Proceedings of ICMRED 2026 Symposium: Financial Innovation, Risk Governance, and the Dynamics of Global Capital Flows
DOI: 10.54254/2754-1169/2026.BL34144

Drivers of Stock Prices in China's State-owned Banks: A LASSO and Elastic Net Approach to the "Volume-Value Divergence" Effect

Yue Shen

Renmin University of China, Beijing, China
syue23@ruc.edu.cn

Abstract. As a cornerstone of the capital market, bank stocks not only serve as the foundation of the financial ecosystem, but reflect market expectations for macroeconomic policies through their price fluctuations. This study focuses on stock price of six major state-owned banks in China from 2025 to 2026, screening for significant variables impacting daily closing prices. Since traditional OLS models often struggle with multicollinearity, this research uses LASSO model for variable selection and prediction, along with 10-fold validation and bootstrap examination to verify coefficient robustness. The results demonstrate that the study model can effectively explain stock price fluctuations using relevant variables and performs stably across different dimensions, which can provide empirical insights for both investors and regulators.

Keywords: Bank Stock Price, LASSO Regression, Bootstrap Sampling, Elastic Net

1. Introduction

As the cornerstone of China's financial system, major state-owned commercial banks plays a central role in monetary policy transmission [1, 2]. Therefore, investigating the determinants of their stock prices is critical for optimizing investment decisions. In 2025, the confluence of heightened geopolitical tensions, especially the resurgence of US-China trade frictions [3], and rising EPU [4] significantly amplified market fluctuations [5]. At the same time, the state's 500-billion RMB capital injection via special government bonds further shifted the banks' capital adequacy landscapes (State Council of the People's Republic of China). Such capital injections are consistent with policy tools that alleviate capital constraints and reshape bank behavior and liquidity creation [6].

In price forecasting, while OLS has historically been the dominant tool, recent scholarship stresses its limitations in handling high-dimensional financial data [7]. Specifically, the inherent multicollinearity among financial indicators often leads to unreliable parameter estimates in OLS models [8, 9]. To overcome these constraints, regularization techniques have emerged as a new paradigm for variable selection. LASSO regression can use an L1 penalty to automatically filter out irrelevant variables by shrinking their coefficients to zero, leaving only the most important predictors in the model [10]. This capability ensures predictive accuracy without sacrificing model

DOI: 10.54254/2754-1169/2026.BL34144

interpretability, providing a more robust empirical framework for deciphering complex market dynamics.

A key nuance in liquidity research lies in the distinction between trading volume and trading value. While often used interchangeably, recent market data reveal frequent "volume-value" divergences [10]. The empirical analysis part finds that for China's state-owned banks, Trading Volume (X1) and Trading Value (X2) exert diametrically opposite effects on stock prices. To address potential instability in LASSO caused by extreme multicollinearity, the study further implements Elastic Net regression (α=0.5) [11, 12]. The results consistently support the signs of these core factors and verify the "Volume-Value Divergence effect", suggesting that price appreciation is driven by institutional capital inflows rather than retail-driven market noise [13]. Validated through 10-fold cross-validation and Bootstrap sampling, these findings enrich market microstructure theory and offer regulators empirical insights into identifying "high-quality liquidity."

2. Model

2.1. Method limitation

When discussing the driving factors of stock prices of large state-owned banks, given the strong correlation between multiple indicators, the traditional ordinary least squares method (OLS) is prone to serious multicollinearity problems in estimation, resulting in increased model variance, and reduced prediction accuracy. In order to solve these problems, this study introduces the LASSO regression model that cand select variables according to data characteristics.

2.2. LASSO model

LASSO performs variable selection by shrinking less relevant coefficients to zero through L1 regularization. The formula of the target function is as follows:

$$\widehat{\beta}_{LASSO} = argmin\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\chi_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\} \quad (1)$$

In this formula, y_i stands for the dependent variable (daily closing price), β_i stands for the regression coefficient for each variable, and λ serves as a tuning parameter.

