Impact analysis of intra-firm factors on share prices of Chinese construction SMEs based on the XGBoost algorithm

Xinyi Liu [1, *], Yeqin Xiong [2]

[1] School of Economics and Management, Wuhan University, Wuhan, China, 430072

[2] School of Management, Huazhong University of Science and Technology, Wuhan, China, 430074

* Corresponding author: lexiliuxy@163.com

Abstract. At this stage, small and medium-sized construction enterprises (construction SMEs) are in a difficult position due to fierce competition and severe homogenization. However, current studies on China's construction industry are mainly from the macro perspective of regions and policies. There is a lack of studies that can guide business managers to improve enterprises' performance. This paper constructs a 5-dimensional and 18-indicator system of factors influencing the share price of construction SMEs from an intra-firm micro perspective. Based on this, several XGBoost models were trained, and the highest accuracy was selected as the criterion for feature importance ranking using R^2 and *RMSE* to output the importance ranking of internal factors influencing the share price of Chinese construction SMEs. The results show that the most critical intra-firm factors for the share price of Chinese construction SMEs are the growth capacity of capital, especially the growth ability of shareholders' equity, and the long-term and short-term solvency. This study provides an actionable starting point for Chinese construction SME managers. It uses the XGBoost algorithm, which has higher predictive accuracy than the linear regression commonly used in existing studies, to obtain more reliable factor importance rankings.

Keywords: XGBoost, Share price, Small and medium-sized construction companies, Identification of factors.

1. Introduction

The construction industry is the sector that repairs existing buildings and engages in building installation and is a pillar industry in China. The current distribution of construction enterprises in China is pyramidal: there are few large enterprises but many small and medium-sized enterprises. China's construction industry has a low barrier to entry, a large number of enterprises, serious homogenization of enterprises, fierce competition, and low profits. Many small and medium-sized enterprises (SMEs) have difficulty surviving and developing.

This paper aims to point out starting points for managers of listed Chinese construction SMEs to improve their business conditions and raise their share prices.

There are few studies on the factors influencing share prices in the construction industry in China. Existing researches are primarily written for investors rather than managers. The methods used are mainly based on linear regression, of which the accuracy fit and prediction is insufficient.

DONG [1] used PCA to reduce dimensionality, studied firms listed on the Chinese SME board based on multiple regression models, and found that financial information is an important influencing factor for firms on the Chinese SME board. ZHU & ZHU [2] studied the influencing factors of A-share listed firms falling below net assets based on multiple regression models and found that profitability is significantly negatively related to net breakage after excluding the industry's influence. DAI [3] used factor analysis to reduce dimensionality and studied the financial indicators affecting share price based on multiple regression. He found that profitability is the core of comprehensive evaluation of the financial status of listed companies, among which earnings per share and return on net assets are the financial indicators that most affect share price return. HE & JIAO [4] studied the influence of various cash flow indicators on the share price of electric power listed companies based on multiple regression. It divided cash flow activities into investment, operation, financing, and distribution. They found that investment activities have the strongest correlation with the share price of electric power companies.

Current researches on the construction industry in China primarily construct evaluation systems from a macro perspective, studying the differences in the level of technology, green and ecological development, competitiveness, and development drivers of the construction industry between regions. Thus, lack of a micro perspective from within the enterprise.

XU et al. [5] studied the current situation and influence factors of eco-economic efficiency among different regions in China. They found that eco-economic efficiency was highest in the east and lowest in the west. They also found that optimizing the energy consumption structure and improving technological development could improve the eco-economic efficiency of China's construction industry. CHEN et al. [6] constructed an evaluation system for technology and innovation levels in China's construction industry. They horizontally compared the current technology and innovation levels among construction, manufacturing, and agriculture. They found that technology and innovation levels in China's construction industry had improved between 2009 and 2018, but the improvement was minor compared with other sectors. SHEN et al. [7] studied the differences in international competitiveness levels of China's construction industry in different regions and the effectiveness of factors influencing international competitiveness levels in enhancing regions with varying levels of competitiveness. XIANG et al. [8] studied the driving factors of quality development in the construction industry to help transform and upgrade the construction industry in China.

