IDENTIFYING KEY FEATURES AFECTING BANK LIQUIDITY RISK: A MACHINE LEARNING REGRESSION APPROACH

PhD. Vu Thi Thu Huong* - MBA. Nguyen Thu Ha* - Tran Thi Thuy Huyen*
Pham Ngoc Khue* - Hoang Thao Nhi* - Nguyen Thi Lan Huong*

Abstract: ***This study investigates the determinants of liquidity risk by applying advanced machine learning techniques to overcome the limitations of traditional econometric methods. Using data collected from the financial statements of 31 Vietnamese commercial banks over the period 2009-2024, the study implements ten regression algorithms classified into three main groups: regularized linear models (Lasso, Ridge, Elastic Net) compared with standard linear regression; tree-based regression models (Decision Tree, Random Forest); and gradient boosting models (Gradient Boosting, XGBoost, LightGBM, and CatBoost). The empirical results quantify and rank the importance of key financial features affecting liquidity risk, thereby providing further evidence of the nonlinear nature of banking data in Vietnam. These findings contribute to the development of quantitative tools that support bank managers and regulators in designing risk monitoring frameworks and early warning systems for liquidity instability.***

• Keywords: *machine learning, regression, liquidity risk, commercial banks, Vietnam.*

Date of receipt: 25th Feb., 2026
Date of delivery revision: 04th Mar., 2026
Date of receipt revision: 20th Mar., 2026
Date of approval: 26th Apr., 2026

DOI: https://doi.org/10.71374/jfar.v26.i3.22

1. Introduction

Liquidity risk is one of the most critical risks affecting the solvency, stability, and overall safety of the commercial banking system. In the context of deep financial integration and increasing exposure to macroeconomic shocks, identifying and managing liquidity risk has become increasingly essential, not only for individual banks but also for the stability of the entire financial system.

Previous studies on liquidity risk determinants have primarily relied on traditional econometric approaches such as linear regression, Ridge, Lasso, and Elastic Net. However, these methods often assume linear relationships between dependent and explanatory variables and face limitations in capturing complex interactions and nonlinear patterns in the data. In practice, banking operations are influenced by multiple financial and macroeconomic factors with inherently nonlinear relationships, making linear models insufficient to fully capture the nature of liquidity risk. Empirical findings by Tran Thi Thanh Tu and Nguyen Thi Hoang Anh (2019) highlight that liquidity risk in Vietnamese banks is multidimensional, requiring flexible risk identification frameworks to adapt to the rapidly evolving financial environment.

The advancement of machine learning methods has introduced new approaches to financial data analysis. Regression algorithms such as Decision Tree, Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost enable the modeling of nonlinear relationships and effectively capture interactions among variables. In addition, model interpretability techniques such as SHAP allow for the quantification of each input variable's contribution, thereby enhancing the transparency of machine learning models.

Although these methods have been widely applied internationally, empirical studies in Vietnam remain limited, particularly in integrating model performance evaluation with feature importance analysis. This gap highlights the need for empirical research to validate the effectiveness of machine learning regression algorithms in the context of Vietnamese commercial banks.

Based on a dataset of financial statements from 31 banks during the period 2009-2024, this study applies a range of regression models, including Linear Regression, Ridge, Lasso, Elastic Net, Decision Tree, Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost. The combination of linear and nonlinear models enables a comprehensive evaluation of liquidity risk modeling under different data conditions. The empirical results indicate that nonlinear models, particularly XGBoost, outperform

** Thuongmai University, Hanoi; email: huong.vtt@tmu.edu.vn (Corresponding author); ha.qttc@tmu.edu.vn; thuyhuyen665@gmail.com; khuepham778@gmail.com; nhih6031@gmail.com; ntl.huong28@gmail.com*

linear models, suggesting their superior ability to capture complex relationships between financial variables and liquidity risk. Furthermore, the study employs SHAP to quantify the contribution of each input feature to the prediction outcomes.