2.3. Elastic Net

While LASSO provides an efficient approach for variable selection by producing sparse models, it exhibits limitations when dealing with highly correlated predictors. To complement this and ensure the robustness of the empirical findings, this study further employs the Elastic Net regularization.

$$\widehat{\beta}_{ENet} = argmin\{\sum_{i=1}^{n}((y_i - \beta_0 - \sum_{j=1}^{p}\chi_{ij}\beta_j))^2 + \lambda[\tfrac{1-\alpha}{2}\sum_{j=1}^{p}\beta_j^2 + \alpha\sum_{j=1}^{p}|\beta_j|]\} \quad (2)$$

In this formula, α stands for the mixing coefficient of LASSO and ridge regression. This dual-model approach allows us not only to identify the most significant drivers of bank stock prices but also to maintain stable and reliable coefficient estimates for highly correlated economic indicators.

DOI: 10.54254/2754-1169/2026.BL34144

3. Results and Analysis

3.1. Data source and processing

This study selects six major state-owned commercial banks in China, including Industrial and Commercial Bank, Construction Bank, Bank of China, Agricultural Bank of Communications and Postal Savings Bank of China, and uses a large amount of stock data from February 2025 to February 2026. All financial data, daily transaction data and market data used in this study come from JoinQuant, a third-party quantitative platform. The platform provides standardized and high-quality data, and ensures the consistency and reliability of data through cross-verification with authoritative financial databases such as CSMAR.

The dependent variable of this study is the daily closing price. According to the asset pricing theory and the microstructure of the market, the independent variables are divided into four groups.

The first set of variables involves transaction dynamics (such as turnover ratio and transaction volume), which is mainly used to reflect market liquidity and the intensity of investors' expectations for macroeconomic policy adjustments. The second group is valuation indicators, including price-to-earnings ratio (P/E) and market-to-net ratio (P/B). The third group of variables focuses on profitability and growth indicators (such as ROA, ROE and revenue growth rate). Among them, the growth of income and profits reflects the resilience and growth potential of banks in the economic cycle. Finally, the Shanghai Composite Index is included to systematically control the impact of overall market trend on a particular stock, given that large-scale bank stocks are heavy-weighted components of the index. Additionally, market capitalization is also introduced to capture firm size, as firms of different scales tend to exhibit distinct risk-return characteristics (Fama-French size effect).

In the data preprocessing stage, extensive data cleaning was carried out, eliminating variables with missing values or empty values accounting for more than 30%, as well as redundant or low-quality variables. Among them, the "Gross Margin" was excluded from the analysis because it was always zero in the entire bank sample, which reflects the uniqueness of the interest income structure of the banking industry. After this rigorous screening process, 12 independent variables were finally retained for subsequent analysis. In order to eliminate the framework impact of different units of measurement - such as the significant numerical difference between market capitalization and turnover ratio - this study uses R language to standardize the Z fraction of all variables to ensure that the LASSO model treats each variable equally in the regularization process.

Table 1. Variables remained

Code	Variable Name
X1	Trading Volume
X2	Trading Value
X3	P/E Ratio
X4	P/B Ratio
X5	Market Capitalization
X	Turnover Ratio
X7	ROE
X8	ROA
X9	Net Profit Margin

DOI: 10.54254/2754-1169/2026.BL34144

Table 1. (continued)

X10	Revenue Year-on-Year
X11	Net Profit Year-on-Year
X12	Shanghai Composite Index

As shown in Figure 1, the variable correlation heat map shows that there are strong multiple collinearities between financial indicators (such as P/E Ratio and P/B Ratio), so this paper uses LASSO regression for analysis. Unlike the ordinary least squares method (OLS), LASSO can compress the coefficients of redundant variables to zero, thus improving the interpretability and stability of the model and achieving effective variable selection while retaining important features.