In this paper, we study the internal factors affecting the share price of Chinese construction SMEs from the micro perspective of the enterprises themselves, based on the XGBoost algorithm, to provide business recommendations for the internal management of construction SMEs in the highly competitive Chinese construction industry. Firstly, in section 2, we construct a system of indicators affecting the share price of Chinese construction SMEs, divided into five dimensions and 18 indicators. Secondly, 49 construction companies with average annual total assets between RMB800 million and RMB4.5 billion were selected as the study population, and the XGBoost algorithm was used to predict and calculate the ranking of the importance of each indicator on the impact of the share price. The importance of impact of each dimension on corporate effectiveness will be determined by the importance of the factors influencing share price under each dimension. This study finds the most critical dimensions and indicators that impact the share price for the reference of internal business managers.

2. Indicator system construction

This paper, 18 indicators were selected from five aspects: Profitability, Condition of Assets, Solvency, Growth Ability, Technology and Innovation as factors influencing the share price of Chinese small and medium-sized construction companies.

2.1. Profitability

The founders of companies set up the business to make profits. Investors buy shares with the expectation that the company will profit to generate income for themselves. Profitability is crucial to running a business: without profitability, the business cannot survive, investors cannot benefit, and the share price is unlikely to rise.

Therefore, this paper has selected indicators in terms of profitability: Net Operating Profit Margin (X_1), Return on Invested Capital (X_2).

2.2. Condition of Assets

Condition of assets examines the efficiency of the business in recovering cash from each asset, the asset structure, and the quality of the assets.

The turnover of each asset category measures the efficiency with which a business recovers its capital. The turnovers can also show the structure of assets and the company's ability to negotiate upstream and downstream. Asset quality is reflected by the ratio of certain assets to others. It reflects the efficiency with which a business generates income from certain assets. If asset utilization is low, i.e., if the assets are inefficient in generating revenue, then the quality of the assets is poor.

To the share price of small and medium-sized Chinese construction companies, which assets are essential in terms of turnover, what assets are important, how much influence does the ability to negotiate with upstream and downstream have, and which asset quality indicators are of most concern?

Therefore, this paper selects indicators in terms of the condition of assets: Inventory Turnover (X_3), Receivables Turnover (X_4), Current Assets Turnover (X_5), the Recovery Ratio of Cash to Total Assets (X_6).

2.3. Solvency

Solvency reflects a company's business risk. Suppose a company does not have enough money to repay loans on time. In that case, it will face a credit crisis, making it more challenging to raise funds or even risk bankruptcy, not to mention continued profitability. But there is a balance to be struck between risk control and the pursuit of profitability: too much risk control is not profitable enough, but only pursuing profitability is prone to capital breakdowns. How much does solvency affect the share price for Chinese construction SMEs? Which part of the asset's solvency has a more critical impact on the share price?

Therefore, this paper has selected indicators in terms of corporate solvency: Debt Asset ratio (X_7), Interest Coverage ratio (X_8), Long-term Debt to Capitalization (X_9), Acid-Test ratio (X_{10}), ratio of Cash to Current Liabilities (X_{11}).

2.4. Growth Ability

Growth capability refers to the ability of a business to continue to grow in the future, mainly including the preservation and growth of assets and revenue. Growth is primarily a function of trend and speed. The trend of growth refers to whether the business will continue to be profitable in the future. The speed of growth refers to the extent to which a company can increase its profitability over a given period. The ability of a business to grow has a bearing on its future viability and profitability expectations and is, therefore, important to managers, shareholders, and potential investors. Expectations about the future growth of a business can also affect the share price.

Therefore, this paper selects indicators in terms of corporate growth ability: Increase rate of Operating Revenue (X_{12}), Operating Profit Growth Rate (X_{13}), Capital Accumulation Rate (X_{14}), Total Assets Growth Rate (X_{15}).