This study contributes in three main aspects. First, methodologically, it demonstrates the effectiveness of machine learning algorithms such as XGBoost, LightGBM, and CatBoost in modeling liquidity risk. Second, empirically, it provides evidence specific to the context of Vietnamese commercial banks. Third, practically, the findings offer a basis for proposing liquidity risk management strategies grounded in the identified key determinants.

The paper is structured into five sections. Following the introduction, Section 2 presents the literature review, including the theoretical background on liquidity risk and the regression algorithms employed. Section 3 describes the data and research methodology. Section 4 reports the empirical results and provides in-depth discussions on the underlying risk structure. Finally, Section 5 concludes the study, discusses policy implications, and suggests directions for future research.

2. Literature Review

2.1. Bank Liquidity Risk

2.1.1. Theoretical Foundations of Bank Liquidity Risk

Commercial banks are financial intermediaries that perform core functions, including deposit mobilization, credit provision, and payment services. Their operational structure is fundamentally based on maturity transformation, whereby short-term, highly liquid liabilities are used to finance long-term, less liquid assets (Nguyen Van Tien, 2015). While this mechanism enhances profitability, it inherently creates mismatches in maturity and cash flows, thereby giving rise to liquidity risk as an intrinsic feature of banking activities.

According to the Basel Committee on Banking Supervision (2008), liquidity risk refers to a bank's inability to meet its financial obligations as they fall due without incurring unacceptable losses or jeopardizing its financial position. Thus, liquidity risk reflects not only short-term solvency but also the overall financial sustainability of the institution.

The theoretical foundation explaining the formation of liquidity risk is rooted in the model of Diamond and Dybvig (1983), which demonstrates that the liquidity creation function of banks is inherently associated with the risk of bank runs driven by changes in depositor expectations. Even solvent banks may face liquidity shortages that disrupt their operations. Therefore, liquidity management represents an optimization problem involving a trade-off between safety and efficiency: excessive liquidity reserves reduce risk but incur opportunity costs, whereas insufficient reserves increase vulnerability to sudden withdrawal shocks (Dietrich & Wanzenried, 2011).

From a microeconomic perspective, liquidity risk is strongly influenced by maturity mismatches and information asymmetry. According to Stiglitz and Weiss (1981), imperfect information regarding asset quality may lead to adverse selection and credit rationing, thereby constraining access to interbank funding and exacerbating liquidity stress. From a macroeconomic perspective, liquidity risk is closely linked to financial cycles and market sentiment. Minsky's (1992) financial instability hypothesis suggests that risks accumulate during periods of economic expansion as institutions increase leverage and adopt riskier financial structures. Meanwhile, Keynes' (1936) liquidity preference theory emphasizes the role of interest rates and expectations in shaping money-holding behavior, thereby generating exogenous liquidity shocks to the banking system.

Overall, liquidity risk is a multidimensional phenomenon driven by both internal and external factors. Empirically, these factors can be grouped into three categories: (i) balance sheet characteristics reflecting maturity transformation and liquidity buffers, such as the loan-to-deposit ratio (LDR), liquid asset ratio (AR), and asset structure (AC); (ii) financial and governance factors, grounded in trade-off and financial distress cost theories (Modigliani & Miller, 1963), including equity capital (EAR), profitability (ROE, NIM), and operational efficiency (CIR); and (iii) market and macroeconomic conditions, including credit growth, inflation, public confidence, and systemic shocks such as financial crises or pandemics. The nonlinear and state-dependent nature of these relationships necessitates the use of advanced analytical approaches, particularly machine learning, to enhance the identification and prediction of liquidity risk in the banking system.

2.1.2. Measurement of Liquidity Risk

In banking risk management practice, measuring liquidity risk requires a combination of balance sheet indicators and composite risk measures based on earnings volatility. One of the most widely used indicators is the loan-to-deposit ratio (LDR).