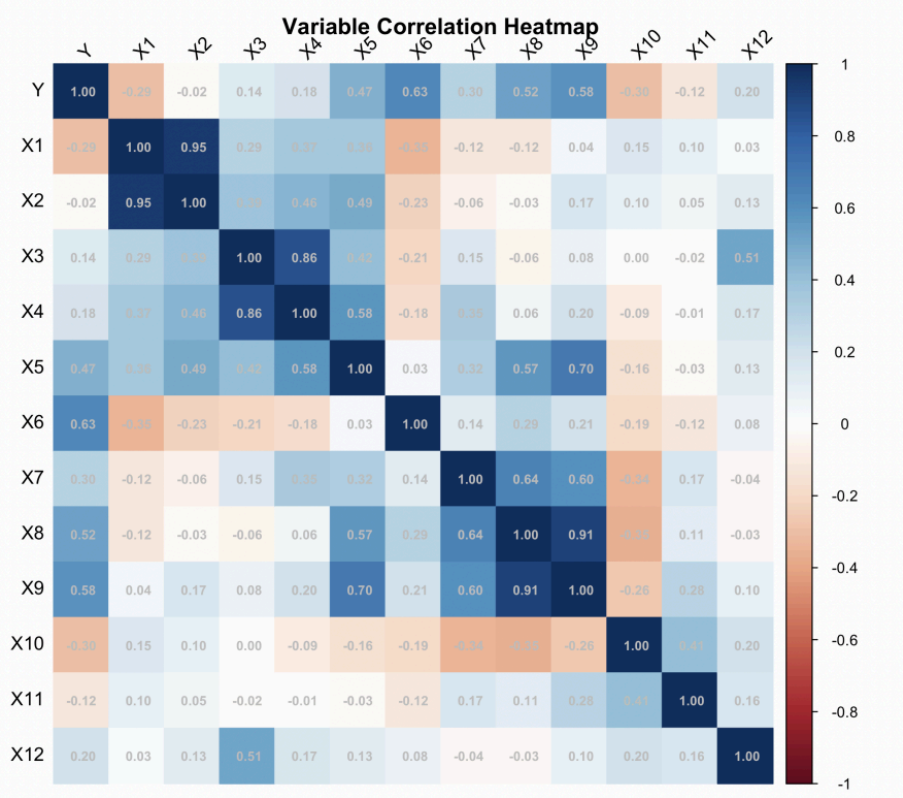

Figure 1. Variable correlation heatmap

It is worth noting that the period of this study (February 2025 to February 2026) coincides with the intensive adjustment period of China's banking regulatory policy. Relevant adjustments include: the Ministry of Finance issued special government bonds on March 31, 2025 to inject capital into state-owned banks to support their development; and the interest rate subsidy policy for eight specific enterprises in the service industry announced on August 14, 2025. To address this, the study first employs a baseline LASSO model focusing on market-driven variables to identify the core structure of price formation. To further account for potential policy shocks, policy-related variables are subsequently introduced in an extended model.

3.2. Analysis results

Before using LASSO model, the glmnet package in R language is applied to convert and standardize continuous independent variables and dependent variables to improve their distribution

DOI: 10.54254/2754-1169/2026.BL34144

characteristics.

As shown in Figure 2, the results of 10-fold cross-validation show that the Mean Square Error (MSE) reaches the minimum value at -log(λ) ≈ 4.5. Under the optimal regularization parameters, the model retains 10 variables with non-zero coefficients, achieving a good balance between model complexity and explanatory power.