2.5. echnology and Innovation

The technology and Innovation section is separated from section 2.4 because of its importance to the company's future sustainability. China's construction industry, with its low level of information technology, low resource utilization, and low productivity, requires a focus on technology and innovation.[9] Focusing on technology level and innovation capability, investing more capital and workforce in technology research and development can improve work efficiency and reduce costs. So, to what extent do the level of technology and innovation capability affect the share price of small and medium-sized construction companies in China?

Therefore, this paper selects indicators in terms of the technology and innovation level of companies: RD Spend Sum Ratio (X_{16}), RD Person Ratio (X_{17}), number of Patents Granted (X_{18}).

In conclusion, Table 1 shows the indicator system of intra-firm factors influencing the share prices of Chinese construction SMEs.

Table 1. Indicator system of intra-firm factors influencing share prices of Chinese construction SMEs

Secondary Indicators	Tertiary indicators	Formula	Unit	Reference
Profitability	X_1	Net Operating Income / Revenue	%	[3]
	X_2	Net Operating Profit After Tax / Invested Capital	%	[3]
Condition of Assets	X_3	COGS / Average Value of Inventory	times	[1]
	X_4	Net Credit Sales / Average Accounts Receivable	times	[1]
	X_5	Total Sales / Average Current Assets	times	[1]
	X_6	Net Operating Cash Flow / Total Assets	%	[1]
Solvency	X_7	Total Debt / Total Assets	%	[1]
	X_8	EBIT / Interest Expense	%	[1]
	X_9	Long-Term Debt / (Long-Term Debt + Shareholders' Equity)	%	[1]
	X_{10}	(Current Assets – Inventory) / Current Debt	%	[1]
	X_{11}	Net Operating Cash Flow / Current Debt	%	[1]
Growth Ability	X_{12}	Revenue Growth / Total Revenue of the previous year	%	[10]
	X_{13}	Operating Profit Growth / Total Revenue of the previous year	%	[10]
	X_{14}	Equity Growth / Beginning Equity	%	[10]
	X_{15}	Asset Growth / Beginning Asset	%	[10]
Technology and Innovation	X_{16}	R&D input / Total Revenue	%	[9]
	X_{17}	Technical Development Staff / Total Workforce	%	[9]
	X_{18}	Number of Patents Granted	item	[9]

3. XGBoost Algorithm

This paper uses XGBoost [11] to rank the intra-firm influence factors on share prices.

Extreme Gradient Boosting (XGBoost) is an ensemble algorithm developed from GBDT (Gradient Boosting Decision Tree) based on CART (Classification and Regression Tree).[12] Chen and Guestrin proposed the XGBoost method in 2016.[13] XGBoost is widely used in industry and Kaggle competitions for its high prediction accuracy and fast learning speed.[14]

XGBoost also excels in ranking feature importance due to its high prediction accuracy. The more a feature improves the model's goodness of fit, the more critical it is. Loss reduction quantifies the enhancement of fit and thus can measure the prediction performance. A larger loss reduction is preferred. The importance of a feature is determined by to what extent the prediction performance descends after destroying the feature by replacing its data with random noise.[15]

Consequently, the validness of feature importance rank highly relates to the accuracy of prediction that the model performs. Therefore, XGBoost with high accuracy is selected to be the predicting model. Furthermore, the feature importance score is automatically calculated as the XGBoost model runs.

The definition function of the XGBoost model based on multiple CARTs is listed as follows:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$
$$F = \{f(x) = w_{q(x)}\}(w \in R^T, q: R^m \to T) \tag{1}$$

Where F symbolizes all CARTs that can be used. w denotes leaf weight, or called leaf score. q maps each instance x_i to the corresponding tree number T. f_k shows the corresponding leaf score w of each instance x_i. T is the number of leaf nodes. $\hat{y}_i$ is the prediction of the model.

To control the complexity of the model, the objective function of the XGBoost model is added by the L2 regularization term:

$$\text{Obt}\ (\phi) = \sum_{i} l(y_i, \hat{y_i}) + \sum_{k} \Omega(f_k)$$
$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2}$$

Where l measures the difference between the prediction $\hat{y_i}$ and the target y_i. $\Omega(f_k)$ is the L2 regularization term used to control the complexity of the model and prevent overfitting.