According to international standards and Vietnam's Circular No. 22/2019/TT-NHNN, LDR is designed to control the extent to which short-term funding is used for long-term lending. It is calculated as:

$$LDR = \frac{\text{Total customer loans}}{\text{Total customer deposits}} \times 100\%$$

Theoretically, a high LDR reflects an aggressive lending strategy and efficient capital utilization; however, as argued by Hannan and Hanweck (1988), it also increases banks' sensitivity to sudden withdrawal demands. In Vietnam, a threshold of 85% is generally considered a prudential limit to balance profitability and liquidity resilience. Nevertheless, static indicators such as LDR are limited in capturing the dynamic nature of insolvency risk arising from asset deterioration or income volatility.

To address these limitations, this study employs the Z-score as a comprehensive measure of financial stability and insolvency risk. Theoretically, the Z-score represents the distance to default, measuring how many standard deviations a bank's return must decline before its equity is fully depleted (Boyd & Runkle, 1993). It is defined as:

$$Z = \frac{ROA + CAP}{\sigma(ROA)}$$

where ROA is return on assets, CAP is the equity-to-asset ratio, and σ(ROA) is the standard deviation of ROA over a given period. A higher Z-score indicates greater financial stability and lower risk. Due to the right-skewed and nonlinear nature of the original Z-score, this study applies a logarithmic transformation $\ln_z = -\ln(Z)$ to normalize the distribution, reduce the impact of outliers, and align the measure positively with risk (Lepetit & Strobel, 2013), thereby improving compatibility with machine learning algorithms.

2.2. Regression Algorithms in Liquidity Risk Analysis

The transition from traditional econometric models to machine learning approaches is necessary to address the nonlinearities and complex multivariate interactions inherent in banking data. In Vietnam, although liquidity risk has attracted considerable research attention, most empirical studies still rely on linear frameworks such as OLS, FEM, and GMM. These approaches face limitations in capturing nonlinear relationships and lack efficient mechanisms for automatic feature selection.

To overcome these limitations, this study implements a set of ten advanced regression algorithms combined with SHAP-based model interpretability techniques (Lundberg & Lee, 2017), providing both high predictive accuracy and economic interpretability.

The algorithms are classified into three main groups. The first group includes linear regression and regularization techniques such as Lasso (Tibshirani, 1996), Ridge (Hoerl & Kennard, 1970), and Elastic Net, which help mitigate multicollinearity and overfitting through L1 and L2 penalties. The second group comprises tree-based models, including Decision Tree and Random Forest. In particular, Random Forest employs bootstrap aggregation (bagging) to enhance model stability and generalization performance, especially in noisy panel data.

The third group, which constitutes the core of this study, includes gradient boosting models such as Gradient Boosting Regressor (Friedman, 2001), XGBoost (Chen & Guestrin, 2016), LightGBM, and CatBoost. These algorithms follow a sequential learning process in which each weak learner iteratively corrects the errors of its predecessors. Notably, XGBoost incorporates regularization and advanced optimization techniques, enabling efficient modeling of nonlinear interactions among 21 financial and macroeconomic variables. The use of multiple algorithms not only improves predictive performance but also facilitates a deeper understanding of the underlying structure of liquidity risk through cross-method comparison, thereby strengthening the empirical foundation for liquidity risk analysis in Vietnamese commercial banks.

3. Methodology

3.1. Data and Variables

This study employs a secondary dataset collected from the annual financial statements of 31 commercial banks in Vietnam over the period 2009-2024. Based on the raw data, the authors construct the variables used in the empirical model. Prior to analysis, the dataset is screened to remove observations with missing values in any variable, ensuring the integrity and consistency of the input data.

The dependent variable, representing liquidity risk, is measured using the Z-score as described in Section 2.1.1.

The independent variables (features) are grouped to capture different dimensions of banking operations. Specifically, the asset structure and credit quality group includes indicators related to loan portfolios and credit risk exposure. The profitability and operational efficiency group reflects banks' performance and cost management. Additionally, variables capturing bank characteristics account for differences in size and lifecycle across institutions. The study also incorporates macroeconomic and market variables to reflect the influence of external conditions on liquidity risk. A detailed description of the 21 features is presented in Table 1.