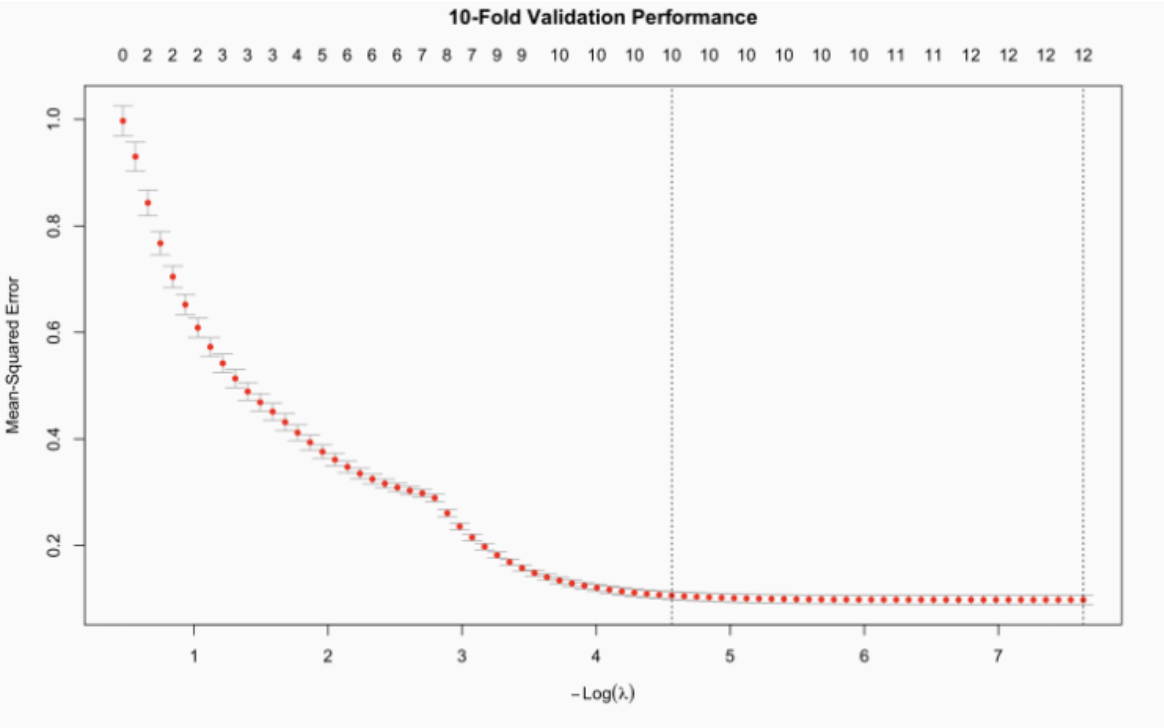

Figure 2. LASSO Parameter selection via 10-fold cross validation

As the regularization parameter varies, the coefficient paths indicate that X1 and X2 deviate from zero at an early stage, suggesting their strong explanatory power. The LASSO Coefficient Plot (Figure 3) further presents the direction and degree of influence of each variable. Notably, although the heat map shows a very high positive correlation (0.95) between X1 and X2, the LASSO model assigns them opposite coefficients: the positive impact of X2 is the most prominent, while X1 shows obvious negative effects. This phenomenon, a complex result of LASSO's feature selection under multicollinearity, reveals a "quality over quantity" feature of liquidity—that stock prices are no longer primarily driven by transaction frequency (X1), but rather by the scale of large capital inflows (X2). From an economic perspective, this "offsetting effect" suggests that after controlling for actual capital inflow, pure trading frequency may reflect short-term market noise or retail selling behavior, thus dragging down prices. Therefore, the rise of bank stocks in 2025–2026 depends mainly on large-scale institutional fund inflows ("high-quality liquidity") rather than frequent short-term transactions.

In addition to X2, the Net Profit Margin (X9) and the Turnover Ratio (X6) are also identified as secondary influencing factors. The positive impact of X9 shows that the market's focus is shifting from simple scale expansion to profit quality and operational efficiency, and tends to choose banks with strong cost control ability. At the same time, the importance of the change of ownership rate shows that the rise in the stock prices of banks in the six major countries is not a passive valuation repair, but depends on the improvement of market liquidity and investor participation.