Similar to GBDT, XGBoost also optimizes the objective function by cumulative iteration. Assuming that the current prediction $\hat{y}_i^{(t)}$ is at the t-th iteration, the objective function is expressed as follows:

$$Obt^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{3}$$

Here, formula (3) is at the t-th iteration. $f_t(x_i)$ is the newly added tree model. $\hat{y}_i^{(t-1)}$ is the prediction after $t-1$ iterations.

Approximating formula (3) by the second-order Taylor expansion:

$$Obt^{(t)} \simeq \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i)) + \Omega(f_i) \tag{4}$$

Where g_i and h_i are respectively the first-order partial derivative and the second-order partial derivative of the loss function l concerning $\hat{y}_i^{(t-1)}$, which is also known as the first and second-order coefficients of the Taylor series.

Removing constant terms for they have no impact on optimizing the model, the objective function is simplified as follows:

$$Obt^{(t)} \simeq \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{5}$$

Define $I_j = \{i \mid q(X_i) = j\}$, and the optimal score of each leaf j is:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{6}$$

Where I_j is the instance set of leaf j.

The corresponding objective function is:

$$Obt^{(t)} = -\frac{1}{2} \sum_{i=1}^{n} \frac{(\sum_{i \in l_j} g_i)^2}{\sum_{i \in l_j} h_i + \lambda} + \gamma T \tag{7}$$

Formula (7) is the objective function used to measure the quality of a particular tree structure. The outcome of formula (7) should be interpreted as: the smaller it is, the better the tree structure.

Having formula (7), we get the following function to decide how to split each leaf:

$$Gain = \frac{1}{2} [\frac{(\sum_{i \in l_L} g_i)^2}{\sum_{i \in l_L} h_i + \lambda} + \frac{(\sum_{i \in l_R} g_i)^2}{\sum_{i \in l_R} h_i + \lambda} - \frac{(\sum_{i \in l} g_i)^2}{\sum_{i \in l} h_i + \lambda}] + \gamma \tag{8}$$

Where $I = I_L \cup I_R$. I_L and I_R are respectively the left and the right subset of the instance set I of a particular leaf after splitting. Formula (8) measures the loss reduction after node splitting. Therefore, it can be used to decide the optimal split of each leaf: the larger the outcome of formula (8) is, the more the loss is reduced, and the better the split.

The above briefly introduces how XGBoost fits data and predicts. Below is how XGBoost calculates the importance score of each feature.

The importance score of the feature $X_t = k$ at every single CART T is:

$$S_k^2(T) = \sum_{t=1}^{J-1} \hat{\tau}_t^2 \tag{9}$$

J represents all nodes of this tree T, and $J-1$ signifies all internal nodes of T. t represents every node of the tree T. X_t is the feature that splits the tree at node t. $\hat{\tau}_t^2$ is the squared loss reduction after node t splits.

Formula (9) means to add up all of the squared loss reduction that a feature generates in a tree. The larger the value is, the more this feature contributes to making the model better fit the data.

XGBoost ranks the feature's importance by adding the squared loss reduction of this feature in each tree, then averaging the sum over the number of trees. The overall importance score of the feature $X_t = k$ is:

$$S_k^2(T) = \frac{1}{M} \sum_{m=1}^{M} \hat{\tau}_l^2(T_m) \tag{10}$$

Where M is the number of trees that feature X is used to split nodes.

Consistent with formula (9), a larger outcome in formula (10) ranks feature X higher.

4. Performance Estimators

The reliability of the feature importance rank depends highly on the model's prediction performance. A feature contributes to a better data fit as applied in splitting nodes. The more a feature enhances the goodness of fit, the more critical it is regarded. If the model's prediction is unreliable, the feature importance correspondingly loses validity.