Table 1. Definition of independent variables in the research model

No.	Variable Name	Explain the variables
Liquidity		
1	AR	Ratio of liquid assets to total assets
2	LDR	Loan-to-deposit ratio
Capital Adequacy		
3	EAR	Ratio of equity to total assets
4	LLR	Loan loss reserve ratio = total risk provisions / total liabilities
5	LOAN_TA	Ratio of liabilities to total assets
6	Credit_G	Credit growth
7	NIM	Net Interest Margin = net interest income / interest-earning assets
8	TSCD_VCSH	Fixed assets to equity
Operating Efficiency		
9	ROE	Return on Equity = net profit after tax / total equity
10	CIR	Cost-to-Income Ratio = total operating expenses / operating income
11	OER	Operating Expense Ratio = total operating expenses / revenue
12	RTR	Receivables-to-total assets ratio
Bank Characteristics		
13	AGE	Age of the bank
14	SIZE	SIZE = ln (Total assets)
Macroeconomic Variables		
15	CPI	Consumer Price Index
16	Covid	Binary variable, equal to 1 in 2020 and 2021, 0 in other years
New Variables		
18	Trust	Depositor confidence index = customer deposit growth / total customer deposits
18	Liquidity	Liquidity growth rate = net cash from operations / total cash at beginning of year
19	AC	Asset conversion ratio = customer loans / investment securities
20	Interbank	Interbank funding dependency ratio = deposits and borrowings from other credit institutions / total liabilities
21	Sen	Liquidity risk sensitivity index = Loan-to-deposit ratio × Consumer Price Index

Source: Authors' research

3.2. Implementation Procedure of Machine Learning Regression Algorithms

Step 1: Data preparation

The initial dataset consists of 487 observations with 22 variables, including one dependent variable (ln_z) and 21 independent variables. After removing observations with missing values, the final dataset contains 344 observations. The data are then split into training and testing sets with an 80/20 ratio, corresponding to 299 and 75 observations, respectively.

Step 2: Model selection and training

Based on the processed dataset, the study estimates regression models across three main groups to compare their ability to model liquidity risk. The first group includes linear models (Linear Regression, Ridge, Lasso, Elastic Net), representing traditional approaches that assume linear relationships. The second group consists of tree-based models (Decision Tree, Random Forest, Gradient Boosting), which are capable of capturing nonlinear relationships and variable interactions. The third group includes advanced boosting models (XGBoost, LightGBM, and CatBoost), designed to more effectively exploit complex data structures and improve predictive performance.

This classification enables a comparative evaluation of different methodological approaches and identifies the most suitable models for analyzing bank liquidity risk.

Step 3: Model performance evaluation

Model performance is assessed using R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), which measure explanatory power and prediction accuracy. The evaluation is conducted across all model groups to examine differences in their ability to capture variations in liquidity risk, particularly between linear and nonlinear approaches.

Step 4: Model selection and result interpretation

Based on the evaluation metrics, the best-performing model is selected for further analysis. The study then applies SHAP (SHapley Additive exPlanations) to quantify the contribution of each input feature, thereby identifying the most important determinants of liquidity risk.

4. Results and Discussion

4.1. Regression results analysis

4.1.1. Linear models

Prior to estimation, the dataset was carefully preprocessed to ensure the quality of model inputs. First, observations with missing values were removed to avoid bias in the estimation. Next, outliers were identified for each variable individually using the interquartile range (IQR) method. To mitigate the influence of extreme values without discarding valuable information, a winsorization technique was applied, replacing outliers with the 1st and 99th percentile values.

The data were then standardized using the StandardScaler method to bring all variables onto a comparable scale, ensuring fairness in the estimation of models that incorporate regularization. Finally, linear regression models were estimated, with the regularization parameter (alpha) optimized through 5-fold cross-validation. The estimation results are presented in Table 2.

The comparison of performance metrics in Table 2 indicates that the Lasso model is the most appropriate among the four models. Although the R^2 on the test set is low for all models, suggesting limited explanatory power, Lasso achieves the lowest RMSE and MAE, and slightly higher CV R^2 than OLS and Ridge, reflecting better generalization ability. Compared to Elastic Net (which is nearly equivalent to Lasso in this case due to l1_ratio = 1), Lasso is more parsimonious while maintaining predictive performance, making it the most reasonable choice.