DOI: 10.54254/2754-1169/2026.BL34144

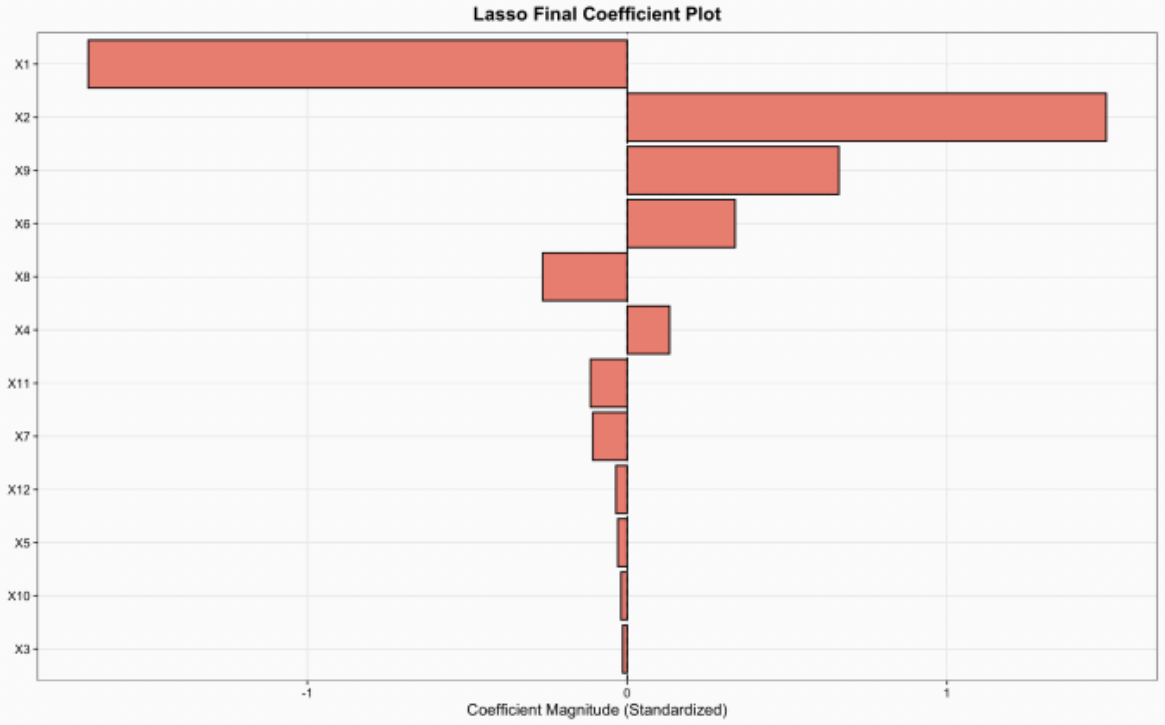

Figure 3. LASSO final coefficient plot

In summary, from 2025 to 2026, the rise in the stock price of state-owned banks is mainly driven by two factors: the improvement of market activity and the reallocation of institutional investor funds, thus forming a dual valuation structure.

It is worth noting that the Price-to-Earnings Ratio (X3) and the Revenue Year-on-Year (X10) have less impact in the LASSO model, reflecting that the pricing mechanism of China's banks is undergoing a fundamental transformation. The weakening of the influence of the P/E ratio indicates that investors do not only trade on short-term profit performance - because these indicators are more susceptible to the macroeconomic environment and policy factors. On the contrary, the focus of market attention is gradually shifting to structural robustness, showing that an asset-oriented valuation mechanism is being formed.

3.3. Robustness test

To ensure the reliability of the LASSO variable selection results and avoid conclusion bias caused by a single sample, the model is further validated with two methods: 10-fold cross validation and Bootstrap sampling.

10- fold Cross-Validation in Figure 4 illustrates that the linear model constructed with LASSO selection variables has excellent and stable prediction. The R2 values (blue line) remain consistently high, with an average of 0.9016 and slight fluctuations, indicating that the variable combination of 12 core factors can stably explain about 90.16% of the stock price variance even when the model is tested on unseen data.

The model also has high prediction accuracy with an average of RMSE (red line) at 0.3105. The absence of extreme high level in RMSE suggests the reliability and prediction ability of this processed model.

