Transactions on Computational and Scientific Methods | Vo. 6, No. 01, 2026
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Early Warning and Prediction of Systemic Financial Risk Using Machine Learning Methods

Fran Pinheiro[1], Jun Kyu Kim[1], Han Bao[1]
[1]Columbia University, New York, USA
*Corresponding author: Han Bao; baohan66221@gmail.com

Abstract: Machine learning techniques offer significant advantages over traditional economic modeling methods, particularly in capturing complex nonlinear relationships, making them well suited for systemic financial risk analysis. This study develops a machine learning-based framework for monitoring and early warning of systemic financial risk, supported by both theoretical analysis and empirical evidence. Early warning indicators are constructed across eight dimensions, including macroeconomic fundamentals, monetary conditions, fiscal status, financial markets, price dynamics, foreign exchange markets, leverage levels, and the banking system. Five representative machine learning models, along with ensemble learning approaches, are employed to forecast systemic financial risk. Empirical results indicate that machine learning models consistently outperform traditional linear models in both in-sample and out-of-sample settings. While the Lasso model achieves superior short-term forecasting performance, the SVM model demonstrates stronger predictive capability over longer horizons. Ensemble models effectively balance predictive accuracy and robustness. Furthermore, partial dependence analysis enhances model interpretability by revealing nonlinear effects and key risk drivers. Exchange rates, money supply, market interest rates, and industrial product prices emerge as critical determinants of systemic financial risk. Targeted monitoring of these variables can support timely risk identification and early intervention.

Keywords: Systemic financial risk; Early warning system; Machine learning; Nonlinear relationships

1. Introduction

In recent years, artificial intelligence technologies-particularly machine learning-have brought profound transformations across many industries, promoting innovation and development in related fields. The financial sector is no exception. With the continuous advancement of financial technology, finance has undergone or is undergoing significant transformations at multiple levels. In the field of financial risk management, as modern financial systems become increasingly complex, the limitations of traditional risk modeling approaches have become more evident. In contrast, machine learning methods are well suited to capturing complex nonlinear relationships among variables. Compared with conventional economic analysis and forecasting techniques, machine learning offers notable potential advantages, enabling more effective modeling and prediction of economic and financial systems, which are complex and open systems. Owing to these advantages, academic research on artificial intelligence and machine learning has expanded rapidly. In top-tier economics and finance journals, an increasing number of studies have applied machine learning models to address economic and financial problems. For example, prior research has examined the effectiveness of machine learning models in predicting bond risk premia [1,2], investigated the application of machine learning in measuring financial market microstructure [3], and analyzed investor decision-making under private information using machine learning techniques [4].

With respect to financial risk monitoring and early warning, early studies have demonstrated that machine learning has considerable potential in predicting currency crises, banking crises, and economic recessions [5-7]. However, financial crises or systemic financial risk events occur relatively infrequently, which leads to limited sample sizes in empirical modeling. This constraint may restrict the learning and predictive capability of machine learning models. Using quantitative indicators of systemic financial risk as prediction targets can help mitigate this limitation. Existing systemic financial risk early warning models developed in the literature are mostly based on linear analytical methods to select relevant indicators for risk prediction [8]. Some studies have constructed indicator systems and applied models such as the Markov regime-switching framework to identify high- and low-risk regimes [9,10]. Only a small number of studies attempt to employ specific machine learning models to forecast systemic financial risk [11-14].

The existing literature mainly selects macroeconomic and financial market indicators to predict systemic financial risk, which can be further categorized into macroeconomic indicators (e.g., economic growth, inflation, interest rates, money supply), banking sector indicators (e.g., total social financing, capital adequacy ratios, non-performing loan ratios), stock market indicators (e.g., returns, volatility, liquidity), bond market indicators (e.g., yield curves, credit spreads, term spreads), and foreign exchange market indicators (e.g., foreign exchange reserves, exchange rates). For instance, a financial stress index constructed by a major central banking institution incorporates 18 financial variables, including interest rates, spreads, and stock market volatility [15]. Other studies have examined the predictive power of more than thirty financial indicators-including economic growth, foreign exchange reserves, money supply, and interest rates-for sovereign debt crises [16]. Additional research has selected indicators from multiple dimensions, such as debt exposure, external sector conditions, macroeconomic performance, and the banking sector, to construct early warning systems for systemic financial risk [17]. Some studies have developed financial stress indices covering multiple subsystems, including banking, bond, equity, foreign exchange, real estate, and derivatives markets, and further synthesized static and dynamic weighted financial stress indices [18]. Related work has also selected indicators from several financial sub-markets-such as credit, capital, foreign exchange, bond, and money markets-to measure financial system pressure [19]. Furthermore, comprehensive indicator systems incorporating banking, securities markets, international reserves, international trade, balance of payments, economic growth, money supply, and household living conditions have been proposed to predict systemic financial risk [20].

Based on this literature, this paper approaches systemic financial risk from a global and endogenous perspective and constructs an indicator system at the macro level of the economic-financial system. The focus is on comparing and optimizing appropriate machine learning methods to develop a more effective nonlinear monitoring and early warning framework for systemic financial risk. Specifically, this study compares mainstream machine learning models-including Lasso regression, support vector machines (SVM), random forests (Random Forest), extreme gradient boosting (XGBoost), and neural networks-as well as their ensemble models.