Therefore, two indicators *RMSE* and R^2 are applied to evaluate the accuracy of the established model by comparing predicted values with actual target values.[14] The established model is trained in the training set, and the indicators are used in the test set to evaluate the predicting performance of the established model.[16]

RMSE (Root Mean Squared Error) is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{11}$$

R^2 (Coefficient of Determination) is defined as:

$$R^2 = \sqrt{1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y_i})^2}} \tag{12}$$

Where y_i is the actual target value, or called the observed value. $\hat{y}_i$ is the predicted value. $\overline{y_i}$ is the mean of target values. N is the size of the test set. *RMSE* measures the errors between predicted values and target values. *RMSE* ranges from 0 to positive infinity. An *RMSE* closer to 0 is preferred. Because a smaller *RMSE* means smaller prediction errors.[17] R^2 measures how well the prediction of the model fits the data. Generally, R^2 ranges from 0 to 1. An R^2 closer to 1 is preferred. A negative R^2 shows that the evaluated model predicts worse than the baseline model, which uses mean values of y_i to predict.

5. Empirical Analysis

5.1. Sample Selection and Data Sources

In this paper, 49 construction companies with total annual assets between $800 million and $4.2 billion were selected from the CSMAR database according to the 2012 edition of the industry classification of the China Securities Regulatory Commission. Data were collected from the CSMAR database for the period 2007 to 2021. The data in this paper are all obtained from the CSMAR database. The annual closing prices of individual stocks in the stock market series of the CSMAR database are used as the dependent variable for each company's share price. All the indicators used as independent variables are obtained from the sub-database of the company research series. The indicators in Technology and Innovation are acquired from the R&D and innovation database of listed companies. All other indicators are taken from the financial indicators analysis database.

Some Financial Indicators Analysis Database indicators have different calculations and are distinguished by the indicator suffixes A, B, and C. Indicators with the suffix B are selected for X_5, X_{12}, and X_{13}. Indicators with the suffix A are selected for X_8, X_{14}, and X_{15}. All other selected indicators have only one calculation method in the database and do not need to be distinguished.

When processing the missing values for indicators X_1 to X_{18}, the required data is added from the company's research series. The missing values are filled in according to the corresponding formula for the indicator. When collating the patent data for X_{18}, only the patent data corresponding to ApplyTypeCode S5204, S5206, and the corresponding patent data for S5202 and S5203 appearing immediately after the above fields are retained and processed as meaning "at least cumulative number of patents owned in the year".

The remaining missing values that could not be filled in directly with the raw data according to the indicator formula were averaged five times using SPSS multiple interpolations to obtain the final data. Descriptive statistics for the data are shown in Table 2.

Table 2. Descriptive statistics of indicators and share price

Indicator / Unit	Min	Max	Mean	Median	Std
StockPrice/CNY	1.310	78.810	14.747	11.040	12.122
X_1 / %	-7.864	3.469	-0.099	0.043	0.816
X_2 / %	-1.754	0.304	0.003	0.050	0.236
X_3 / times	0.088	9013.155	39.893	3.406	484.035
X_4 / times	0.070	728.652	16.579	2.589	68.479
X_5 / times	0.059	2.446	0.779	0.696	0.466
X_6 / %	-0.258	0.325	0.004	0.004	0.078
X_7 / %	0.028	1.081	0.529	0.534	0.206
X_8 / %	-983.173	3153.235	17.445	2.305	218.956
X_9 / %	0.000	1.508	0.111	0.023	0.174
X_{10} / %	0.122	34.878	1.611	1.172	2.472
X_{11} / %	-5.936	1.786	-0.001	0.008	0.408
X_{12} / %	-0.834	363.068	1.399	0.059	19.545
X_{13} / %	7197.141	12463.762	22.821	-21.500	1035.040
X_{14} / %	-1.160	10.500	0.256	0.061	0.969
X_{15} / %	-0.758	12.727	0.229	0.114	0.822
X_{16} / %	0.020	15.651	3.310	3.290	1.521
X_{17} / %	2.031	87.419	19.276	18.983	8.233
X_{18} / item	1.000	515.000	98.667	98.000	43.183

5.2. Model Training

After selecting and processing the data, the model is set up, and the parameters are adjusted to obtain multiple models for training and prediction. The model with the best prediction is selected for importance ranking according to the estimators in section 4.

