THE INFLUENCE OF FACTORS ON THE PROBABILITY OF INCURRING BAD DEBT EXCEEDING THE THRESHOLD AT JOINT-STOCK COMMERCIAL BANKS

Thi Van Anh Pham [*], Van Binh Nguyen [**], Huong Quynh Doan [***], Thi Hoai Thu Ho [****]

* Faculty of Postgraduate Education, Academy of Finance, Hanoi, Vietnam
** Faculty of Public Finance, Academy of Finance, Hanoi, Vietnam
*** Faculty of Corporate Finance, Academy of Finance, Hanoi, Vietnam
**** *Corresponding author,* Faculty of Economics, Academy of Finance, Hanoi, Vietnam
Contact details: Faculty of Economics, Academy of Finance, No. 58 Le Van Hien, Dong Ngac Ward, Hanoi 120418, Vietnam

How to cite this paper: Pham, T. V. A., Nguyen, V. B., Doan, H. Q., & Ho, T. H. T. (2026). The influence of factors on the probability of incurring bad debt exceeding the threshold at joint-stock commercial banks. *Risk Governance and Control: Financial Markets & Institutions, 16*(1), 152–162. https://doi.org/10.22495/rgcv16i1p13

ISSN Online: 2077-4303
ISSN Print: 2077-429X

Received: 23.11.2025
Revised: 15.02.2026; 02.03.2026
Accepted: 11.03.2026

JEL Classification: E31, G21, G30, G32
DOI: 10.22495/rgcv16i1p13

Abstract

The study aims to analyze the factors affecting the probability of bad debt exceeding the threshold at listed joint-stock commercial banks in Vietnam during the period 2012–2024. The study examines the relationship between the probability of bad debt exceeding the threshold (Lyra et al., 2015) and independent variables such as credit growth rate, bank size, business efficiency, liquidity, and macroeconomic variables such as inflation rate, economic growth, and money supply by using a binary logit model. The results show that rapid credit growth, low business efficiency, and poor liquidity are factors that significantly increase the probability of bad debt exceeding the threshold. On the contrary, large size and high profitability have an impact on reducing this probability. The study provides important empirical evidence for bank managers in controlling credit risks and contributing to ensuring the safety of the banking and financial system in Vietnam (Chang et al., 2025). The study contributes to the application of the logit model to predict the possibility of bad debt exceeding the threshold, while clearly identifying key factors to enable banks to enhance early warning and control credit risks more effectively.

Keywords: Joint-Stock Commercial Banks, Stock Market, Internal Financial Factors, Macroeconomic Factors, Non-Performing Loans

Authors' individual contribution: Conceptualization — T.V.A.P and T.H.T.H.; Methodology — T.V.A.P.; Validation — V.B.N. and T.H.T.H.; Formal Analysis — T.V.A.P. and V.B.N.; Investigation — V.B.N.; Resources — T.H.T.H.; Data Curation — H.Q.D. and T.H.T.H.; Writing — H.Q.D. and T.H.T.H.; Supervision — H.Q.D.; Project Administration — T.V.A.P.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

Acknowledgements: This research is supported by the Academy of Finance, Vietnam [Grant Number: Decision No. 1422/QĐ-HVTC].

1. INTRODUCTION

Non-performing loans (NPLs) have consistently been regarded as one of the central issues in commercial banks' risk management, particularly in the context of a volatile business environment and sustained credit expansion pressures. In Vietnam, numerous listed commercial banks have experienced situations in which the NPL ratio exceeds the prudential threshold of 3% due to both internal bank-specific factors and macroeconomic conditions (State Bank of Vietnam, 2013). A prolonged increase in NPLs not only undermines banks' profitability and liquidity but also poses systemic risks to the entire financial system, given the heavy reliance of the national financial structure on the banking sector. In Vietnam, listed commercial banks play a dominant role in credit allocation; therefore, effective credit risk monitoring is of critical importance to maintaining financial stability (Anwar et al., 2023; Das & Ghosh, 2007; Saleh & Afifa, 2020).

Though many previous studies have focused on factors affecting the bad debt ratio, most of these studies only consider the level of change in the bad debt ratio as a continuous variable, without analyzing the possibility of bad debt exceeding the threshold, a clearly defined quantitative warning threshold in banking supervision.

The distinguishing feature of this study lies in its departure from the traditional continuous approach. Instead of examining the magnitude of NPLs, the study employs a binary logit model to estimate the probability that the NPL ratio exceeds a prudential threshold, thereby providing a foundation for the development of an early warning system. Under the conventional approach, the scale or severity of problematic loans is typically measured as the percentage of NPLs relative to total loans and treated as a continuous variable (Naili & Lahrichi, 2022; Sharma et al., 2024; Hess et al., 2009). In contrast, this study adopts a prudential supervision framework. Rather than focusing on how large the NPL ratio is, it shifts attention to identifying when the banking system enters a risk state.

From a regulatory perspective, an NPL ratio exceeding the warning threshold may signal the need for timely intervention and appropriate supervisory measures. Therefore, analyzing the probability of threshold exceedance offers a more policy-relevant perspective than merely assessing the level of NPLs. By transforming NPLs into a binary risk variable based on a prudential cutoff, the study moves from an explanatory framework toward a classification and early warning approach.

The main research question of the study is as follows:

RQ: What internal bank-specific factors and macroeconomic variables affect the probability that the NPL ratio exceeds the prudential threshold among listed commercial banks in Vietnam?

The paper is focused on the following objectives:

- to examine the impact of bank-specific characteristics and macroeconomic variables on the probability of threshold exceedance;
- to evaluate model performance using classification accuracy, the receiver operating characteristic (ROC) curve, and the optimal cutoff point;
- to assess the applicability of the model in supporting risk management and policy formulation.

Thus, the study simultaneously examines the effects of internal bank-specific factors (asset-to-loan ratio, profitability — return on assets (ROA), bank size, credit growth) and macroeconomic variables (economic growth, inflation, money supply growth) on the probability of NPL exceeding the threshold. The direction of the effects of these factors is not always clear; therefore, empirical testing is necessary.