DOI: 10.54254/2754-1169/2026.BL34144

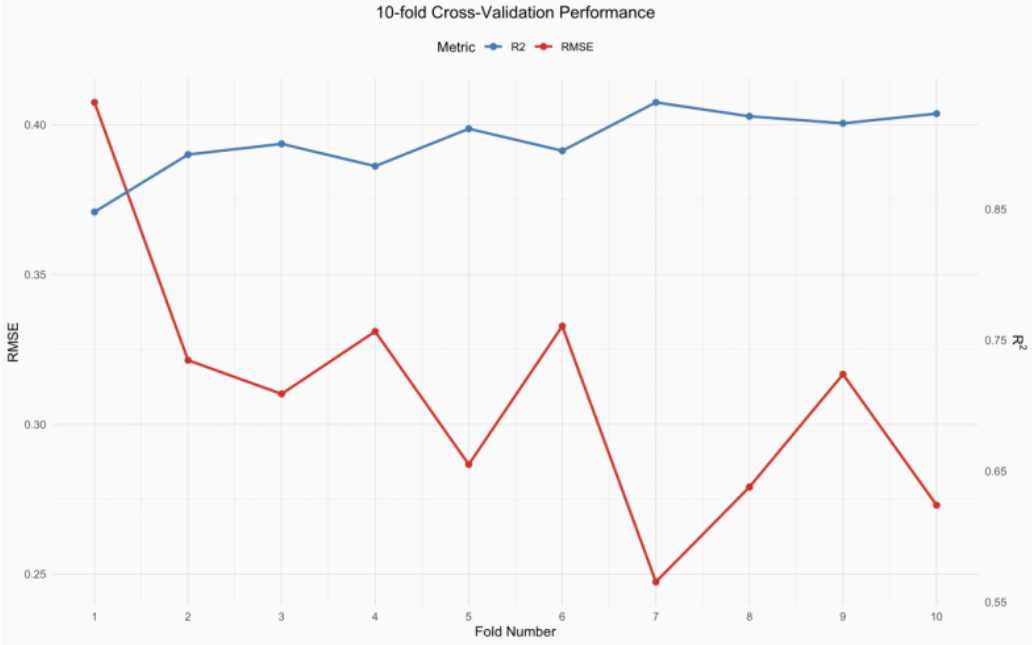

Figure 4. 10-fold cross-validation performance

The prediction ability of the model can be visually verified by comparing the predicted value of the model with the actual standardized stock price (Figure 5). Although there are slight fluctuations in the tail area of extremely undervalued (below -1.5) and extremely overvalued (above 1.0), the overall trend is still stable. These small fluctuations may partly reflect the short-term market disturbances caused by policy factors, but do not affect the judgment of the overall pricing trend.

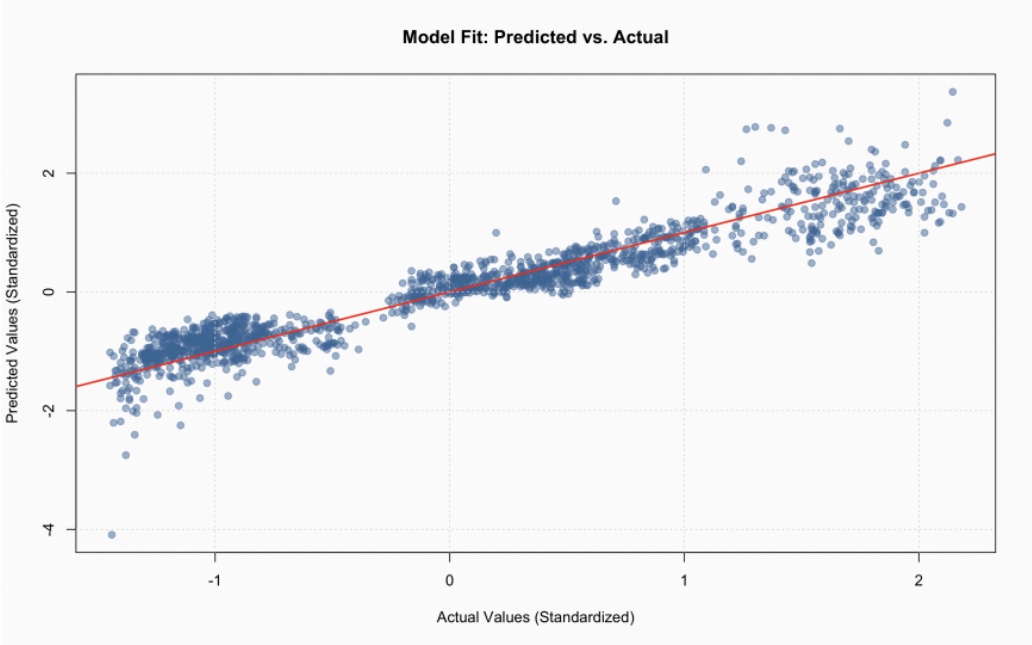

Figure 5. Predicted vs. actual

In view of the relatively small sample size of this study, the statistical significance and stability of the coefficient estimation of each variable were tested through 100 Bootstrap samplings. As shown

DOI: 10.54254/2754-1169/2026.BL34144

in Figure 6, the box line diagram of the Trading Value (X2) and the Trading Volume (X1) is located on both sides of the red zero reference line, of which the digits are 1.5 and -1.7 respectively. At the same time, the narrower quartile spacing (blue box) and the beard line away from the zero axis indicate that these two factors show good stability in different data sets.In contrast, some variables such as P/E Ratio (X3) and Market Capitalization (X5), the estimated coefficients either overlap with the zero axis or are extremely close to the zero axis, indicating that they are not significant in the LASSO model.