This paper uses 70% of the overall data for training and 30% for testing. The data is shuffled before training. A gradient boosting tree (gbtree) was used as the base learner. The learning rate is set to 0.1. The L1 regular term and the minimum weight of samples in leaf nodes are set to 0. The L2 regular term, sample feature sampling rate, tree feature sampling rate, and node feature sampling rate are all set to 1, and the maximum depth of the tree is set to 10. Ten-fold cross-validation was used to achieve better generalization results.

Since the number of base learners (gbtrees) is the parameter that most affects the model's performance, changing the other parameters hardly affects the model's prediction performance. In this paper, we only change the number of base learners (gbtrees) from 100 to 1000 in steps of 100, training ten models while keeping the other parameters mentioned unchanged. The respective R^2 and *RMSE* are shown in Figure 1.

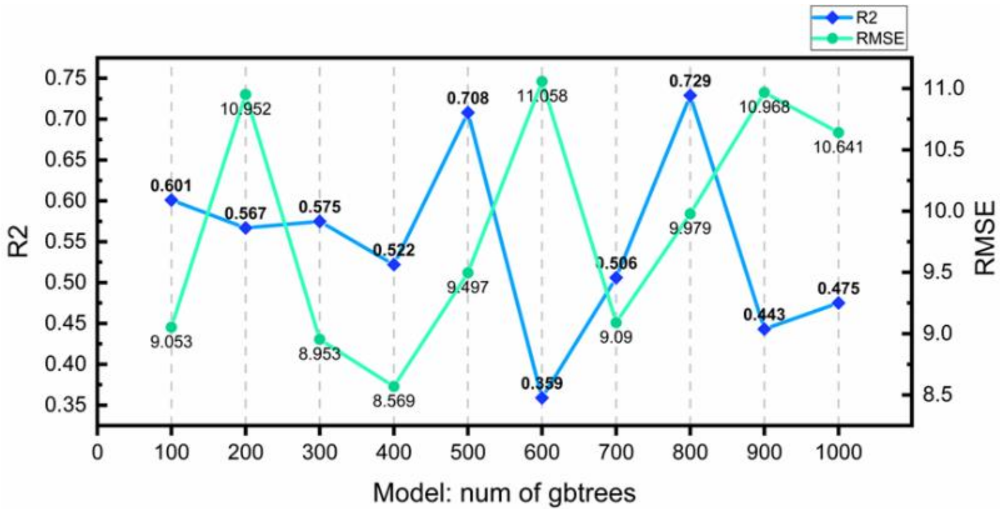

Figure 1. R^2 and *RMSE* of different models

RMSEs do not differ significantly between models. With the model trained by the 800 base learners (gbtrees) having the largest R^2 of 0.729 and the best prediction, so this set of results was chosen to rank the importance of the features. The model's predicted values trained by the 800 base learners on the test set are compared with the true values in Figure 2.