Using panel data of listed commercial banks, the study employs a logit model and evaluates predictive performance (similar to Odunlami and Nwonu, 2025). The empirical results show that the asset-to-loan ratio, ROA, bank size, and inflation are statistically significant in explaining the probability of exceeding the NPL threshold. The model reflects an acceptable level of goodness-of-fit and classification performance, supporting its applicability in risk monitoring (Bhandary & Ghosh, 2025; Mileris, 2012).

The contributions of the study are reflected in three aspects. First, it proposes a probability-of-threshold-exceedance approach instead of merely analyzing the level of NPL. Second, it integrates internal and macroeconomic factors within an early warning risk framework. Third, it provides policy implications based on statistically significant quantitative results. The important contribution of the study lies not only in identifying the determinants of NPL but also in quantifying the probability of breaching the risk threshold — an important input for an early warning system.

This particular study is divided into the following sections. Section 1 presents the background of the study, the research question, and the contributions of the study. Section 2 systematizes the relevant studies, develops the theoretical foundation, and builds the research hypotheses. Section 3 presents the research methodology, including data description, sample selection, variables, and the logistic model. Section 4 reports the empirical results. Section 5 provides discussion and policy implications. Section 6 provides the conclusion and directions for future research.

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

2.1. Theoretical background

Research on NPLs is generally divided into two main strands: 1) analyzing the level of NPLs as a continuous variable using linear regression models; and 2) examining risk states through classification models. Traditional studies often focus on the relationship between bank characteristics and the NPL ratio, including several representative studies as follows: Dimitrios et al. (2016) explored the determinants of NPLs using a sample of European banks from 1990 to 2015. The objective of the study was to identify the main factors affecting NPLs in the euro area banking system during the period 1990Q1–2015Q2 using the generalized method of moments (GMM) method. In addition to the bank-specific and country-specific variables examined in previous studies, the roles of income tax and the output gap were tested for the first time and were found to be statistically significant. The research findings provide valuable implications for fiscal policy formulation and macro-prudential supervision. Hassan et al. (2019) compared liquidity risk between Islamic banks and conventional banks during the period 2007–2015 in Islamic countries. The results show that credit risk and liquidity risk have an inverse relationship. For Islamic banks, liquidity risk also reduces stability. Overall, Islamic banks manage risks more effectively than conventional banks (Akram & Rahman, 2018). Kwashie et al. (2022) examined the impact of credit risk management on the financial performance of micro-finance institutions (MFIs) in Nigeria. The results show that factors such as proactive risk assessment, thorough credit appraisal, loan monitoring, and timely debt collection play an important role in improving performance. Credit risk management is further improved by clear regulations, skilled personnel, and appropriate technological infrastructure. However, MFIs still face challenges in credit information, legal framework, and operational constraints. The study recommends increased credit information sharing, stakeholder collaboration, and capacity building to improve performance and sustainability. Naili and Lahrichi (2022) employed the system GMM method using data from developing countries. It presents important factors affecting NPL, including gross domestic product (GDP) growth, unemployment rate, bank capitalization, operating efficiency, ownership concentration, inflation, public debt, and bank size. While credit growth, bank diversification, and interbank competition do not have significant impacts. The results show that NPL is mainly explained by macroeconomic factors and bank-specific characteristics, with notable differences in the degree of impact. This study makes important contributions both theoretically and practically, and provides clear evidence on early indicators of future bad loans. Sharma et al. (2024) used a macroeconomic-based forecasting model and applied algorithms such as decision trees and statistical methods. The results show that inflation and unemployment rates are important macroeconomic indicators, which have a positive relationship with credit risk, while national savings, national debt, and national income have a negative relationship. The forecast accuracy is about 95%. Dao and Minh (2025) conducted research on management practices and employed the structural equation modeling (SEM) approach using real-world data from Vietnam. Identify five main factors affecting credit risk, including: legal environment and credit policy, quality of management and internal control, macroeconomic and political environment, technology and risk management system, asset quality, and customer characteristics, in which legal environment and credit policy have the strongest influence.

Recently, many studies have shifted to emphasizing the role of predictive models in credit risk analysis. For instance, Barasa et al. (2025) applied a logistic regression model, thereby reinforcing the role of this model as a reliable method for probability forecasting. Similarly, Bhandary and Ghosh (2025) reported comparable results.

Thus, the traditional approach does not adequately capture the condition of "risk threshold exceedance", which is of particular concern to supervisory authorities. Recent studies have employed logit or probit models to develop early warning systems, focusing on the probability of high credit risk occurrence. This approach enables the classification of banks into risky and non-risky groups, while also evaluating predictive performance through ROC and area under the curve (AUC) (El Khair Ghoujdam et al., 2024) and the Youden index.

Based on the international compiled studies, it is possible to identify some gaps in the above studies that have not been fully addressed:

- A lack of studies applying credit risk forecasting models based on prudential thresholds.
- A limited number of studies that fully integrate both macroeconomic and micro (bank-specific) factors into quantitative forecasting models. Some studies remain confined to qualitative analysis or examine the two groups of factors separately.
- A lack of practical tools to propose warning thresholds and support risk management decisions. International studies rarely provide specific guidance for establishing operational credit risk warning thresholds within early warning systems.

In Vietnam, studies specifically focusing on the probability of threshold exceedance remain limited. This gap creates an opportunity to employ binary or logistic classification models to predict the probability of threshold exceedance and to propose timely supervisory and control policies.

2.2. Research hypotheses

Based on the existing studies, the research proposes that the determinants affecting the probability of NPLs exceeding the threshold include both internal bank-specific factors and macroeconomic variables, including the asset-to-loan ratio (*ALR*), credit growth rate (*LG*), profitability (*ROA*), bank size (*SIZE*), economic growth (*GP*), inflation (*INF*), money supply growth (*M2*).

Asset-to-loan ratio (*ALR*): The ratio of total assets to total loans, reflecting the degree of asset coverage for financial obligations and the bank's risk-absorbing capacity. A higher *ALR* may reduce the risk of insolvency and limit the likelihood of NPLs. However, a high *ALR* may also reflect asset expansion, including credit growth, which may generate potential risks if asset quality is not strictly controlled. Therefore, this relationship remains an empirical issue that requires further testing. The hypothesis is:

H1: The asset-to-loan ratio has a significant effect on the probability of NPLs exceeding the threshold.