Table 2. Estimation results and performance of linear regression models

Variables	OLS	Lasso	Ridge	Elastic Net
AR	-2.2248	-1.2351	-2.0804	-1.2351
LDR	0.4271	0.0000	0.0908	0.0000
EAR	-0.9562	0.5498	0.0001	0.5498
LLR	-2.2582	0.0000	-1.2842	0.0000
LOAN_TA	-3.1220	-1.9761	-2.5669	-1.9761
Credit_G	-0.8290	-0.4007	-0.7046	-0.4007
NIM	-3.2373	0.0000	-0.8030	0.0000
ROE	7.4731	6.1649	6.7336	6.1649
CIR	0.2274	0.1231	0.1930	0.1231
OER	-6.1297	-4.8911	-5.7369	-4.8911
RTR	-1.5315	-1.3814	-1.9442	-1.3814
SIZE	-0.1633	-0.1195	-0.1287	-0.1195
CPI	0.1019	0.0269	0.0629	0.0269
trust	1.1534	0.6597	1.0487	0.6597
Liquidity	0.0167	0.0000	0.0090	0.0000
AC	0.0122	0.0000	0.0120	0.0000
interbank	0.4195	0.0000	0.5253	0.0000
TSCD_VCSH	1.0922	0.5866	0.9703	0.5866
sen	-0.0677	0.0000	-0.0288	0.0000
COVID	-0.2324	-0.1804	-0.2205	-0.1804
AGE	-0.0065	-0.0027	-0.0062	-0.0027
Intercept	0.4968	-1.0131	-0.4087	-1.0131
Train_R²	0.3141	0.2913	0.3110	0.2913
Test_R²	0.0129	0.0253	0.0073	0.0253
CV_R²_mean	0.1464	0.1799	0.1682	0.1799
CV_R²_std	0.1333	0.1008	0.1212	0.1008
RMSE	0.9878	0.9816	0.9906	0.9816
MAE	0.7992	0.7728	0.7914	0.7728
MAPE_%	21.1519	20.7052	21.0255	20.7052
Alpha		0.0266	10.0000	0.0266

Source: Authors' estimations using Python

Examining the coefficients in the Lasso model shows that the feature selection mechanism works effectively by shrinking some coefficients to zero. This makes the model more compact and easier to interpret while preserving its predictive capability.

Specifically, ROE has the strongest positive effect, playing a key role in explaining the dependent variable. In contrast, OER has the largest negative impact, indicating that lower operational efficiency reduces performance. Variables such as LOAN_TA, RTR, and AR also have moderate negative effects, while trust, TSCD_VCSH, and EAR exhibit smaller positive impacts. Meanwhile, CPI, AGE, and CIR are nearly insignificant, with coefficients close to zero. This confirms that the Lasso model effectively selects relevant variables, retaining the most important predictors and eliminating less informative ones, thereby improving both interpretability and predictive performance.

In addition, variables such as LDR, LLR, NIM, Liquidity, AC, interbank, and sen are shrunk to zero in the Lasso model, suggesting that they provide little predictive value for liquidity risk in a linear framework.

Although Lasso outperforms OLS, Ridge, and Elastic Net in both feature selection and prediction, the overall performance of linear models remains relatively low. This highlights the limitation of the linearity assumption in the context of bank liquidity risk analysis. To address this limitation, the next steps will focus on: (i) applying nonlinear machine learning models to improve predictive performance; (ii) using SHAP (SHapley Additive exPlanations) to interpret results from nonlinear models and maintain transparency; and (iii) identifying key features that drive bank liquidity risk.

4.1.2. Comparison of Regression Algorithm Performance

To achieve an objective assessment of liquidity risk detection capability in the Vietnamese commercial banking system, the study synthesized and compared results from 10 experimental algorithms. The comparison focused on key criteria including explanatory power (R^2), prediction error magnitude (MAE, RMSE), and stability through cross-validation (CV R^2).