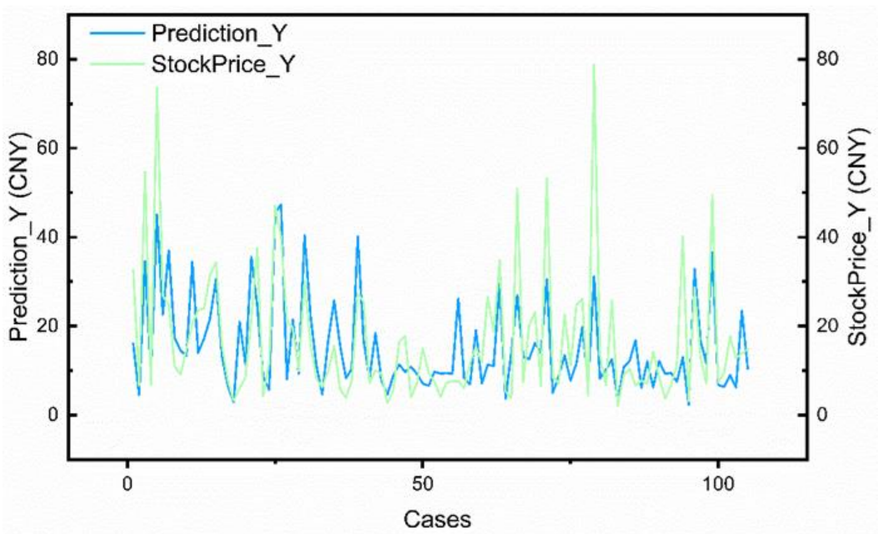

Figure 2. Prediction vs. True value (800 gbtrees)

5.3. Results and Analysis

Based on the above, the model with the best prediction results obtained from 800 base learners was used for feature importance ranking. The results are shown in Figure 3.

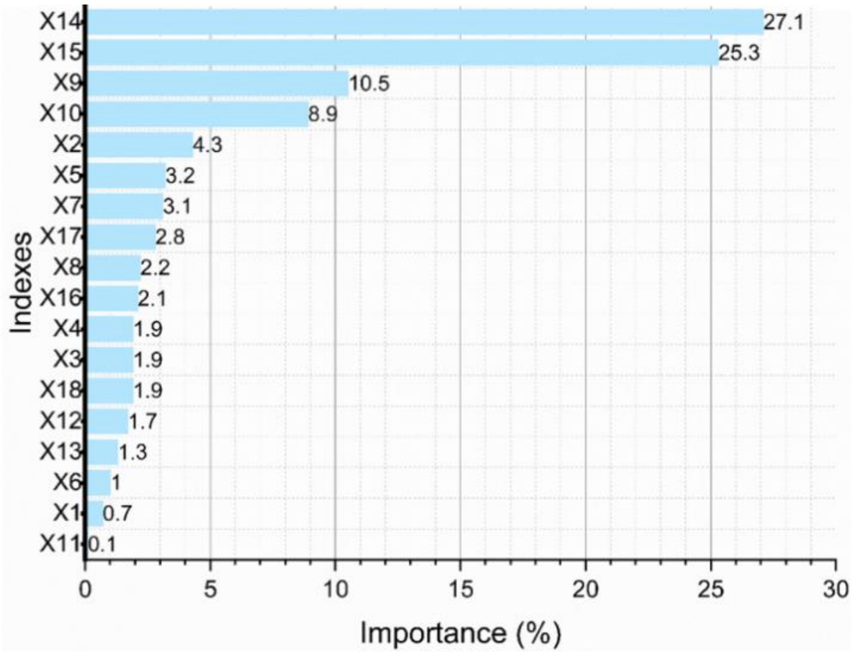

Figure 3. Importance Ranking

According to the graph, the four most important factors affecting the share price of Chinese small and medium-sized construction companies are X_{14} Capital Accumulation Rate (27.1%), X_{15} Total Assets Growth Rate (25.3%), X_9 Long-term Debt to Capitalization Ratio (10.5%), and X_{10} Quick Ratio (8.9%).

X_{14} and X_{15}, with a total of 52.4% importance, are both part of the growth ability of corporate capital. The first-ranked X_{14} refers specifically to equity's growth rate and trend, while the second-ranked X_{15} reflects total assets' growth rate and trend. X_9 and X_{10}, with a total of 19.4% importance, reflect long- and short-term solvency, respectively.

The top four indicators accounted for 71.8% of the importance, with the other indicators being much less important than these four.

Therefore, to improve the share prices of construction SMEs in China, managers could focus most on capital accumulation, especially the accumulation of shareholders' equity, and on the long-term and short-term solvency of the companies.

6. Conclusion

In the study of share price influencing factors, this paper uses the more accurate XGBoost algorithm to obtain a more reliable ranking of the importance of the features than the current commonly used linear regression. The importance ranking of the internal factors influencing the share price of Chinese small and medium-sized construction companies obtained in this paper is:

(1) The top four indicators that have the greatest impact on the share price of Chinese construction SMEs account for 71.8% of the importance of these indicators. The other indicators are much less important than these four indicators. Thus, to improve share prices, managers of Chinese construction SMEs could focus on capital growth, especially the growth of shareholders' equity and the company's long- and short-term solvency.