Compared with the existing literature on financial risk prediction and early warning, the incremental contributions and innovations of this paper are as follows. First, while most studies employ a single machine learning model to predict systemic financial risk, this paper proposes a nonlinear early warning system based on the comparative analysis of multiple machine learning models and further integrates optimized model combinations. Empirical results demonstrate that the proposed system achieves more accurate estimation and prediction of systemic financial risk. Second, due to the “black-box” nature of machine learning, few studies conduct comprehensive and effective analyses of variable importance within machine learning models. This paper applies multiple interpretable modeling techniques to partially open the black box of machine learning, thereby enhancing the economic interpretability and stability of the models. Specifically, feature importance measures (including SDT) and partial dependence plots (PDP) are employed to evaluate the relative importance of variables, enabling an examination of the internal structure of the prediction models and improving their credibility and reliability.

2. Models and Methods

Mainstream machine learning models include regularized regression methods, support vector machines, tree-based models (such as decision trees, random forests, and boosting models), as well as neural networks. No single machine learning model consistently outperforms others across all forecasting tasks. Therefore, for a given prediction problem, it is necessary to select the most appropriate model based on practical considerations. In this study, Lasso regression, support vector machines (SVM), random forests, extreme gradient boosting (XGBoost), neural networks, and ensemble modeling approaches are selected as candidate methods. By comparing individual models and constructing ensemble models, this study aims to identify optimal modeling strategies for systemic financial risk forecasting and early warning.

2.1 Machine Learning Models

To account for the nonlinear and high-dimensional nature of systemic financial risk, this study applies a diverse suite of machine learning models, including Lasso, ridge regression, elastic net, support vector machines (SVM), random forests (RF), extreme gradient boosting (XGBoost), and neural networks. The decision to incorporate neural and deep learning models is supported by Xu et al. [21], who demonstrated that attention-based deep architectures can significantly enhance forecasting performance in financial time series by dynamically capturing latent dependencies and abrupt shifts in risk dynamics.

Recognizing that single models may offer limited robustness across complex economic regimes, we also adopt ensemble learning methods to aggregate model outputs. Inspired by the ensemble logic used in recent interpretable financial AI frameworks (Long et al. [22]), we design two types of ensemble strategies. The first relies on weighting-based averaging, including both equal-weight and DAR-weight schemes. This approach aligns with the growing consensus that model aggregation can mitigate overfitting and balance performance across different data segments. The second ensemble strategy draws inspiration from the multi-layer frameworks utilized in distributed anomaly detection systems, as exemplified by Feng et al. [23], who demonstrated the potential of hierarchical architectures to integrate heterogeneous model outputs for improved decision-making. Building on this idea, we design a two-layer ensemble model in which predictions from multiple base learners serve as inputs to a linear regression meta-model. This architecture enables the incorporation of diverse predictive signals in a structured manner.

2.2 Model Evaluation Metrics

To comprehensively evaluate the predictive performance of the proposed models, this study adopts a set of widely used evaluation metrics that capture both numerical accuracy and directional consistency of forecasts. Specifically, root mean square error (RMSE) and mean absolute error (MAE) are employed to measure the magnitude of prediction errors, with RMSE placing greater emphasis on large deviations and MAE providing a robust assessment of average absolute discrepancies. In addition, Theil's inequality coefficient (Theil-U) is included as a scale-independent metric to assess relative forecasting efficiency compared with naive benchmark models.

Beyond error magnitude, directional accuracy (DAR) is incorporated to evaluate the model's ability to correctly predict the direction of change in the target variable, which is particularly important in decision-making scenarios where the sign of variation carries substantial practical implications. Furthermore, the coefficient of determination (R^2) is reported to quantify the proportion of variance in the observed data that can be explained by the predictive model, thereby providing an overall measure of goodness-of-fit.

Together, these metrics offer a comprehensive and balanced evaluation framework by jointly assessing prediction precision, stability, explanatory power, and directional reliability. The formal definitions and calculation procedures of these evaluation metrics are presented in the formulas below:

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

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i-y_i|$$

$$\text{Theil-U}=\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(y_i-y_{i-1})^2}$$

$$\mathrm{DAR}=\frac{1}{n}\sum_{i=1}^{n}a_i$$

$$a_i=\begin{cases}1, & (y_{i+1}-y_i)(\hat{y}_{i+1}-y_i)>0\\ 0, & \text{otherwise}\end{cases}$$

$$y_i=\gamma_0+\gamma_1\hat{y}_i+\varepsilon_i$$

Smaller values of RMSE, MAE, and Theil-U indicate better predictive performance. Directional accuracy (DAR) ranges between 0 and 1, with larger values reflecting higher accuracy in predicting the direction of change. The coefficient of determination R^2 evaluates overall model fit, where higher values correspond to stronger predictive capability.

2.3 Variable Importance Analysis Methods

Commonly used interpretable machine learning techniques include feature importance measures, surrogate decision trees (Surrogate Decision Trees, SDT), partial dependence plots (Partial Dependence Plot, PDP), and Shapley values. Among these approaches, PDP not only allows for the identification of the relative importance of variables but also enables the examination of their nonlinear effects on model outputs. Given these advantages, this study adopts partial dependence plots as the primary tool for analyzing variable importance.

The core idea of the PDP approach is to evaluate how changes in the values of a specific feature influence the model's predicted output while holding all other features constant. Specifically, one feature is fixed at a given constant value, while the remaining features are kept unchanged, generating a new set of model inputs. The trained model is then used to produce predictions based on these modified inputs. By comparing the resulting predictions with the original model outputs, the marginal effect of the selected feature can be observed, thereby revealing its relative importance and influence on the prediction results.