Credit growth rate (*LG*): High credit growth often comes with limited risk control. The hypothesis is:

H2: Credit growth rate has a positive impact on the probability of bad debt exceeding the threshold.

Profitability (*ROA*): Efficient banks tend to manage risks better. The hypothesis is:

H3: Profitability has a negative impact on the probability of NPLs exceeding the threshold.

Bank size (*SIZE*): Large banks tend to have good risk management systems. The hypothesis is:

H4: Bank size has a negative impact on the probability of NPLs exceeding the threshold.

Economic growth (*GP*): High economic growth increases income, increasing customers' ability to repay debt. The hypothesis is:

H5: Economic growth has a negative impact on the probability of bad debt exceeding the threshold.

Inflation (*INF*): Inflation affects credit quality through two different transmission channels: 1) high inflation increases borrowing costs, thereby reducing borrowers' repayment capacity; and 2) inflation reduces the real value of nominal debt, so if borrowers' income adjusts sufficiently in line with price levels, inflation may reduce the real debt burden. The hypothesis is:

H6: Inflation has a significant effect on the probability of NPLs exceeding the threshold.

Money supply growth (*M2*): Overexpansion of the money supply can lead to uncontrolled credit. The hypothesis is:

H7: Money supply growth has a positive impact on the probability of NPLs exceeding the threshold.

3. RESEARCH METHODOLOGY

3.3. Data collection and sampling

The study uses unbalanced panel data of 195 observations, collected from 15 listed commercial banks on the Vietnamese stock market during the period 2012–2024. The data is compiled from reliable sources such as audited financial statements, annual reports of banks, the State Securities Commission Portal, and macroeconomic data from the State Bank of Vietnam (SBV) and the General Statistics Office. After collection, the data is processed, checked for reliability, and standardized to serve quantitative analysis in the risk forecasting model.

From 2012 to 2024, Vietnam's economy underwent a strong restructuring process, particularly in the banking sector. After the crisis, the banking system had to address a large volume of bad debt and undertake the merger of weak institutions. The economy subsequently stabilized, and banks experienced steady growth. Since 2020, COVID-19 has slowed economic growth while accelerating digital transformation in the banking industry. By 2024, Vietnam's banking system will have become more capable of supporting credit risk management and facilitating the formulation of appropriate policy decisions.

3.2. Data analysis

3.2.1. Dependent variable

The binary variable *NPL_class* reflects the probability of falling into the high NPL group:

- *NPL_class* = 1 if the bad debt ratio (NPL) > 3% (warning threshold according to international and Vietnamese standards);
- *NPL_class* = 0 if NPL ≤ 3%.

The optimal threshold for classifying bad debt in the study is 0.01835 (1.835%).

Basis for choosing the optimal threshold:

1) Based on actual data distribution, the data is not evenly distributed around the percent mark as usual, and most banks have NPL < 3%, so choosing a threshold of 3% will cause most of the sample to fall into the same group, and then the model may lose its ability to distinguish.

2) Based on the assessment of model accuracy, the optimal threshold selection method:

- optimize ROC/AUC: find the cutoff point for the best classification;
- optimize the sum of sensitivity and specificity: clearly distinguish the two groups;
- minimize misclassification: choose the cutoff to minimize false predictions.

The study focuses on selecting the optimal ROC/AUC method to determine the threshold score for the best classification.

3.2.2. Independent variable

The independent variables include both bank-internal factors and macro factors, as presented in Table 1.

Table 1. Variable description

Symbol	*Variable name*	*Form*	*Expectation*
ALR	Asset-to-loan ratio	(-)	A higher level of financial soundness is associated with a lower probability of NPLs.
LG	Loan growth	(-)	Strong credit growth may come with risks.
ROA	Return on assets	(-)	High business efficiency, low bad debt.
SIZE	Bank size	(-)	Large banks often control risks better.
GP	Economic growth	(-)	Economic growth will reduce bad debt.
INF	Inflation rate	(+)	High inflation increases the risk of bad debt.
M2	Broad money growth	(+)	An increased money supply can stimulate credit or increase risk.

3.3. Research model

Unlike linear regression models that estimate changes in the NPL ratio, this study employs a binary logistic model to estimate the probability that the NPL ratio exceeds a prudential threshold. This choice reflects the nonlinear nature of the accumulation process within the banking system. From a supervisory perspective, credit risk does not increase in a linear or steady manner. When pressures become sufficiently large and the NPL ratio surpasses a critical limit, a risk state emerges. By transforming the continuous NPL ratio into a binary variable, the model captures the nonlinear relationship between the explanatory variables and the probability of threshold exceedance, thereby aligning with the objectives of classification and early warning.

The study uses a binary logistic regression model to predict the probability of a bank falling into the high bad debt group, with the binary dependent variable *NPL_class*. In this study, to assess the probability of bad debt exceeding the threshold at listed commercial banks, the study uses NPL as the base variable. Based on the practical warning threshold of 1.835%, the NPL variable is classified into the binary variable *NPL_class*, with the value:

- *NPL_class* = 1 if NPL > 0.01835;
- *NPL_class* = 0 if NPL ≤ 0.01835.

The choice of threshold 1.835% is to reflect the risk of bad debt exceeding the standard in the context of the Vietnamese credit market. This is the threshold used to construct the binary variable *NPL_class* from the continuous variable NPL to distinguish between "bad debt exceeding the threshold" and "bad debt not exceeding the threshold".

After building a logistic regression model to predict the probability of *NPL_class* = 1, the study determined the optimal cut-off for the predicted probability (p-hat). The results showed that the threshold of 0.55195 is the optimal cut-off point, determined based on the criterion of maximizing the Youden index (Sensitivity + Specificity – 1), ensuring a balance between the ability to detect correctly and avoid misclassification. This threshold is used to evaluate the effectiveness of the model.

Although the binary model is appropriate for estimating the probability of threshold exceedance, alternative approaches may also be employed, such as panel probit models or dynamic panel techniques using system GMM, which can better control for endogeneity and persistent credit risk. Future research may consider these methods in order to enhance the robustness of the findings.