Table 3: Performance Metrics of Regression Models

Model	Train R²	Test R²	CV R² (5-fold)	Test RMSE	Test MAE
Linear Regression Models					
Linear Regression	0.3141	0.0129	0.1464	0.9878	0.7992
Ridge	0.3110	0.0073	0.1682	0.9906	0.7914
ElasticNet	0.2913	0.0253	0.1799	0.9816	0.7728
Lasso	0.2913	0.0253	0.1799	0.9816	0.7728
Tree-based Regression Models					
Decision Tree	0.9347	-0.8989	-0.4610	1.3948	1.009
Random Forest	0.8036	-0.0130	0.1769	1.0187	0.766
Boosting Regression Models					
XGBoost	0.9893	0.1040	0.1308	0.9581	0.718
LightGBM	0.8978	0.0330	0.1871	0.9954	0.771
CatBoost	0.8118	0.0180	0.2603	1.0031	0.752
Gradient Boosting	0.9954	-0.0050	0.1421	1.0150	0.809

Source: Authors' calculations

From the performance evaluation of the ten regression models in the Table 3, a clear divergence can be observed between linear models and nonlinear machine learning approaches. Linear models exhibit very low Test R^2 values (≈0.01-0.03), indicating limited explanatory power, while Decision Tree and Gradient Boosting suffer from severe overfitting, as reflected in negative Test R^2 values. Random Forest shows some improvement in stability but still fails to achieve strong predictive performance.

In contrast, boosting models demonstrate clear advantages. XGBoost achieves the highest Test R^2 (0.1040) and the lowest RMSE and MAE, indicating superior predictive accuracy. LightGBM maintains stable performance with relatively high CV R^2 (0.1871) and low prediction errors, while CatBoost stands out in terms of stability, with the highest CV R^2 (0.2603), reflecting strong generalization ability.

Based on a balanced consideration of predictive accuracy, interpretability, and stability, the three most suitable models for analyzing bank liquidity risk are XGBoost, LightGBM, and CatBoost.

4.2. Identifying important features using SHAP

After establishing the predictive performance and stability of nonlinear algorithms, this study applies the SHAP (SHapley Additive exPlanations) technique, an advanced approach grounded in cooperative game theory, to overcome the black-box limitations of machine learning. This method enables the quantification of each feature's contribution and the identification of nonlinear impact thresholds within Vietnamese banking data. The corresponding Feature Importance and Summary Plots are presented in the Figure 1-3.

Figure 1: SHAP analysis results for the XGBoost model

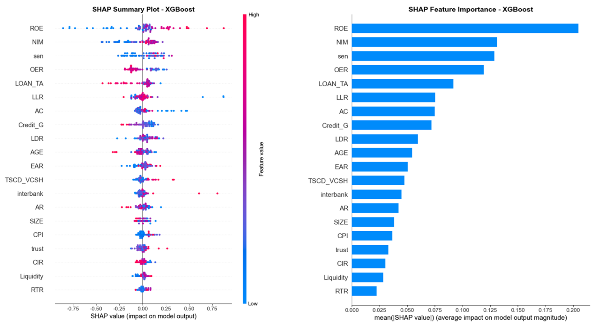


Figure 2: SHAP analysis results for the LightGBM model

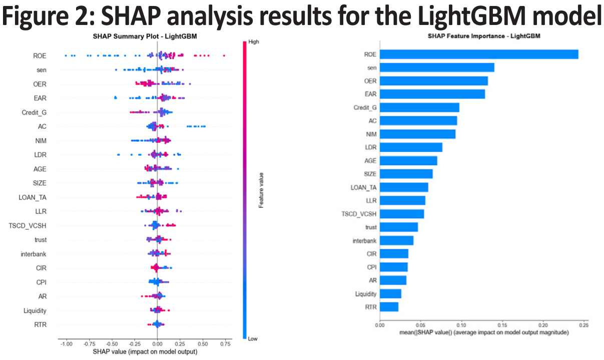
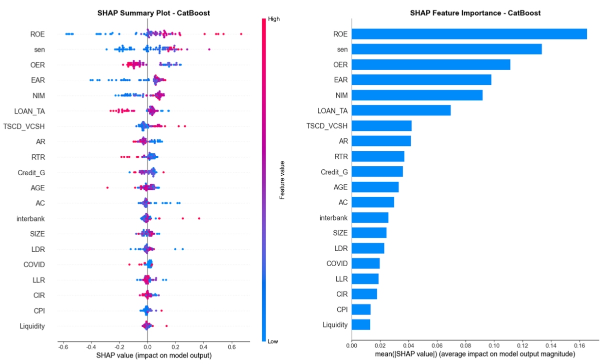