In this study, partial dependence analysis is applied to the random forest, XGBoost, and SVM models to examine and compare the contribution of individual features across different machine learning frameworks.

3. Variable Selection and Description

3.1 Risk Prediction and Early Warning Indicator System

Existing studies on systemic financial risk prediction generally select macroeconomic and financial market indicators, which mainly include macroeconomic indicators, banking sector indicators, stock market indicators, bond market indicators, and foreign exchange market indicators. Drawing on prior research [20], and taking data availability and timeliness into account, this study constructs a comprehensive early warning indicator system for systemic financial risk.

Specifically, indicators are selected from eight dimensions: macroeconomic fundamentals, monetary conditions, fiscal conditions, securities and interest rate markets, price indices, foreign exchange and exchange rate markets, leverage ratios, and the banking system. In total, 42 indicators are chosen to form the initial indicator set for systemic financial risk prediction and early warning. The complete list of indicators and their corresponding classifications is presented in Table 1.

Table 1: Early Warning Indicator System for Systemic Financial Risk

Indicator Dimension	Indicator Name	Symbol	Indicator Dimension	Indicator Name	Symbol
Macroeconomic Fundamentals	Industrial value added (YoY)	VAI	Price Indices	CPI (YoY)	CPI
	Fixed asset investment completed (cumulative YoY)	FixInv		PPI of industrial products (YoY)	PPI
	Urban real estate development investment (YoY)	RealEs		Retail price index (YoY)	RPI
	Total retail sales of consumer goods (YoY)	Retail		National housing climate index	House
	Urban unemployment rate	Employ		Export value (YoY)	Export
	Purchasing Managers' Index	PMI		Import value (YoY)	Import
	OECD composite leading indicator	OECD	Foreign Exchange and Exchange Rate	Growth rate of foreign exchange reserves	ForeEx
	Business climate index	Prosper		Foreign direct investment (YoY)	FDI
	Consumer confidence index	Consum		Foreign assets / GDP	Foreign
Monetary Conditions	Total social financing	Social		Net capital inflow	Hot
	M1 (YoY)	M1		RMB real effective exchange rate index	RMBEx
	M2 (YoY)	M2	Leverage Ratios	Household sector debt / GDP	HouDebt

	Monetization ratio (M2/GDP)	M2/GDP		Non-financial corporate sector debt / GDP	Nonfin
Fiscal Conditions	Fiscal budget revenue (YoY)	BudRev		Government sector debt / GDP	GovDebt
	Fiscal budget expenditure (YoY)	BuEx		Real economy sector debt / GDP	RealDebt
Securities and Interest Rate Markets	Total market value of listed companies / GDP	MV/GDP		Financial sector asset-side debt / GDP	FinDebt
	Stock market trading volume (YoY)	Volume		Financial sector liability-side debt / GDP	LiaDebt
	Interbank pledged repo weighted average rate	Repoe	Banking System	Total deposits (YoY)	Deposit
	Government bond yield (1-year maturity)	Bond1Y		Total loans (YoY)	Loans
	Government bond yield (10-year maturity)	Bond10Y		Loan-to-deposit ratio	LTD
	Corporate bond yield (1-year maturity)	Corp1Y		Non-performing loan ratio of commercial banks	BLR

3.2 Systemic Financial Risk

Systemic financial risk is measured using CoVaR and MES. Based on the estimated CoVaR and MES values of all financial institutions, cross-sectional averages are computed to construct a macro-level indicator of systemic financial risk. According to the secondary industry classification, as of the end of 2024, the listed financial institutions included banks, insurance companies, and securities firms. Due to data availability constraints, all indicators are constructed using observations from January 2018 onward.

The computation of CoVaR and MES relies on daily data, while the market benchmark is proxied by the Wind All-A index. Daily observations are aggregated into monthly values by averaging within each month. All early warning indicators are measured at a monthly frequency. Missing values in some indicators are addressed using linear smoothing methods. Given the relatively high data requirements of machine learning models, all variables are subjected to 95% winsorization and standardization prior to model estimation. All raw data are obtained from the Wind database.

Figure 1 illustrates the dynamic evolution of systemic financial risk from January 2018 to December 2024. The results show a strong co-movement between CoVaR and MES, indicating a high degree of consistency between the two measures. Both indicators exhibit pronounced increases during periods of heightened financial stress, followed by gradual normalization, suggesting their effectiveness in capturing cyclical fluctuations in systemic financial risk.