3.3.1. General regression model

The logit model can be determined as follows in Eq. (1).

$$Log_e[\frac{P(NPL\_class = 1)}{1 - P(NPL\_class = 1)}] = \beta_0 + \beta_1 ALR + \beta_2 LG + \beta_3 ROA + \beta_4 SIZE + \beta_5 GP + \beta_6 INF + \beta_7 M2 \quad (1)$$

where,

- *P (NPL_class = 1)*: probability of a bank being in the high NPL group;
- *ALR*: assets to loan ratio;
- *LG*: credit growth rate;
- *ROA*: return on assets;
- *SIZE*: bank size (logarithm of total assets);
- *GP*: GDP growth;
- *INF*: inflation rate (CPI);
- *M2*: money supply growth rate.

3.3.2. Summary statistics

The statistics presented in Table 2 show the following results: *NPL_class* — target variable, right skewed distribution; *ALR* — low volatility, needs further testing; *LG* — important variable, appropriate deviation; *ROA* — small, low dispersion; *SIZE* — quite stable; *GP* — macro variable with moderate dispersion; *INF* — macro variable, quite volatile; *M2* — good dispersion.

Table 2. Descriptive statistic

Variables	*Average*	*Standard deviation*	*Min*	*Max*
NPL_class	2.24%	1.67%	0.5%	-14.7%
ALR	1.09	0.036	1.04	-1.27
LG	0.20	0.197	-0.46	-1.73
ROA	~0.01	0.007	-0.043	-0.032
SIZE	8.46	0.44	7.33	-9.44
GP	5.83%	1.90%	2.55%	-9.81%
INF	3.07%	1.97%	0.63%	-9.91%
M2	14.3%	5.3%	6.2%	-24.7%

Source: Authors' elaboration.

4. RESEARCH RESULTS

The logistic regression model is used to estimate the probability of a binary dependent variable based on quantitative or qualitative independent variables to assess the impact of each independent variable on the probability of an event occurring.

The logit model table shows that the results of running four iterations show that the model converges well because after four stable iterations, the log-likelihood decreases continuously.

Table 3. Log likelihood results for each iteration

Iteration	*Log likelihood*
0	-135.16114
1	-111.03059
2	-110.87087
3	-110.87075
4	-110.87075

Note: NPL_class: Dependent variable (1 = high risk, 0 = low risk).
Source: Authors' elaboration using Stata software.

Table 4. Logistic regression reflects the overall model information

Variable	*Coefficient*	*Std. err.*	*z*	*p > |z|*	*[95% conf. interval]*
ALR	17.03354	6.868049	2.48	0.013	3.572414 30.49467
LG	-0.7928953	0.9204048	-0.86	0.389	-2.596856 1.011065
ROA	-71.16482	37.37023	-1.90	0.057	-144.4091 2.079485
SIZE	-1.292494	0.5231464	-2.47	0.013	-2.317843 -0.2671463
GP	7.042241	8.857149	0.80	0.427	-10.31745 24.40193
INF	26.58283	12.28566	2.16	0.030	2.503376 50.66228
M2	1.354659	4.577313	0.30	0.767	0.-7.61671 10.32603
_cons	-8.334859	9.928187	-0.84	0.401	-27.79375 11.12403

Note: Dependent variable: NPL_class; Number of obs. = 195; Prob. > chi^2 = 0.0000; Likelihood ratio $chi^2(7)$ = 48.58; Log likelihood = -110.87075; Pseudo R^2 = 0.2797.
Source: Authors' elaboration using Stata software.

Likelihood ratio chi^2 (7) = 48.58 shows the overall model fit. Prob. > chi^2 = 0.0000 is statistically significant for the whole model. Pseudo R^2 0.2797 is good with the logit model (27.97%).

ALR has a p-value = 0.013: statistically significant, reflecting that when *ALR* increases, the probability of falling into the high bad debt group increases sharply. *LG* is not statistically significant and has no clear impact. *ROA* has a p-value = 0.057, which is almost statistically significant, reflecting that when *ROA* increases, the probability of high bad debt decreases sharply. *SIZE* is statistically significant, reflecting that larger banks are less risky. *GP* is not statistically significant. *INF* has a p-value = 0.030, which is statistically significant, reflecting that when inflation increases, credit risk increases. *M2* is not statistically significant.

Thus, the variables with high statistical significance include: *ALR, SIZE,* and *INF. ROA* is close to the threshold of 0.05, so it should be monitored further. *LG, GP,* and *M2* are not statistically significant.

Table 5. Logistic regression shows the meaning of the odds ratio

Variable	*Odds ratio*	*Std. err.*	*z*	*p > |z|*	*[95% conf. interval]*
ALR	2.50e+07	1.72e+08	2.48	0.013	35.60242 1.75e+13
LG	0.4525327	0.4165133	-0.86	0.389	0.0745075 2.748527
ROA	1.24e-31	4.63e-30	-1.90	0.057	1.92e-63 8.000348
SIZE	0.274585	0.1436482	-2.47	0.013	0.0984858 0.7655611
GP	1143.948	10132.12	0.80	0.427	0.0000331 3.96e+10
INF	3.51e+11	4.31e+12	2.16	0.030	12.22369 1.01e+22
M2	3.875438	17.73909	0.30	0.767	0.0004922 30516.63
_cons	0.00024	0.0023828	-0.84	0.401	8.50e-13 67780.5

Note: Dependent variable: NPL_class; Number of obs. = 195; Prob. > chi^2 = 0.0000; Likelihood ratio chi^2(7) = 48.58; Log likelihood = -110.87075; Pseudo R^2 = 0.2797; _cons estimates baseline odds.
Source: Authors' elaboration using Stata software.

Table 5 shows the meaning of the odds ratio: the odds ratio shows the change in the probability of an event (here, falling into the high bad debt group) when the independent variable increases by 1 unit. If the odds ratio > 1, the variable increases the probability of falling into the high bad debt group; if the odds ratio < 1, the variable reduces the probability of falling into the high bad debt group. Specifically for each variable, we have:

- *ALR:* Increasing ALR strongly increases the probability of a bank falling into the high bad debt group.
- *LG*: Not statistically significant.
- *ROA*: Increasing *ROA* strongly reduces the probability of falling into the bad debt group.
- *SIZE*: The larger the bank, the lower the probability of having high bad debt.
- *GP*: Not statistically significant.
- *INF*: Increasing inflation strongly increases credit risk.
- *M2*: not statistically significant.