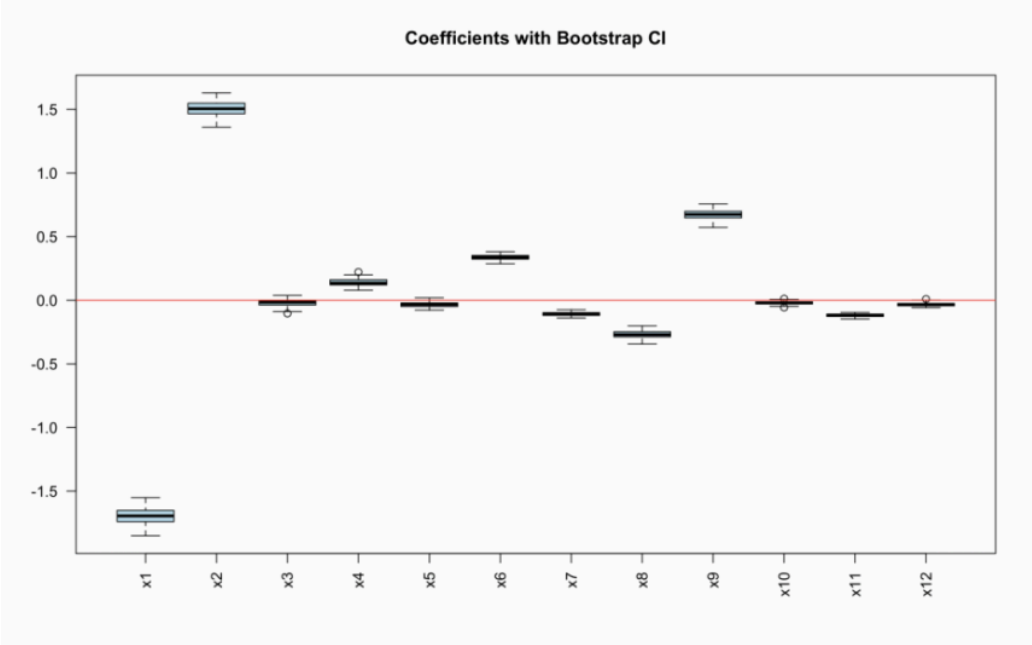

Figure 6. Coefficients with bootstrap

4. Further Analysis

4.1. Extended policy-augmented model

To further verify whether the identified "volume-value divergence" is a persistent statistical regularity or merely a byproduct of major regulatory shifts, this study further introduces two policy-related variables to control for exogenous policy shocks during the sample period.The results show that both policy variables are retained in the LASSO model with positive coefficients (0.2772 and 0.0979), indicating that policy shocks partly contribute to explaining stock price fluctuations. More importantly, after introducing policy controls, the coefficients of Trading Volume (X1) and Trading Value (X2) remain highly stable in both sign and magnitude. This suggests that the "Volume-Value Divergence" effect is not driven by policy-induced fluctuations, but reflects a more intrinsic market mechanism.

In addition, the MSE decreases from 0.1747 to 0.1705, showing a moderate improvement in model performance.

4.2. Model refinement with elastic net

In order to further examine the robustness of variable selection under high multicollinearity, this paper introduced Elastic Net Regression for comparative analysis. This study set α as 0.5 to strike a

DOI: 10.54254/2754-1169/2026.BL34144

balance between sparsity and group effects.

The MSE convergence path of the Elastic Net model exhibits a similar trajectory to that of the LASSO model (Figure 2), thus the redundant plot is omitted for brevity. Moreover, the regression result (Table 2) indicates that the signs for Trading Volume (X1) and Trading Value (X2) remain opposite, matching the LASSO results. Besides, it is worth noting that Market Capitalization (X5) still shows a remarkable positive effect.

On the whole, the results of Elastic Net support the LASSO model in terms of the direction and relative importance of core variables, indicating that the conclusion regarding liquidity structure has a certain degree of robustness.

Table 2. Regression coefficient

Variable	Coefficient
(Intercept)	-0.0128663176
X1	-0.298480584
X2	0.2927235902
X3	0.0620854563
X4	0.1324833205
X5	0.9356835179
X6	-0.0584916781
X7	0.1078790516
X8	-0.2160939272
X9	0.1106821798
X10	-0.0003337108
X11	0.0000000000
X12	-0.0072884128

5. Conclusion remarks

This study uses LASSO model, elastic net, 10-fold cross validation and Bootstrap sampling to find drivers of stock prices in China's state-owned banks under high-dimentional and complex environment, successfully find a model of twelve variables with strong explanatory power and stability. Besides, it is noteworthy that the study identifies the "Volume-Value Divergence" between the two relevant driving forces, trading value and trading volume despite policy impact, which is quite counterintuitive.