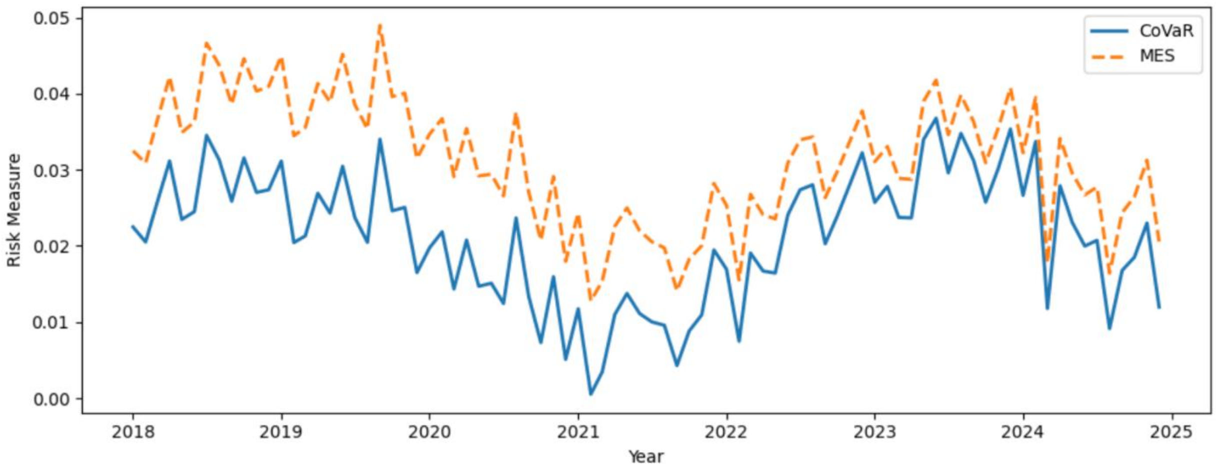

Figure 1. Trend of Systemic Financial Risk

3.3 Predictive Performance for CoVaR

Given the highly similar dynamic patterns observed for CoVaR and MES, this study focuses primarily on CoVaR as the main target variable, while MES is used for robustness analysis. Data from January 2018 to December 2023 are used for in-sample model estimation and parameter tuning, whereas data from January 2023 to December 2024 are reserved for out-of-sample evaluation. The dependent variable is the one-period-ahead CoVaR. Explanatory variables include the lagged CoVaR and 42 early warning indicators observed from period t-11 to period *t*, corresponding to information from the previous twelve months. This setup allows the models to forecast systemic financial risk one month ahead.

All models are estimated using optimized hyperparameters and are evaluated based on both in-sample and out-of-sample performance. Table 2 reports the prediction results. Columns 1 through 5 present the forecasting performance of XGBoost, random forest, SVM, Lasso, and artificial neural network models, respectively. Columns 6 and 7 report the results of two ensemble models. Specifically, Ensemble 1 combines the predictions of XGBoost, random forest, and SVM, while Ensemble 2 integrates XGBoost, random forest, SVM, and Lasso. For comparison, Column 8 reports the prediction results obtained from a linear OLS regression model.

It is worth noting that the ensemble models are constructed by computing the simple arithmetic average of predictions from individual learners, which is conceptually similar to the bagging strategy commonly used in random forest models. Compared with single-model forecasts, ensemble approaches are generally able to achieve improvements in both predictive accuracy and stability. In this study, multiple ensemble configurations based on two, three, and four individual models (excluding neural networks) are examined. Overall, the ensemble model incorporating four base learners demonstrates the best predictive performance and robustness. To further illustrate the advantages of ensemble learning, the predictive performance of ensembles combining XGBoost, random forest, and SVM is also reported in subsequent analysis.

From the in-sample results shown in Panel A, all models exhibit relatively strong fitting performance. In particular, the XGBoost model achieves the lowest RMSE and a directional accuracy (DAR) close to unity, with the coefficient of determination approaching one. Although the OLS regression model performs worse than machine learning models in terms of in-sample fit, it still attains a relatively high R^2 value of approximately 0.88. However, based on the out-of-sample results reported in Panel B, the Lasso model delivers the best forecasting performance, with RMSE, MAE, and Theil-U values smaller than those of other models, as well as superior DAR and R^2. The SVM model ranks second in terms of predictive accuracy.

In contrast, the artificial neural network exhibits very weak out-of-sample forecasting performance, characterized by relatively large RMSE, MAE, and Theil-U values, as well as a very low R^2 of approximately 0.02, which is even lower than that of the OLS model. This result indicates severe overfitting of the neural network model in the in-sample period, likely due to limited sample size and the curse of dimensionality. Consequently, neural networks are excluded from subsequent analysis.

Comparing the two ensemble models, Ensemble 1 outperforms the SVM model, indicating that although XGBoost and random forest individually exhibit weaker out-of-sample performance, combining them with SVM can enhance overall predictive accuracy. Ensemble 2 further improves forecasting performance by incorporating the Lasso model, leading to more stable and accurate predictions than Ensemble 1.

Overall, the results suggest that selected machine learning models-particularly SVM and Lasso-are effective in forecasting one-period-ahead systemic financial risk. Moreover, compared with most single-model approaches (except Lasso), ensemble models provide more robust and stable predictive performance.

Figure 2 presents the out-of-sample fitted curves of three models-Lasso, Ensemble 2, and OLS-where the solid line represents the realized value of systemic financial risk. It can be observed that the fitted trajectories produced by the machine learning models closely track the actual dynamics of systemic financial risk. In particular, the fitted series generated by the Lasso model exhibits a high correlation of 0.95 with the realized series, while the fitted series from Ensemble 2 also shows a strong correlation of 0.93. In contrast, the fitted curve obtained from the OLS model deviates substantially from the actual series.