4.1. Model quality check

The ROC curve in Figure 1 shows the relationship between true positive rate (Sensitivity) and false positive rate (1 – Specificity).

The optimal cut-off (cut-off = 0.55195) is determined at the point closest to the upper left corner, which optimizes the overall accuracy.

The ROC curve has a sharp curve, indicating that the model discriminates well between the two groups.

The predicted probability distribution clearly divides the two groups: NPL = 0 (low risk) and NPL = 1 (high risk).

The cut at 0.55195 separates the two distributions quite clearly, showing the effectiveness of the chosen classification threshold.

The distribution density has overlap, but most of the observations of the two groups are concentrated on two different sides of the threshold.

From Table 6, we can see the following:

- *ALR*: Statistically significant, reflecting that when *ALR* increases by 1 unit, the probability of a bank falling into the bad debt group increases by nearly 3.3 percentage points, showing a large and positive impact.
- *LG*: Not statistically significant.
- *ROA*: Statistically significant, reflecting that when *ROA* increases by 1%, the probability of a bank falling into the bad debt group decreases by nearly 13.8%, showing a negative, strong, and reasonable impact.
- *SIZE*: Statistically significant, reflecting that the larger the bank, the lower the probability of high bad debt.
- *GP*: Not statistically significant.
- *INF:* Statistically significant, reflecting that rising inflation significantly increases the risk of bad debt.
- *M2*: Not statistically significant.

Thus, the variables that have a strong and statistically significant impact on the possibility of high bad debt are: *ALR* (in the same direction); *ROA* (in the opposite direction); *SIZE* (in the opposite direction); *INF* (in the same direction). Some variables, such as *LG, GP,* and *M2,* do not have clear statistical significance and may cause interference. The results of the marginal coefficient analysis show that, when the loan/total asset ratio (*ALR*) increases by 1 unit, the probability that a bank falls into the high bad debt group increases by about 3.3%, with statistical significance at the 1% level. On the contrary, profitability (*ROA*) and bank size (*SIZE*) significantly reduce the possibility of a bank falling into the high bad debt group.

Figure 1. Model evaluation visualization: ROC curve and predicted probability

Figure 1a. ROC curve and optimal cut-off

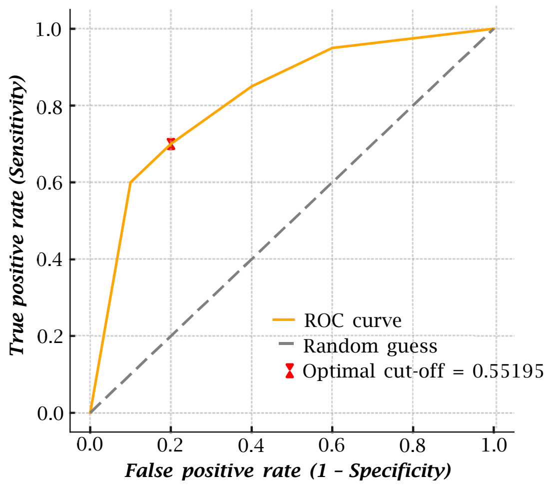

Figure 1b. Predicted probability distribution by class

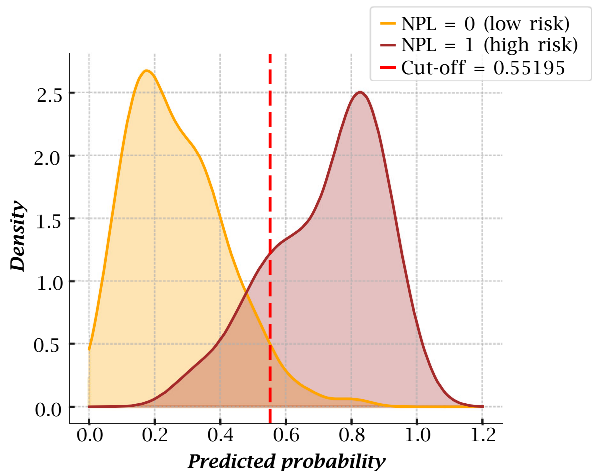

Table 6. Marginal effects of independent variables: margins, dy/dx*

Variable	*dy/dx*	*Std. err.*	*z*	*p > |z|*	*[95% conf. interval]*
ALR	3.298593	1.254418	2.63	0.009	8399795 5.757207
LG	-0.1535464	0.1770224	-0.87	0.386	-0.5005038 0.1934111
ROA	-13.78127	6.994048	-1.97	0.049	-27.48935 -0.0731856
SIZE	-0.2502952	0.0954951	-2.62	0.009	-0.4374621 -0.0631283
GP	1.36375	1.704509	0.80	0.424	-1.977027 4.704526
INF	5.147839	2.278545	2.26	0.024	6819731 9.613705
M2	0.2623334	0.8857373	0.30	0.767	-1.47368 1.998347

*Note: * marginal effects are calculated for all independent variables in the model. Number of obs. = 195; Variance-covariance estimator model: Observed information matrix; Expression: Pr(NPL_class), predict(); dy/dx wrt: ALR, LG, ROA, SIZE, GP, INF, M2.*
Source: Authors' elaboration using Stata software.

4.2. Validation and assessment of the suitability of the logistic model

Pearson chi² test with the following assumptions:

- Assumption 1: The model fits the data well.
- Assumption 2: The model does not fit the data well.

Table 7. Model fit: Goodness-of-fit test after logistic model

Variable	*No. of obs.*	*No. of covariate patterns*	*Pearson chi² (187)*	*Prob. > chi²*
NPL_class	195	195	185.36	0.5202

Source: Authors' elaboration using Stata software.

Table 7 analyzes the detailed results of the model fit test, showing that:

- Pearson chi² (187) = 185.36, with p-value = 0.5202: the model fits the data well.
- p-value = 0.5202 > 0.05: do not reject Assumption 1.

Conclusion: As the p-value = 0.5205 exceeds the 0.05 significance level, Assumption 1 cannot be rejected, suggesting that the model fits the data well and exhibits no evidence of serious bias

Thus, the logistic regression model built on financial quantitative variables has demonstrated its effectiveness in predicting the possibility of bad debt exceeding the threshold. The evaluation indicators and visual charts both reinforce the accuracy and stability of the model.