However, it should be mentioned that under the condition of strong correlation variables, the size of the coefficient may still be affected by the model setting, so the relevant economic explanation should remain cautious. Besides, the sample period is short in this study. For future research, more casual inference methods and dynamic modeling frameworks can be induced to make more accurate conclusions.

References

[1] Mo, Y., Sun, W., Ding, Y., & Wang, L. (2025). Divergent impacts of quantity versus price-based monetary policies on banking systemic risk: Evidence from China. PloS one, 20(5), e0322709. https://doi.org/10.1371/journal.pone.0322

DOI: 10.54254/2754-1169/2026.BL34144

[2] Huang, Min & Jiang, Hai, 2025. "Shadow banking, macroprudential policy and banks' systemic risk, "Research in International Business and Finance, Elsevier, vol. 77(PB).
[3] Chen, Y., Kang, R., & Liu, L. (2026). Geopolitical tensions and systemic vulnerability in the banking sector: Evidence from China. Finance Research Letters, 94, 109679. https: //doi.org/10.1016/j.frl.2026.109679
[4] Feng, Yue. (2025). Measuring Political Risk in the Stock Market : An Empirical Study of Economic Policy Uncertainty and the Volatility Index. Advances in Economics, Management and Political Sciences. 216. 129-143. 10.54254/2754-1169/2025.GL27009.
[5] Zhao, X., Hu, Q., Song, Y., & Huang, J. (2025).Systemic risk spillovers incorporating investor sentiment: Evidence from an improved TENET analysis.https: //doi.org/10.1016/j.econmod.2025.107184
[6] Weifeng Yu, Huiqi Wu, Zhihui Song, Capital constraint relief and liquidity creation in small and medium-sized banks: Evidence from a quasi-natural experiment in China, International Review of Economics & Finance, Volume 103, 2025, 104405, ISSN 1059-0560, https: //doi.org/10.1016/j.iref.2025.104405.
[7] Ogunbona, Babafemi & Balogun, Folorunsho & Famuagun, Kayode. (2025). Solving Multicolinearity Problem in a Linear Regression: A Comparative Study of Ordinary Least Squares and Partial Least Squares Regression. 1. 66-75. 10.5281/zenodo.15556948.
[8] Huynh, T.T., Khoa, B.T. (2026). Predictive Analytics in Stock Markets: An Integrated Approach Using OLS and Lasso Regression. In: Motahhir, S., Bossoufi, B., Guerrero, J.M. (eds) Digital Technologies and Applications. ICDTA 2025. Lecture Notes in Networks and Systems, vol 1639. Springer, Cham. https: //doi.org/10.1007/978-3-032-07718-9_43
[9] Huynh, T.T., Khoa, B.T. (2025). Determinants of Vietnam's VN-Index: Analyzing the Interplay of Global and Domestic Factors Using Machine Learning. In: Yaseen, S.G. (eds) Applied Artificial Intelligence in Business. Studies in Systems, Decision and Control, vol 597. Springer, Cham. https: //doi.org/10.1007/978-3-031-90271-0_2
[10] Kesse, Godfred Ahenkroa, "Variable Selection using Lasso Regression" (2025). Data Science and Data Mining. 28.https: //stars.library.ucf.edu/data-science-mining/28
[11] Ahiduzzaman, Md, "Comparative Analysis of LASSO, Ridge, and Elastic Net for Variable Selection in High-Dimensional Maize Data" (2025). Data Science and Data Mining. 49.https: //stars.library.ucf.edu/data-science-mining/49
[12] Hui Zou, Trevor Hastie, Regularization and Variable Selection Via the Elastic Net, Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 67, Issue 2, April 2005, Pages 301–320, https: //doi.org/10.1111/j.1467-9868.2005.00503.x
[13] Cui J, Wei Q, Gao X. How Retail vs. Institutional Investor Sentiment Differ in Affecting Chinese Stock Returns? Journal of Risk and Financial Management. 2025; 18(2): 95. https: //doi.org/10.3390/jrfm18020095