(2) The top two most influential indicators focus on the growth ability of a company's capital, with equity's growth ability being particularly important. The top indicator of importance is X_{14} Capital Accumulation Rate (27.1%), and the second is X_{15} Total Assets Growth Rate (25.3%).

(3) The third and fourth most important indicators are solvency indicators, which focus on a company's long- and short-term solvency, respectively. The third-ranked indicator, X_9 Long-term Debt to Capitalization (10.5%), reflects long-term solvency, while the fourth-ranked indicator, X_{10} Acid-Test ratio (8.9%), reflects short-term solvency.

References

[1] Dong Xinxin. An empirical study on the factors influencing share price based on principal component analysis: empirical data from the SME board [J]. Communication of Finance and Accounting, 2011 (36): 64 - 67.

[2] Zhu Xiping, Zhu Wei. Analysis of factors influencing the share price of listed companies breaking net[J]. Communication of Finance and Accounting, 2015 (33): 91 - 94.

[3] Dai Qingwen. Analysis of factors influencing the share price return of listed companies in China [J]. Friends of Accounting, 2012 (10): 114 - 116.

[4] He Qinhua,Jiao Anyong. Analysis of factors influencing share prices of power listed companies[J]. Finance and Accounting Monthly,2008(35):41 - 42.

[5] Xu Shuitai, Ma Caiwei, Zhang Shenglong, Yuan Beifei. Measurement of eco-economic efficiency and influencing factors in China's construction industry [J]. Environmental Pollution & Control, 2022, 44 (06): 833 - 840.

[6] Chen Ke, Gao Liangfei,Yu Hongliang. A study on the level of technological innovation in China's construction industry: a cross-time and cross-industry perspective [J]. Science and Technology Management Research, 2021, 41 (19): 56 - 61.

[7] Ji Hongmei, Wang Xiaoping, Luo Biao. Comprehensive evaluation and analysis of regional construction industry science and technology competitiveness: an empirical study based on panel data of Chinese provinces and cities[J]. World Sci-Tech R & D, 2012, 34 (02): 349 - 354.

[8] Xiang Yong, Zheng Mao, Dai Tianhui. Research on the dynamic factors and influencing mechanism of high-quality development of China's construction industry [J]. Construction Economy, 2019, 40 (12): 15 - 20.

[9] WU Xiang-Hua,CHU Xin-Yi. Research on the efficiency and influencing factors of the construction industry's science and technology innovation of listed enterprises: based on data envelopment analysis and Tobin's model [J]. Science and Technology Management Research, 2022, 42 (13): 89 - 96.

[10] Liu W, Ma Q, Liu X. Research on the dynamic evolution and its influencing factors of stock correlation network in the Chinese new energy market[J]. Finance Research Letters, 2022, 45: 102138.

[11] Vuong P H, Dat T T, Mai T K, et al. Stock-price forecasting based on XGBoost and LSTM [J]. Computer Systems Science and Engineering, 2022, 40 (1): 237 - 246.

[12] Zhou J, Qiu Y, Khandelwal M, et al. Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations[J]. International Journal of Rock Mechanics and Mining Sciences, 2021, 145: 104856.

[13] Parsa A B, Movahedi A, Taghipour H, et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis [J]. Accident Analysis & Prevention, 2020, 136: 105405.

[14] Dong W, Huang Y, Lehane B, et al. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring [J]. Automation in Construction, 2020, 114: 103155.

[15] Zheng H, Yuan J, Chen L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation [J]. Energies, 2017, 10 (8): 1168.

[16] Zhou J, Qiu Y, Zhu S, et al. Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization[J]. Underground Space, 2021, 6 (5): 506 - 515.

[17] Zhang W, Wu C, Zhong H, et al. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization [J]. Geoscience Frontiers, 2021, 12 (1): 469 - 477.