4.3. Improvement of the model

Implemented through feature selection: remove insignificant variables ($p > 0.1$) such as *LG*, *GP*, *M2*, to increase model stability: Reduce noise, improve interpretability.

Rerun the logistic model with only variables that have clear statistical significance ($p < 0.05$): *ALR*, *ROA*, *SIZE*, and *INF*.

Table 8. Logistic regression

Panel A: Log likelihood results for each iteration					
Iteration			*Log likelihood*		
0			-135.16114		
1			-111.63204		
2			-111.47483		
3			-111.47472		
4			-111.47472		
Panel B: Logistic regression					
Variable	*Coefficient*	*Std. err.*	*z*	*p > |z|*	*[95% conf. interval]*
ALR	17.75557	6.794115	2.61	0.009	4.439346 31.07179
ROA	-81.94622	36.56364	-2.24	0.025	-153.6096 -10.28281
SIZE	-1.238087	0.4826377	-2.57	0.010	-2.18404 -0.2921345
INF	31.33965	9.942976	3.15	0.002	11.85178 50.82753
_cons	-9.198563	9.379489	-0.98	0.327	-27.58202 9.184899

Note: Dependent variable: NPL_class; Number of obs. = 195; Likelihood ratio chi²(4) = 47.37; Log likelihood = -111.47472; Pseudo R² = 0.3752.
Source: Authors' elaboration using Stata software.

Table 9 shows that the logistic model estimation process performed through the maximum likelihood estimation method has converged at iteration (because the log-likelihood does not change anymore). The improvement from -135.16 to -111.47 shows that the independent variables have made the model more suitable than the empty model (with only constants). This shows that the estimation algorithm has found the optimal set of parameters, and the model is eligible to continue to evaluate the quality through tests such as the likelihood ratio test (likelihood ratio chi²) and pseudo R^2.

Results of model quality assessment:

- Pseudo $R^2 = 0.3752$: relatively stable with logistic model.
- Prob. > chi^2 = 0.0000: the model is statistically significant overall.
- All variables in the model are highly statistically significant ($p < 0.05$).
- When *ALR* increases, the bank is more likely to have bad debt or significantly increases the probability of bad debt.
- When *ROA* increases, the bank is better at managing cash flow, reducing the risk of bad debt.
- *SIZE*: the larger the bank, the lower the probability of bad debt (due to better control capacity).
- *INF*: a high inflation environment increases the risk of bad debt.

Conclusion: The reduced model is considered more effective because it retains only variables with high statistical significance while eliminating statistically insignificant variables. The reduced regression model is built to eliminate variables with no statistical significance, helping to simplify and improve the accuracy of the model. The results show that the four independent variables *ALR, ROA, SIZE,* and *GP* still play an important role in predicting the probability of bad debt exceeding the threshold.

In addition, the marginal effects plot shows the marginal impact of each variable on the probability of bad debt occurrence. In particular, the impact of *INF* and *ALR* is the strongest in the positive direction, while *ROA* and *SIZE* have a negative impact. This strengthens the model's validity and supports the empirical arguments. In addition, the predicted probability distribution plot and the confusion matrix confirm that the cutoff threshold of 0.55195 is appropriate for classification. The model achieves a good balance between sensitivity and specificity, as evidenced by the ROC plot with a fairly good AUC.

The reduced regression equation has the following logit form:

$$Log(P/(1-P)) = \beta_0 + \beta_1 ALR + \beta_2 ROA + \beta_3 SIZE + \beta_4 INF$$
$$Log(P/(1-P)) = -9.198563 + 17.75557\,ALR - 81.94622\,ROA - 1.238087\,SIZE + 31.33965\,INF \quad (2)$$

4.4. Researching the lag effect of policy

Table 9 indicates that both *INF* and *INF_lag1* exert positive effects and are statistically significant at the 5% level. Current inflation (*INF*) increases the probability of exceeding the NPL threshold, as rising prices immediately elevate funding costs and repayment pressure on borrowers. In contrast, the one-year lag effect shows that the influence of inflation on credit risk persists over time, not only in the current year but also for at least one subsequent year. This finding confirms that the impact of macroeconomic variables on NPLs is characterized by a noticeable lagged effect.

Table 9. Logistic regression with lagged variables (Part 1)

Panel A: Log likelihood results for each iteration	
Iteration	*Log likelihood*
0	-124.58866
1	-99.170641
2	-98.81722
3	-98.815832
4	-98.815832

Table 9. Logistic regression with lagged variables (Part 2)

Panel B: Logistic regression					
Variable	***Coefficient***	***Std. err.***	***z***	***p > |z|***	***[95% conf. interval]***
ALR	15.05172	6.884974	2.19	0.029	1.557415 28.54602
ROA	-59.79226	36.96325	-1.62	0.106	-132.2389 12.65438
SIZE	-1.346039	0.5221745	-2.58	0.010	-2.369482 -0.3225953
INF	39.78685	18.20681	2.19	0.029	4.102166 75.47154
INF_lag1	30.62289	12.74623	2.40	0.016	5.640725 55.60505
_cons	-6.780131	9.569096	-0.71	0.479	-25.53521 11.97495

Note: Dependent variable: NPL_class. The number of observations decreases due to the loss of data in the first year when constructing the lagged variable.
Source: Authors' elaboration using Stata software.

Although *ROA* is generally expected to be an important determinant of credit risk, the results of the lagged regression model indicate that *ROA* becomes statistically insignificant once the *INF_lag1* variable is incorporated. This may be because the effect of *ROA* on NPLs is partially "absorbed" by inflation or, in other words, overshadowed when a strong macroeconomic variable is added to the model. Nevertheless, the *ROA* coefficient remains negative, as theoretically expected, confirming its economically meaningful role in reducing the probability of NPL occurrence. However, under conditions of lagged inflation, the magnitude of this effect is not sufficiently strong to achieve statistical significance. This modeling outcome accurately reflects the nature of credit risk in Vietnam, where pronounced macroeconomic fluctuations tend to dominate bank-level determinants in explaining variations in NPL formation.

Table 10. Comparison between the baseline model and the lagged model

Model	***Log-likelihood***	***Parameter***	***Akaike information criterion (AIC)***	***Bayesian information criterion (BIC)***
Model 1: *INF*	-111.47	5	232.95	249.31
Model 2: *INF + INF_lag1*	-98.82	6	209.63	228.77

Source: Authors' elaboration.