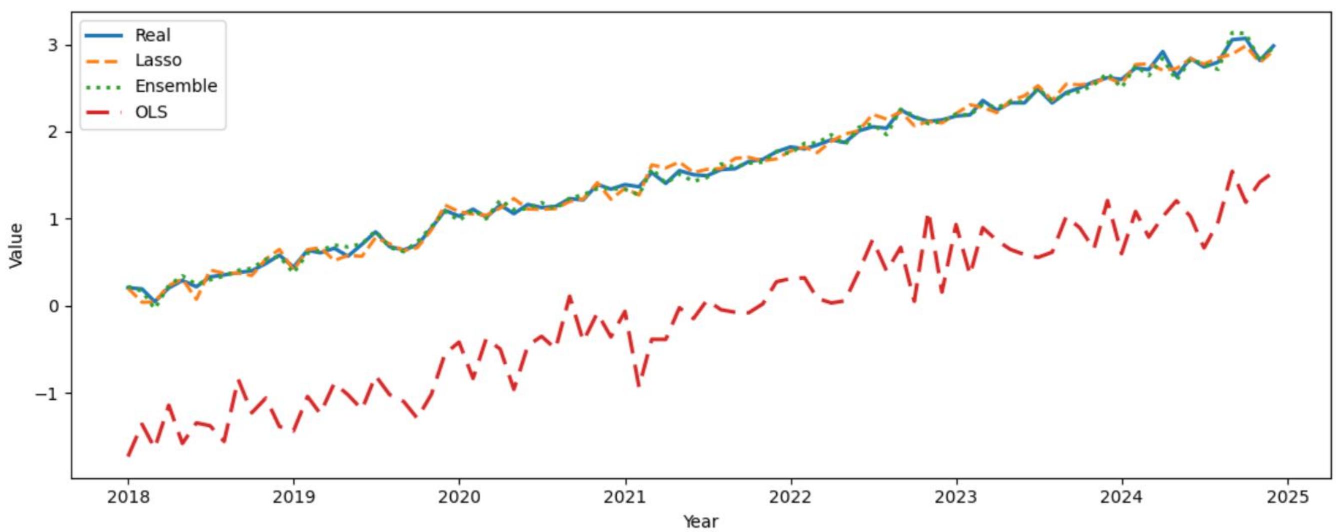

Figure 2. Out-of-Sample Fitting Results of CoVaR

These results provide clear evidence of the effectiveness of machine learning models in forecasting systemic financial risk, whereas traditional linear models exhibit very limited out-of-sample predictive capability.

Table 2: In-Sample and Out-of-Sample Fitting Results for CoVaR

	XGB	RF	SVM	Lasso	ANN	Ensemble	Ensemble2	OLS
Panel A: In-sample								

RMSE	0.0028	0.1085	0.0953	0.2481	0.0128	0.061	0.101	0.3
MAE	0.0022	0.0718	0.0914	0.1728	0.0038	0.0513	0.0732	0.2403
Theil-U	0.0013	0.0517	0.0459	0.1211	0.006	0.029	0.0485	0.1439
DAR	0.992	0.896	0.92	0.744	0.984	0.936	0.904	0.624
R^2	0.9999	0.9841	0.9877	0.917	0.9998	0.995	0.9862	0.8786
Panel B: Out-of-sample								
RMSE	0.5212	0.474	0.3748	0.2177	1.6514	0.3482	0.3081	1.9311
MAE	0.4638	0.4101	0.3168	0.1491	1.5452	0.2818	0.2458	1.8042
Theil-U	0.3628	0.2962	0.3224	0.1622	0.7384	0.2562	0.2284	0.7553
DAR	0.5714	0.6857	0.6286	0.7143	0.5714	0.6571	0.6571	0.4
R^2	0.7764	0.7785	0.8237	0.9088	0.0222	0.832	0.8599	0.1484

3.4 Variable Importance Analysis

The Lasso model allows direct assessment of variable importance through estimated coefficients. In the fitted Lasso model, only five variables have non-zero coefficients: the year-on-year growth rate of average daily stock trading volume (lagged by two periods, coefficient 0.0406), the year-on-year change in PPI (lagged by one period, coefficient-0.002), the yield of one-year government bonds (lagged by one period, coefficient-0.0036), the yield of ten-year government bonds (lagged by twelve periods, coefficient 0.0002), and lagged CoVaR (lagged by one period, coefficient 0.8973). These results indicate strong persistence in systemic financial risk and highlight the substantial influence of stock market liquidity, producer prices, and government bond yields on risk dynamics.

For tree-based models and kernel-based methods such as random forest, XGBoost, and SVM, variable importance cannot be obtained directly from model parameters. Therefore, this study employs partial dependence plots (PDP) to interpret these machine learning models and to mitigate their black-box nature.

The core idea of the PDP approach is to fix a specific variable at different values and examine its impact on the predictive performance of the originally fitted model. Since all variables are standardized prior to model estimation, each variable is sequentially assigned values of -1.5, -1, -0.5, 0, 0.5, 1, and 1.5. The resulting changes in model performance-measured by RMSE-are then evaluated. A rapid increase in RMSE indicates relatively high importance of the corresponding variable, whereas a small change in RMSE suggests lower importance. In addition to assessing relative importance, PDPs also reveal nonlinear effects by illustrating how model sensitivity varies across different values of a given variable.

Figure 3 reports the PDP curves of two representative variables-the RMB real effective exchange rate and the yield of one-year corporate bonds. The results show pronounced nonlinear effects in the SVM model. When the variable values are close to zero, RMSE remains relatively low, indicating limited impact on model performance. However, as the variable values move away from zero-particularly toward the left tail-their importance increases substantially. Both variables also exhibit strong nonlinear patterns in the XGBoost model. Specifically, the RMB exchange rate has a limited effect when its value is below 0.5, but once it exceeds 1, it significantly deteriorates model performance. The influence of the one-year corporate bond yield is concentrated mainly in the interval between -0.5 and 0. Similar nonlinear relationships are observed in the random forest model, although their magnitude is weaker compared with SVM and XGBoost.