Figure 3: SHAP analysis results for the CatBoost model



Comparative analysis from the experimental plots reveals a high degree of consistency among the algorithms regarding the system of determinants of liquidity risk. Based on training logs and average rank scores, the study establishes the Top 10 most important features summarized as follows:

Table 4: Top 10 Most Important Features (Synthesized from 3 SHAP Models)

No	Features	Mean rank	SHAP (XGBoost)	SHAP (LightGBM)	SHAP (CatBoost)
1	ROE	1.00	0.2047	0.2434	0.1653
2	Sen	2.33	0.1287	0.1404	0.1336
3	OER	3.33	0.1192	0.1325	0.1115
4	NIM	4.67	0.1310	0.0928	0.0918
5	EAR	6.33	0.0504	0.1289	0.0980
6	LOAN_TA	7.33	0.0918	0.0592	0.0695
7	Credit_G	7.67	0.0720	0.0975	0.0359
8	AC	8.33	0.0748	0.0948	0.0298
9	AGE	10.00	0.0543	0.0704	0.0330
10	LDR	10.67	0.0597	0.0765	0.0229

Source: Authors' calculations

The main findings can be summarized by importance ranking and direction of impact as follows:

(i) Operational efficiency (ROE, OER): ROE is the most important factor with a negative relationship, indicating that profitability plays a key role in maintaining liquidity stability. In contrast, OER has a positive impact, suggesting that higher operating costs increase liquidity pressure.

(ii) Risk sensitivity variables (Sen, AC): Sen (LDR * CPI) ranks second, showing that liquidity risk is jointly influenced by inflationary pressure and funding intensity. AC reflects asset flexibility, where a less liquid asset structure increases risk.

(iii) Credit risk and liability structure (NIM, LOAN_TA, Credit_G): NIM supports liquidity through improved earning asset quality, while LOAN_TA and Credit_G signal risks associated with rapid credit expansion and reliance on external funding.

(iv) Capital adequacy and bank characteristics (EAR, AGE, LDR): EAR and LDR capture fundamental financial strength, while AGE highlights the role of experience in liquidity management.

Overall, the results emphasize that bank liquidity risk is primarily driven by internal management capacity and operational efficiency, with macroeconomic factors potentially amplifying effects in a nonlinear manner.

5. Conclusion and Implications

5.1. Conclusions

The study evaluated factors affecting the liquidity risk of the Vietnamese commercial banking system during the 2009-2024 period by applying 10 different machine learning algorithms. Empirical results show that models in the Boosting group, particularly XGBoost, LightGBM, and CatBoost, have superior predictive capability and data explanatory power compared to traditional linear regression models. The application of these algorithms enables the detection of multidimensional interactions among financial indicators, thereby providing highly reliable empirical evidence on system safety.

Based on the SHAP explanation mechanism, the study identified the ranking and impact direction of key features. Results indicate that operational efficiency and internal profitability are the primary factors directly governing liquidity risk. Among these, Return on Equity (ROE) plays the most significant explanatory role in the model structure. Subsequent features including the Sen index, operating expense ratio (OER), earnings asset ratio (EAR), and net interest margin (NIM) constitute the top five factors with the most profound influence on the dependent variable ln_z. Regarding impact direction, ROE, Sen, EAR, and NIM exhibit a positive relationship with the liquidity safety index, while OER shows a clear negative impact.