Based on the results in Table 10, Model 2 exhibits lower AIC and BIC values, indicating a better fit and superior predictive capability. This provides further confirmation that Model 2 is the most appropriate specification for the data.

The reduction in AIC from 232.9 to 209.6 represents a statistically meaningful improvement of approximately 14%, demonstrating that the inclusion of *INF_lag1* enhances the model's ability to explain the probability of NPL occurrence more effectively than the current inflation variable (*INF*).

The regression model with lagged effects is described in Eq. (3).

$$Log(P/(1-P)) = \beta_0 + \beta_1 ALR + \beta_2 SIZE + \beta_3 INF + \beta_4 INF\_lag1$$
$$Log(P/(1-P)) = -6.780131 + 15.05172\, ALR - 1.346039\, SIZE + 39.78685\, INF + 30.62289\, INF\_lag1 \quad (3)$$

5. DISCUSSION OF RESEARCH RESULTS

The reduced logistic regression results with the dependent variable *NPL_class* show that the model is statistically significant and indicates good classification ability (AUC = 0.768). The independent variables, including *ALR, ROA, SIZE*, and *INF*, are all significant at the 95% confidence level, reflecting a significant role in influencing the probability of incurring bad debt exceeding the threshold.

The *ALR* variable has a positive and statistically significant coefficient, indicating that a larger asset size does not necessarily imply better asset quality. In a highly competitive environment, banks experiencing asset expansion may accelerate credit growth in order to maximize profits, thereby accepting a higher level of risk. Therefore, asset growth may be accompanied by an increase in risky loans, leading to a higher probability of exceeding the NPL threshold. This is similar to the study of Hassan et al. (2019).

The *SIZE* variable also has a negative coefficient, indicating that larger banks are often better able to control risks thanks to better governance and resources, thereby reducing the probability of bad debts exceeding the threshold (Hassan et al., 2019).

The *INF* variable has a positive and statistically significant impact, reflecting the adverse influence of macroeconomic factors. This indicates that during the study period, the effect of inflation in increasing borrowing costs and reducing repayment capacity outweighed its effect in lowering the real value of nominal debt. This impact is consistent with the characteristics of developing economies such as Vietnam, where inflation is often accompanied by interest rate adjustments and macroeconomic volatility (Sharma et al., 2024).

The lagged inflation variable (*INF_lag1*) is also statistically significant and exhibits a strong effect, indicating that credit risk tends to accumulate and materialize after approximately one year. This lagged impact reflects the typical weakening cycle of firms: rising costs reduce profitability, which subsequently deteriorates cash flows and heightens the likelihood of overdue payments, ultimately contributing to the formation of NPLs in the banking sector.

Based on the final model, which indicates the statistically significant effects of *ALR, SIZE, INF*, and *INF_lag1* on the probability of exceeding the NPL threshold, the study proposes several policy implications for regulatory authorities and Vietnamese commercial banks as follows.

First, strengthening asset structure management. The positive and statistically significant coefficient of the asset-to-loan ratio suggests that bank safety should not be assessed solely on the basis of asset size; rather, it is necessary to examine asset structure, the proportion of risk-weighted assets, and the degree of credit concentration in order to properly evaluate underlying risk exposure. Excessive concentration of assets in lending activities may increase the probability of exceeding the NPL threshold. Therefore, commercial banks should: 1) avoid excessive concentration of assets in loan portfolios and diversify investment channels; 2) enhance the rigor and quality of credit appraisal, particularly during periods of credit expansion; and 3) allocate and utilize assets flexibly to maintain a balanced and prudent asset structure.

From a regulatory perspective, supervisory authorities should strengthen close monitoring of banks with persistently high asset-to-loan ratio levels, especially during periods of rapid credit growth.

Second, enhance financial capacity and management at small-scale banks. The findings suggest that larger banks demonstrate stronger resilience to credit risk. Therefore:

- The SBV should encourage restructuring and capital strengthening among smaller banks to improve system-wide stability.
- Small-scale banks need to invest heavily in risk management systems, customer data infrastructure, and analytical technologies.
- Forming alliances or promoting consolidation among weaker institutions should be considered to enhance overall financial strength.

This result does not imply that large banks are risk-free; rather, it emphasizes the role of structural capacity in reducing the probability of threshold exceedance.

Third, conducting a stable monetary policy to control credit risk. Both current inflation and one-period lagged inflation are positive and statistically significant, indicating that inflation increases the probability of exceeding the NPL threshold. The significance of the lagged variable reflects the persistent nature of this effect; therefore, inflation stability should be maintained. In addition, the SBV should incorporate lagged inflation dynamics into the macroprudential early warning system for the banking sector. At the same time, monetary policy should target medium-term inflation stability rather than reacting solely to short-term fluctuations in current inflation.

Finally, policy coordination between SBV and the Government should be reinforced to support macroeconomic stability.

Credit risk is inherently cyclical and influenced by the macroeconomic environment. Therefore, macroeconomic stabilization policies play a fundamental role in limiting the probability of NPLs exceeding the prudential threshold. Strengthening policy coordination helps to reduce systemic risk, mitigate excessive credit cycles, enhance the sustainability of the banking system.

Thus, the study employs a threshold exceedance probability model (rather than analyzing the magnitude of NPLs) to quantify the probability that a bank enters a risk state. At the same time, it demonstrates the persistent effect of inflation on credit risk. Accordingly, the study provides empirical evidence from listed commercial banks in Vietnam by employing a dynamic logistic model, thereby enriching the literature on credit risk in emerging markets.

6. CONCLUSION

The study employs a logistic regression model to analyze the factors influencing the probability of exceeding the NPL threshold among listed commercial banks in Vietnam, incorporating both macroeconomic and bank-specific variables. The empirical results indicate that the final model — including *ALR, ROA, SIZE*, current inflation (*INF*), and lagged inflation (*INF_lag1*) — is the optimal specification, offering better goodness of fit and stronger explanatory power compared with earlier model versions. The findings reveal that banks' credit risk is strongly affected by asset structure and the broader macroeconomic environment.