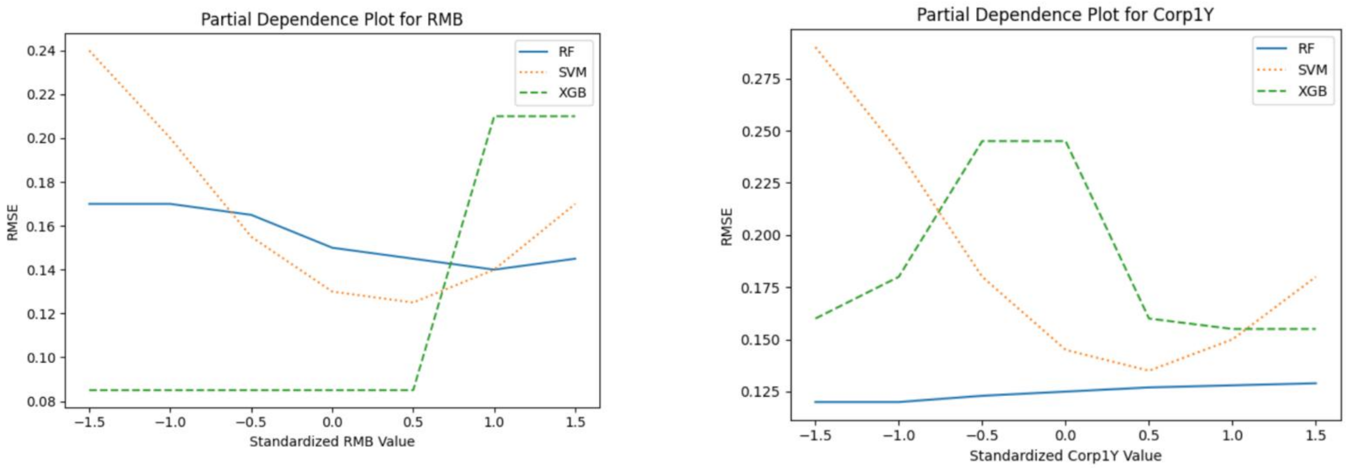

Figure 3. Partial dependence plots (PDPs) for selected variables.

Overall, Figure 3 clearly demonstrates that the relationships between explanatory variables and systemic financial risk are not purely linear but exhibit substantial nonlinear characteristics.

To further evaluate relative importance across variables, the effects at different values are averaged for each variable. Figure 4 presents the ten most influential variables, excluding CoVaR itself. PPI, the RMB exchange rate, the one-year corporate bond yield, and the one-year government bond yield consistently exert strong effects across all models. In addition, foreign assets, the business climate index, and CPI show high importance in the XGBoost model, while consumer confidence and the loan-to-deposit ratio play a more prominent role in the SVM model.

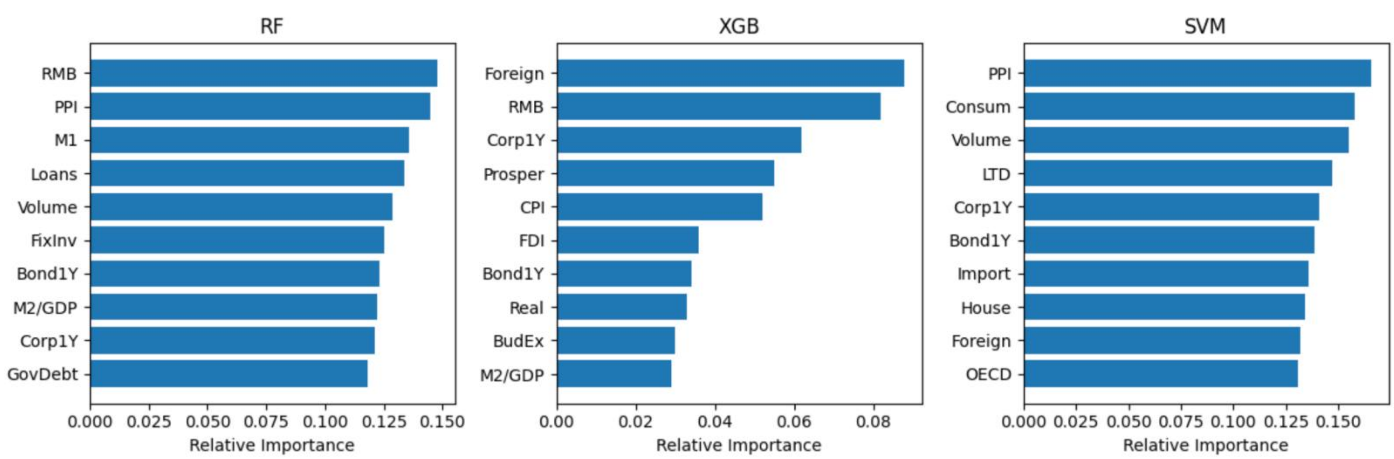

Figure 4. Relative variable importance based on partial dependence analysis.

3.5 Analysis Based on MES

In this subsection, CoVaR is replaced by MES as the dependent variable, and the same modeling and evaluation procedures are applied for robustness analysis (results omitted for brevity). The empirical findings are largely consistent with those reported in Table 2. Both machine learning models and the OLS model exhibit strong in-sample fitting performance, with XGBoost again achieving the best results. In particular, the RMSE decreases to 0.0038, directional accuracy (DAR) reaches 0.97, and the coefficient of determination approaches unity.