The research results emphasize the need to innovate liquidity risk supervision methods through modern analytical tools. Early detection of fluctuations in core financial features such as ROE and EAR enables managers and regulators to establish proactive response scenarios tailored to the actual operations of each banking institution. These findings not only add empirical evidence on the nonlinear nature of financial data in Vietnam but also provide an important quantitative foundation for risk management.

5.2. Policy and management implications

Based on the results obtained, the study proposes several specific policy and management implications for relevant stakeholders as follows:

For state management agencies and the State Bank of Vietnam, there is a need to strengthen supervision of indicators reflecting short-term liquidity fulfillment capability and capital adequacy levels. Empirical results show that variables related to asset-liability structure and equity play important roles in explaining risk fluctuations. Therefore, improving the supervisory framework to focus on capital quality, controlling credit risk, and managing maturity mismatches is necessary to limit liquidity risk accumulation in the system. Additionally, integrating data analysis tools and forecasting models into the supervision process could enhance early warning capabilities.

For commercial banks, liquidity risk management needs to be approached comprehensively. Banks should maintain reasonable equity ratios, strictly control asset quality and non-performing loans, while optimizing the maturity structure between assets and liabilities. Enhancing profitability also helps strengthen internal resources, creating financial buffers to absorb liquidity shocks. Furthermore, applying data analysis tools and machine learning models in internal monitoring can support early identification of rising risk trends, rather than only reacting after risks materialize.

For the interbank market, the research findings imply that connectivity and interdependence among credit institutions may amplify liquidity risks under adverse conditions. Therefore, it is necessary to enhance information transparency, standardize short-term borrowing and lending mechanisms, and develop early warning systems for abnormal fluctuations in the money market, thereby limiting the risk of liquidity contagion among credit institutions when adverse shocks occur.

Limitations and future research

This study relies mainly on quantitative annual data and does not fully capture qualitative factors or short-term dynamics. Future research should incorporate macroeconomic and qualitative variables, utilize higher-frequency data, and explore hybrid models to improve predictive performance.

References:

Basel Committee on Banking Supervision. (2008). Principles for Sound Liquidity Risk Management and Supervision. Bank for International Settlements.

Boyd, J. H., & Runkle, D. E. (1993). Size and performance of banking firms: The evidence from complementaries and co-insurance. Economica, 60(237), 47-67. https://doi.org/10.2307/2554955

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. Journal of Political Economy, 91(3), 401-419. https://doi.org/10.1086/261155

Dietrich, A., & Wanzenried, G. (2011). Determinants of bank profitability before and during the crisis: Evidence from Switzerland. Journal of International Financial Markets, Institutions and Money, 21(3), 307-327. https://doi.org/10.1016/j.intfin.2010.11.002

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

Hannan, T. H., & Hanweck, G. A. (1988). Bank insolvency risk and the market for large certificates of deposit. Journal of Money, Credit and Banking, 20(2), 203-211. https://doi.org/10.2307/1992111

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67. https://doi.org/10.1080/00401706.1970.10488634

Keynes, J. M. (1936). The General Theory of Employment, Interest, and Money. Macmillan. https://doi.org/10.1007/978-3-319-70344-2

Lepetit, L., & Strobel, F. (2013). Bank insolvency risk and Z-score measures: A refinement. Economics Letters, 120(1), 78-80. https://doi.org/10.1016/j.intfin.2013.01.004.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774. https://doi.org/10.5555/3295222.3295410

Minsky, H. P. (1992). The financial instability hypothesis. Levy Economics Institute Working Paper, (74). https://doi.org/10.2139/ssrn.161024

Modigliani, F., & Miller, M. H. (1963). Corporate income taxes and the cost of capital: A correction. The American Economic Review, 53(3), 433-443. https://doi.org/10.2307/1809167

Nguyen Van Tien (2015). Risk Management in Banking Business. Statistical Publishing House.

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. The American Economic Review, 71(3), 393-410. https://doi.org/10.2307/1802787

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tran Thị Thanh Tu, & Nguyen Thi Hoang Anh (2019). Factors Affecting Liquidity Risk of Vietnamese Commercial Banks. Journal of Banking Science and Training, (206), 1-12.