Sea," *Atmospheric Science Letters*, vol. 17, no. 5, pp. 326–333, 2016.

[3] T. Gao, L. Li, T. Chen, L. Shi, Y. Yang, and G. Li, "DNA-oriented shaping of cell features for the detection of rare disseminated tumor cells," *Analytical Chemistry*, vol. 91, no. 1, pp. 1126–1132, 2018.

[4] Y. Wang, A. D'Ariano, J. Yin, L. Meng, T. Tang, and B. Ning, "Passenger demand oriented train scheduling and rolling stock circulation planning for an urban rail transit line," *Transportation Research Part B: Methodological*, vol. 118, pp. 193–227, 2018.

[5] C. Shiyin and M. A. Zhiyu, "Influencing factors of rural land internal circulation market in Guangdong province," *Asian Agricultural Research*, vol. 12, no. 4, pp. 32–37, 2021.

[6] E. Jernström, V. Karvonen, T. Kässi, A. Kraslawski, and J. Hallikas, "The main factors affecting the entry of SMEs into bio-based industry," *Journal of Cleaner Production*, vol. 141, pp. 1–10, 2017.

[7] M. Yang, X. Zhao, and T. Meng, "What are the driving factors of pesticide overuse in vegetable production? Evidence from Chinese farmers," *China Agricultural Economic Review*, vol. 11, no. 4, pp. 672–687, 2019.

[8] M. Wang and C. Feng, "Impacts of oriented technologies and economic factors on China's industrial climate mitigation," *Journal of Cleaner Production*, vol. 233, pp. 1016–1028, 2019.

[9] G. De Pietro, L. Gallo, R. J. Howlett, and L. C. Jain, "Smart innovation, systems and technologies," in *Intelligent Interactive Multimedia Systems and Services: Proceedings of 2018 Conference*, vol. 98, pp. 53–63, 2019.

[10] J. Chrząszcz, Ł. Skonieczny, G. Protaziuk, M. Kryszkiewicz, and H. Rybinski, "Studies in big data," in *Intelligent Methods and Big Data in Industrial Applications*, vol. 40 of Studies in Big Data, pp. 103–113, Springer International Publishing, Cham, 2019.

[11] H. S. Behera and D. P. Mohapatra, "Advances in intelligent systems and computing," in *Computational Intelligence in Data Mining—Volume 2: Proceedings of the International Conference on CIDM, 5-6 December 2015*, vol. 411, pp. 229–237, Cape Town, South Africa, 2016.

[12] A. Akay, A. Dragomir, and B. E. Erlandsson, "Assessing antidepressants using intelligent data monitoring and mining of online fora," *IEEE Journal of Biomedical & Health Informatics*, vol. 20, no. 4, pp. 977–986, 2016.

[13] H. AbouEisha, T. Amin, I. Chikalov, S. Hussain, and M. Moshkov, "Intelligent systems reference library," in *Extensions of Dynamic Programming for Combinatorial Optimization and Data Mining*, vol. 146 of Intelligent Systems Reference Library, pp. 85–104, Cham, 2019.

[14] Y. Y. Jo, S. W. Kim, S. W. Cho, D. H. Bae, and H. Oh, "High-performance data mining with intelligent SSD," *Cluster Computing*, vol. 20, no. 2, pp. 1155–1166, 2017.

[15] J. Mizera-Pietraszko, P. Pichappan, and L. Mohamed, "Advances in intelligent systems and computing," in *Lecture Notes in Real-Time Intelligent Systems*, vol. 756 of Advances in Intelligent Systems and Computing, pp. 332–341, Springer International Publishing, Cham, 2019.

[16] M. Nematollahi, R. Akbari, S. Nikeghbalian, and C. Salehnasab, "Classification models to predict survival of kidney transplant recipients using two intelligent techniques of data mining and logistic regression," *International Journal of Organ Transplantation Medicine*, vol. 8, no. 2, pp. 119–122, 2017.

[17] J. Zhang, S. O. Williams, and H. Wang, "Intelligent computing system based on pattern recognition and data mining algorithms," *Sustainable Computing: Informatics and Systems*, vol. 20, pp. 192–202, 2018.

[18] V. B. Aggarwal, V. Bhatnagar, and D. K. Mishra, "Advances in intelligent systems and computing," in *Big Data Analytics: Proceedings of CSI 2015*, vol. 654, pp. 717–726, New Delhi, India, 2018.

[19] H. S. Behera, J. Nayak, B. Naik, and D. Pelusi, "Advances in intelligent systems and computing," in *Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017*, vol. 711, pp. 697–707, Beijing, China, 2019.

[20] R. Ghazali, M. M. Deris, N. M. Nawi, and J. H. Abawajy, "Advances in intelligent systems and computing," in *Recent Advances on Soft Computing and Data Mining*, vol. 700 of Advances in Intelligent Systems and Computing, pp. 318–329, Springer International Publishing, Cham, 2018.

[21] M. Alenezi and B. Qureshi, "Advances in intelligent systems and computing," in *5th International Symposium on Data Mining Applications*, vol. 753 of New Feature Extraction Approach Based on Adaptive Fuzzy Systems for Reliable Biometric Identification, pp. 210–222, Springer International Publishing, Cham, 2018.

[22] S. Koren, "How to compare and cluster every known genome in about an hour," *Neural Computation*, vol. 27, no. 7, pp. 1373–1404, 2016.

[23] H. Liu, Y. Wang, J. Liu, J. Yang, Y. Chen, and H. V. Poor, "Authenticating users through fine-grained channel information," *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 251–264, 2018.

[24] S. M. Sombolestan, A. Rasooli, and S. Khodaygan, "Optimal path-planning for mobile robots to find a hidden target in an unknown environment based on machine learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 5, pp. 1841–1850, 2019.

[25] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, no. C, pp. 104–116, 2017.