However, in the out-of-sample evaluation, the predictive performance of XGBoost, SVM, and OLS deteriorates substantially, whereas random forest and Lasso exhibit stronger forecasting ability, with Lasso performing particularly well. Ensemble models continue to demonstrate superior performance. The Ensemble model significantly outperforms random forest, while Ensemble2 performs slightly worse than Lasso. These results further confirm that ensemble learning is an effective approach for improving predictive accuracy and robustness.

Overall, machine learning models-especially Lasso and ensemble-based approaches-are effective in forecasting systemic financial risk measured by MES. Although traditional linear models can achieve reasonable in-sample fit, their out-of-sample predictive capability remains limited.

Variable importance analysis based on PDP is also conducted for MES-based prediction models (figures omitted). The results are largely consistent with those reported in Figure 4. The RMB exchange rate, M1, and the one-year corporate bond yield remain among the most influential variables across all three models. In contrast to CoVaR-based prediction, PPI plays a less prominent role, while the OECD leading indicator and bank lending variables become more important. This difference may stem from the distinct definitions of CoVaR and MES. Specifically, CoVaR primarily captures extreme downside risk, whereas MES measures the expected loss conditional on extreme market downturns, reflecting average losses beyond tail risk.

Taken together, regardless of whether CoVaR or MES is used to quantify systemic financial risk, machine learning models exhibit strong predictive capability and similar patterns in variable importance. These findings suggest that early warning systems should integrate information from multiple risk measures and indicators to enable more refined monitoring and timely mitigation of systemic financial risk.

3.6 Medium- and Long-Term Forecasting

Motivated by the strong short-term forecasting performance of machine learning models, this subsection extends the prediction horizon to examine their ability to forecast medium- and long-term systemic financial risk. Specifically, three-month-ahead and six-month-ahead forecasts are conducted.

For three-month-ahead prediction, the lagged values of CoVaR and all early warning indicators from period t-11to period ttt are used to predict systemic financial risk at period t+3. Similarly, for six-month-ahead prediction, the same set of lagged explanatory variables is used to forecast systemic financial risk at period t+6. The modeling strategy is extended accordingly. Table 3 reports the out-of-sample forecasting results.

Compared with the one-period-ahead forecasts reported in Table 2, most models exhibit weaker performance in medium- and long-term prediction, as reflected by higher RMSE, MAE, and Theil-U values, as well as lower DAR and R^2. However, the SVM model demonstrates relatively stronger performance in longer-horizon forecasting than in short-term prediction. Both ensemble models continue to deliver stable performance, particularly Ensemble2, which integrates four different base learners. Ensemble2 exhibits strong robustness across different forecasting horizons, performing consistently well in both three-month-ahead and six-month-ahead predictions.

Overall, these results indicate that while predictive accuracy generally declines as the forecast horizon increases, ensemble learning methods-especially those incorporating a diverse set of base models-offer superior stability and robustness in medium- and long-term forecasting of systemic financial risk.

Table 3: Medium- and Long-Term Forecasting Performance Based on CoVaR

Panel A: 3-Month-Ahead Forecast							
Metric	XGB	RF	SVM	Lasso	Ensemble	Ensemble2	OLS
RMSE	0.7461	0.7336	0.4124	0.5458	0.6054	0.5815	2.0493
MAE	0.6686	0.6573	0.3476	0.4693	0.5521	0.5302	1.9621
Theil-U	0.5745	0.5903	0.4219	0.4595	0.5292	0.5075	0.826
DAR	0.4545	0.5152	0.6364	0.6364	0.5758	0.6061	0.4545
R^2	0.4394	0.5797	0.7456	0.6259	0.6385	0.6546	0.1424
Panel B: 6-Month-Ahead Forecast							
Metric	XGB	RF	SVM	Lasso	Ensemble	Ensemble2	OLS
RMSE	1.2625	0.9189	0.3511	0.5047	0.8178	0.5655	2.2177
MAE	1.138	0.8371	0.305	0.3948	0.7217	0.5129	2.1605
Theil-U	0.8529	0.7934	0.5181	0.4211	0.7857	0.6857	0.8935
DAR	0.5	0.5667	0.6333	0.6667	0.6	0.6	0.6333
R^2	0.0027	0.2443	0.6691	0.4699	0.1908	0.3512	0.1154
Panel C: 12-Month-Ahead Forecast							
Metric	XGB	RF	SVM	Lasso	Ensemble	Ensemble2	OLS
RMSE	0.8602	0.6603	0.309	1.0048	0.572	0.541	2.5066
MAE	0.7494	0.5609	0.2413	0.8308	0.4769	0.4798	2.4722
Theil-U	0.7723	0.9116	0.4047	0.5937	0.7404	0.5654	0.9647
DAR	0.4583	0.4583	0.625	0.5417	0.4167	0.5833	0.625
R^2	0.2609	0.2019	0.1377	0.0507	0.2341	0.1549	0.0014

4. Conclusion

This study proposes a systemic financial risk monitoring and early warning framework based on machine learning techniques, and provides both theoretical justification and empirical validation from multiple perspectives. Specifically, using monthly macroeconomic and financial data from January 2018 to December 2024, early warning indicators are constructed from eight dimensions, including economic fundamentals, monetary supply, fiscal conditions, securities and interest rate markets, price indices, foreign exchange and exchange rate markets, leverage, and the banking system. Both linear regression and five representative machine learning models are employed to predict systemic financial risk.