The main contribution of the study lies primarily in shifting the analytical focus from the magnitude of NPLs to the probability of exceeding a prudential safety threshold. Based on these findings, the study highlights the critical importance of controlling credit growth, strengthening the financial capacity of smaller banks, and particularly closely monitoring inflation dynamics and their lagged impacts on the banking system. The proposed policy implications aim to support regulatory authorities and commercial banks in enhancing risk management effectiveness, stabilizing credit quality, and fostering the long-term sustainability of the financial banking system amid an increasingly volatile economic environment.

Although the study makes certain contributions, it still has several limitations. First, the research sample is limited to the listed joint-stock commercial banks. Second, potential endogeneity issues have not been entirely eliminated within the static logistic model. Third, the macroeconomic variables considered are primarily focused on inflation and its lagged value. Future research may expand the model by incorporating additional systemic risk indicators and dynamic specifications in order to provide a more comprehensive and in-depth assessment.

REFERENCES

Akram, H., & Rahman, K. u. (2018). Credit risk management: A comparative study of Islamic banks and conventional banks in Pakistan. *ISRA International Journal of Islamic Finance, 10*(2), 185–205. https://doi.org/10.1108/IJIF-09-2017-0030

Anwar, C. J., Suhendra, I., Didu, S., Sayektiyani, A., & Kholishoh, L. N. (2023). The impact of monetary policy and credit risk on bank credit behavior: An analysis of banks listed on the Indonesian Stock Exchange. *Cogent Economics & Finance, 11*(1), Article 2220250. https://doi.org/10.1080/23322039.2023.2220250

Barasa, I. N., Wanyonyi, S. W., & Kololi, M. M. (2025). Application of logistic regression in enhancing digital credit risk management in commercial banks. *Asian Journal of Probability and Statistics, 27*(2), 13–26. https://doi.org/10.9734/ajpas/2025/v27i2710

Bhandary, R., & Ghosh, B. K. (2025). Credit card default prediction: An empirical analysis on data-driven classifiers with logistic regression. *Journal of Risk and Financial Management, 18*(1), Article 23. https://doi.org/10.3390/jrfm18010023

Chang, V., Xu, Q. A., Akinloye, S. H., Benson, V., & Hall, K. (2025). Prediction of bank credit worthiness through credit risk analysis: An explainable machine learning study. *Annals of Operations Research, 354*, 247–271. https://doi.org/10.1007/s10479-024-06134-x

Dao, L. T., & Minh, N. Q. (2025). Critical factors affecting bank credit risk: A case study of commercial banks in Vietnam. *Ianna Journal of Interdisciplinary Studies, 7*(1), 477–486. https://iannajournalofinterdisciplinarystudies.com/index.php/1/article/view/525

Das, A., & Ghosh, S. (2007). *Determinants of credit risk in Indian state-owned banks: An empirical investigation* (MPRA Paper No. 17301). Munich Personal RePEc Archive (MPRA). https://mpra.ub.uni-muenchen.de/17301/1/MPRA_paper_17301.pdf

Dimitrios, A., Helen, L., & Mike, T. (2016). Determinants of non-performing loans: Evidence from Euro-area countries. *Finance Research Letters 18*, 116–119, https://doi.org/10.1016/j.frl.2016.04.008

El Khair Ghoujdam, M., Chaabita, R., Elkhalfi, O., Zehraoui, K., Elalaoui, H., & Idamia, S. (2024). Consumer credit risk analysis through artificial intelligence: A comparative study between the classical approach of logistic regression and advanced machine learning techniques. *Cogent Economics & Finance, 12*(1), Article 2414926. https://doi.org/10.1080/23322039.2024.2414926

Hassan, M. K., Khan, A., & Paltrinieri, A. (2019). Liquidity risk, credit risk and stability in Islamic and conventional banks. *Research in International Business and Finance, 48*, 17–31. https://doi.org/10.1016/j.ribaf.2018.10.006

Hess, K., Grimes, A., & Holmes, M. (2009). Credit losses in Australian banking. *Economic Record, 85*(270), 331–343. https://doi.org/10.1111/j.1475-4932.2009.00551.x

Kwashie, A. A., Baidoo, S. T., & Ayesu, E. K. (2022). Investigating the impact of credit risk on financial performance of commercial banks in Ghana. *Cogent Economics & Finance, 10*(1), Article 2109281. https://doi.org/10.1080/23322039.2022.2109281

Lyra, M., Onwunta, A., & Winker, P. (2015). Threshold accepting for credit risk assessment and validation. *Journal of Banking Regulation, 16*(2), 130–145. https://doi.org/10.1057/jbr.2013.26

Mileris, R. (2012). Macroeconomic determinants of loan portfolio credit risk in banks. *Inzinerine Ekonomika-Engineering Economics, 23*(5), 496–504. https://doi.org/10.5755/j01.ee.23.5.1890

Naili, M., & Lahrichi, Y. (2022). Banks' credit risk, systematic determinants and specific factors: Recent evidence from emerging markets. *Heliyon, 8*(2), Article e08960. https://doi.org/10.1016/j.heliyon.2022.e08960

Odunlami, B. G., & Nwonu, B. (2025). Credit risk prediction using ensemble and linear machine learning models. *Journal of Advanced Artificial Intelligence, 2*(1). https://jaaionline.org/archives/volume2/number1/credit-risk-prediction-using-ensemble-and-linear-machine-learning-models/

Saleh, I., & Afifa, A. M. (2020). The effect of credit risk, liquidity risk and bank capital on bank profitability: Evidence from an emerging market. *Cogent Economics & Finance, 8*(1), Article 1814509. https://doi.org/10.1080/23322039.2020.1814509

Sharma, H., Andhalkar, A., Ajao, O., & Ogunleye, B. (2024). Analysing the influence of macroeconomic factors on credit risk in the UK banking sector. *Analytics, 3*(1), 63–83. https://doi.org/10.3390/analytics3010005

State Bank of Vietnam. (2013). Circular No. 02/2013/TT-NHNN providing the Classification of Assets, Risk Provisioning Levels and Methods and Use of Provisions for Handling Risks in Operations of Credit Institutions and Foreign Bank Branches. https://english.luatvietnam.vn/circular-no-02-2013-tt-nhnn-dated-january-21-2013-of-the-state-bank-of-vietnam-providing-the-classification-of-assets-risk-provisioning-levels-and-76429-doc1.html