The empirical results yield several key findings. First, compared with the OLS model, machine learning models that are capable of capturing nonlinear relationships demonstrate superior predictive performance in both in-sample fitting and out-of-sample forecasting. Second, regardless of whether systemic financial risk is measured by CoVaR or MES, the Lasso model consistently exhibits strong predictive ability. Although XGBoost, random forest, and neural network models achieve outstanding in-sample performance, their out-of-sample performance is relatively unstable. This suggests that when sample size is limited, more complex machine learning models are more prone to overfitting, which in turn leads to weaker out-of-sample predictive accuracy. Third, the SVM model performs particularly well in multi-step-ahead forecasting. Moreover, compared with single machine learning models (except SVM), ensemble models are able to outperform most individual predictors, thereby enhancing the robustness and reliability of forecasting results.

Finally, partial dependence plots (PDPs) are shown to be effective in improving the interpretability of machine learning models by revealing nonlinear effects of key variables on systemic financial risk. Among all early warning indicators, exchange rates, monetary supply, market interest rates, and industrial price indices emerge as the most influential factors. Strengthening the monitoring of these variables can therefore contribute to earlier detection and more effective prevention of systemic financial risk.

References

[1] D. Bianchi, M. Büchner and A. Tamoni, "Bond risk premiums with machine learning", The Review of Financial Studies, vol. 34, no. 2, pp. 1046-1089, 2021.

[2] H. R. Seddighi and W. Nian, "The Chinese stock exchange market: operations and efficiency", Applied Financial Economics, vol. 14, no. 11, pp. 785-797, 2004.

[3] D. Easley, M. López de Prado, M. O'Hara and Z. Zhang, "Microstructure in the machine age", The Review of Financial Studies, vol. 34, no. 7, pp. 3316-3363, 2021.

[4] P. Akey, V. Grégoire and C. Martineau, "Price revelation from insider trading: Evidence from hacked earnings news", Journal of Financial Economics, vol. 143, no. 3, pp. 1162-1184, 2022.

[5] A. Zamanian, M. Aslani and M. Hematfar, "A Spatial-Deep Learning Hybrid Model for Cryptocurrency Pricing and Optimal Trading Strategy Design for Capital Management", Business, Marketing, and Finance Open, pp. 1-40, 2025.

[6] G. Kou, X. Chao, Y. Peng, F. E. Alsaadi and E. Herrera-Viedma, "Machine learning methods for systemic risk analysis in financial sectors", Technol. Econ. Dev. Econ., vol. 25, no. 5, pp. 716-742, 2019.

[7] A. Samitas, E. Kampouris and D. Kenourgios, "Machine learning as an early warning system to predict financial crisis", International Review of Financial Analysis, vol. 71, Art. no. 101507, 2020.

[8] V. Acharya, R. Engle and M. Richardson, "Capital shortfall: A new approach to ranking and regulating systemic risks", American Economic Review, vol. 102, no. 3, pp. 59-64, 2012.

[9] A. Tobias and M. K. Brunnermeier, "CoVaR", The American Economic Review, vol. 106, no. 7, pp. 1705-1741, 2016.

[10] D. Bisias, M. Flood, A. W. Lo and S. Valavanis, "A survey of systemic risk analytics", Annu. Rev. Financ. Econ., vol. 4, no. 1, pp. 255-296, 2012.

[11] J. Danielsson, M. Valenzuela and I. Zer, "Learning from history: Volatility and financial crises", The Review of Financial Studies, vol. 31, no. 7, pp. 2774-2805, 2018.

[12] G. Kou, Y. Peng and G. Wang, "Evaluation of clustering algorithms for financial risk analysis using MCDM methods", Information Sciences, vol. 275, pp. 1-12, 2014.

[13] M. Drehmann and M. Juselius, "Evaluating early warning indicators of banking crises: Satisfying policy requirements", International Journal of Forecasting, vol. 30, no. 3, pp. 759-780, 2014.

[14] J. Sirignano and R. Cont, "Universal features of price formation in financial markets: perspectives from deep learning", Machine Learning and AI in Finance, Routledge, pp. 5-15, 2021.

[15] M. Billio, M. Getmansky, A. W. Lo and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors", Journal of Financial Economics, vol. 104, no. 3, pp. 535-559, 2012.

[16] C. R. Proa ń o, C. Schoder and W. Semmler, "Journal of International Money and Finance", Journal of International Money and Finance, vol. 45, pp. 17-37, 2014.

[17] R. Amirzadeh, A. Nazari, D. Thiruvady and M. S. Ee, "Modelling determinants of cryptocurrency prices: A bayesian network approach", arXiv preprint, arXiv:2303.16148, 2023.

[18] A. Namaki, R. Eyvazloo and S. Ramtinnia, "A systematic review of early warning systems in finance", arXiv preprint, arXiv:2310.00490, 2023.

[19] L. Alessi and C. Detken, "Identifying excessive credit growth and leverage", Journal of Financial Stability, vol. 35, pp. 215-225, 2018.

[20] M. L. Duca and T. A. Peltonen, "Assessing systemic risks and predicting systemic events", Journal of Banking & Finance, vol. 37, no. 7, pp. 2183-2195, 2013.